Online Appendix to
"Revealing temptation through menu choice: field evidence"

Séverine Toussaert

February 10, 2019

Table of Contents

A. Reimbursement program

A.1 Descriptive statistics about the rank orderings

Figure 1 shows the proportion of times a given reimbursement option was assigned each of the 7 ranks. Options G, GO and GOR were assigned rank 1 most of the time. Option GO is a top choice for most respondents, with 78% assigning it either rank 1 or 2. The distribution of ranks is more spread out for G and GOR: while both were assigned either rank 1, 2 or 3 by at least 70% of respondents, GOR was assigned rank 4 by 19% (the second most frequent choice with rank 2), and G was assigned rank 4 or 6 by 18%. Option O is the option for which the distribution of ranks is the most evenly spread: although most of the mass is distributed around rank 3 to 5, rank 1 and 6 were each assigned at least 10% of the time. Options GR, OR and R have modal ranks at 5, 6 and 7 respectively, with a distribution more concentrated towards the lower ranks; in particular, R was assigned one of the 3 lowest ranks more than 80% of the time. Overall, option GO seems to be the most popular option followed by GOR and G while R is, as expected, the least popular option.

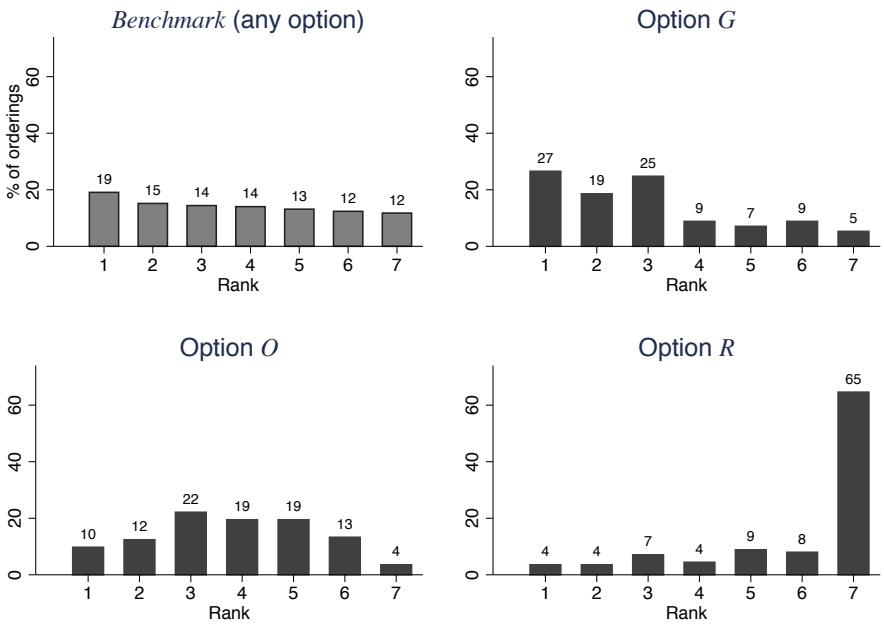
Tables 1 & 2 show the distribution of indifferences. Given the iterative elicitation procedure, participants with a strict preference ordering completed their ranking in 7 steps, while it took only one step for those who were fully indifferent. 82.3% of respondents gave a strict ordering of the 7 options (thus completing their ranking in 7 steps), with the remaining expressing at least one indifference. Among those who expressed indifferences, 30% assigned rank 1 to three reimbursement options and a different rank to each of the remaining options; this is the most common pattern, corresponding to 6 of the 7 participants who completed their ranking in 5 steps. The most common indifferences involved options comparing G and O. Slightly over 10% of respondents assigned the same rank to G and GO. About 7% (resp. 6%) were indifferent between G and O (resp. GO and O).

Table 3 presents the distribution of preference orderings over the 3 singleton options, G, O and R. Nearly 60% expressed the ordering $G \succ O \succ R$ by ranking G strictly above O and O strictly above R. R is strictly dominated by G for 86.7% (98) of participants and by O for 83.2% (94). The preference for G vs. O is slightly more split, with 69.0% (78) of participants strictly preferring G to O.

Table 4 presents the distribution of top choices that would be observed if option GO was removed from the set of options, that is, looking at the restriction of $\succeq$ to $\mathcal{M} \setminus \{GO\}$. This exercise assumes that participants satisfy some form of Independence of Irrelevant Alternatives when ordering items from $\mathcal{M} \setminus \{GO\}$. The take-up rate of commitment to eating healthier foods would drop to 32% (compared to 48% when either G or GO could be a top choice), while 43% would now strictly prefer the unrestricted coverage, GOR.

FIGURE 1: Distribution of ranks by reimbursement option

(A) Singleton options



(B) Doubleton and tripleton options

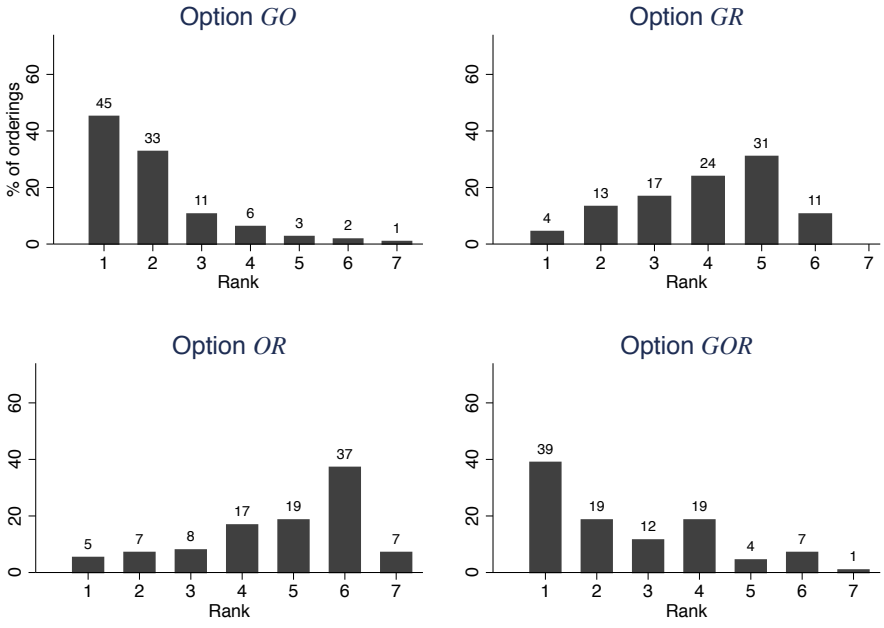


TABLE 1: Proportion of indifferences

Number of steps to complete ordering	Actual sample % (N)	Coarseness of the preference ordering
1	1.8 (2)	$\Rightarrow$ full indifference $\sim$
2	0.0 (0)	
3	2.7 (3)	
4	3.5 (4)	
5	6.2 (7)	
6	3.5 (4)	
7	82.3 (93)	$\Rightarrow$ strict preference $\succ$
% (freq.) of indifferences	4.3 (103/2373)	

Notes: With $|\mathcal{M}|$=7, there are $\binom{7}{2} = 21$ binary comparisons between reimbursement options per individual. Summing over $N = 113$, the total number of binary comparisons is therefore 2373 in the actual sample.

TABLE 2: List of binary comparisons with indifferences

Binary comparison	% (freq.) of participants
$G \sim GO$	10.6 (12/113)
$G \sim O$	7.1 (8/113)
$GO \sim O$, $OR \sim R$	6.2 (7/113)
$G \sim GOR$, $GO \sim GOR$, $O \sim R$	5.3 (6/113)
$O \sim OR$, $O \sim GR$, $O \sim GOR$, $GR \sim OR$, $GR \sim GOR$	4.4 (5/113)
$GO \sim OR$, $GO \sim GR$	3.5 (4/113)
$G \sim GR$, $R \sim GR$, $R \sim GOR$, $OR \sim GOR$	2.6 (3/113)
$G \sim R$, $G \sim OR$, $GO \sim R$	1.8 (2/113)
% (freq.) of $\sim$ in full sample	4.4 (104/2373)

Notes: With $|\mathcal{M}|$=7, there are $\binom{7}{2} = 21$ binary comparisons between reimbursement options for each individual. Summing over all 113 respondents, the total number of binary comparisons is therefore 2373.

TABLE 3: Distribution of preferences over G, O and R

Ordering	% (N)
$G \succ O \succ R$	58.4 (66)
$G \succ R \succ O$	4.5 (5)
$O \succ G \succ R$	15.9 (18)
$O \succ R \succ G$	3.5 (4)
$R \succ G \succ O$	3.5 (4)
$R \succ O \succ G$	3.5 (4)
$G \sim O \succ R$	5.3 (6)
$G \sim R \succ O$	0.0 (0)
$O \sim R \succ G$	0.9 (1)
$G \succ O \sim R$	2.7 (3)
$O \succ G \sim R$	0.0 (0)
$R \succ G \sim O$	0.0 (0)
$G \sim O \sim R$	1.8 (2)
Total	100 (113)

TABLE 4: Distribution of top choices in the absence of GO

Top option	Actual sample % (N)	Benchmark %	p-value
G	31.9 (36)	14.5	< 0.001
GOR	43.4 (49)	14.5	< 0.001
Other	24.8 (28)	71.0	< 0.001
Total	100 (113)	100.0	

Notes: "Other" refers to participants who either (i) had their unique top in the set $\{O, R, GR, OR\}$ or (ii) assigned rank 1 to several options. Reported p-values are the result of a two-sided binomial test that the observed frequency is equal to the benchmark frequency.

A.2 Strength and structure of temptation

A.2.1. Global Temptation Index

In the main text, I present the distribution of a *Global Temptation Index*, which measures the frequency with which an individual prefers to exclude a given food category from the coverage when comparing two nested options. Formally, I define the Global Temptation

Index for food category $x \in \{G, O, R\}$ as

$$GT_{-x} = \sum_{\mathcal{M}_x} \mathbb{1}_{\{M \setminus \{x\} \succ M\}}$$

where $\mathcal{M}_x = \{M \in \mathcal{M} \mid x \in M \text{ and } M \neq \{x\}\}$. This index is based on the definition of temptation as an option that triggers a desire for commitment i.e., option m is a *temptation* in menu M if $M \setminus \{x\} \succ M$. One could argue that this definition is not tight enough for it does not take into account an individual's preferences over the singletons. For instance, should R be considered as a temptation for the individual if $G \succ GR$ but $R \succ G$? In this case, it is plausible that preference for commitment captures other concerns than temptation.[1] Below I therefore consider a tighter definition of the Global Temptation Index that requires additional restrictions on preferences over the singleton options G, O and R. Formally, say that x is a *temptation* in menu M if (i) $M \setminus \{x\} \succ M$ and (ii) $\exists\, x' \in M \setminus \{x\}$ such that $\{x'\} \succ \{x\}$. In other words, x is a temptation in M if the DM would prefer to exclude x from M and there is at least one other option in M that normatively beats it. A stricter version of the Global Temptation Index for food category $x \in \{G, O, R\}$ is then given by

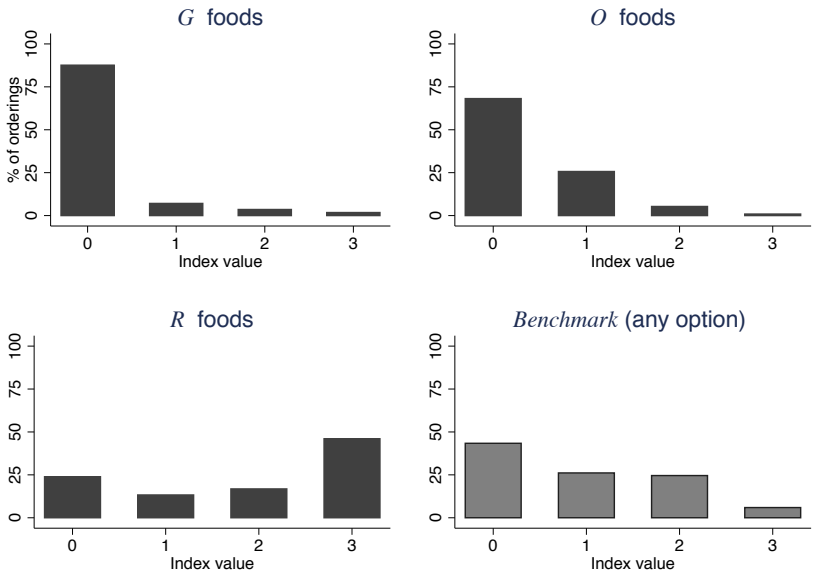
$$\widetilde{GT}_{-x} = \sum_{\widetilde{\mathcal{M}}_x} \mathbb{1}_{\{M \setminus \{x\} \succ M\}}$$

where $\widetilde{\mathcal{M}}_x = \{M \in \mathcal{M} \mid x \in M \text{ and } \exists\, x' \in M \setminus \{x\} \text{ such that } \{x'\} \succ \{x\}\}$. Thus, $\widetilde{GT}_{-R} = GT_{-R}$ if $G \succ R$ and $O \succ R$ (as is the case for 90/113 respondents), but the two indices may differ otherwise. Figure **2** shows the distribution of this alternative index of global temptation for G, O and R; as a point of comparison, the distribution generated under the random benchmark is shown in the fourth quadrant (see main text for a definition of the random benchmark). The two indices differ from each other in relatively few instances: 7 for option R, 10 for O and 16 for G. For 89 of the 113 respondents, there is no discrepancy between the two indices on any of the 3 options. With a more restrictive definition of temptation, G and O are not tempting at all ($\widetilde{GT}_{-x} = 0$) for respectively 87.6% (99) and 68.1% (77) of respondents (relative to 73.5% and 63.7% with the original index). Furthermore, the distributions of $\widetilde{GT}_{-G}$ and $\widetilde{GT}_{-O}$ are now significantly different from each other on a Kolmogorov-Smirnov

[1] While other interpretations appear more likely in this case, temptation concerns cannot be entirely discarded, as G and R contain multiple items. For instance, consider a DM who is tempted by sugar in general (including fruits and pastries) and would like to maximize his protein intake by keeping items that contain meat (e.g., turkey burger). If maximizing protein intake is the main criterion for this DM, it could be that $R \succ G$. In addition, if he expects to be more likely to succumb to sugar if more sugary items are available, he might express the preference $G \succ GR$ and $R \succ GR$. Looking at Survey 2 responses, fruits appear particularly tempting; furthermore, some participants expressed the desire to keep protein in their diet.

test ($D = 0.195$, $p = 0.028$). On the other hand, the proportion of respondents for whom R is globally tempting ($\widetilde{GT}_{-R} = 3$) remains unchanged at 46% (52) under this alternative definition.

FIGURE 2: Temptation value of G, O and R foods (stricter definition of temptation)



Notes: For each category $x \in \{G, O, R\}$, "Index value" refers to the value of the Global Temptation Index $\widetilde{GT}_{-x} = \sum_{\widetilde{\mathcal{M}}_x} \mathbb{1}_{\{M\setminus\{x\} \succ M\}} \in \{0, 1, 2, 3\}$ where $\widetilde{\mathcal{M}}_x = \{M \in \mathcal{M} \mid x \in M \text{ and } \exists\, x' \in M\setminus\{x\} \text{ such that } \{x'\} \succ \{x\}\}$.

As a robustness check, I perform the regression analysis discussed in Section 4.2 of the main text on the subsample of respondents for whom GT_{-R} coincides with $\widetilde{GT}_{-R}$, that is, those for whom $G \succ R$ and $O \succ R$. The results are presented below in Sections **B.2** (goal setting), **B.3** (goal achievement), and **B.4** (other measures of attendance). The effect of the GT_{-R} index is very similar on this subsample for all outcomes except contract take-up, where it loses predictive power in all regressions. It is however important to note that, while the original definition might appear not restrictive enough, the alternative definition proposed here is quite strict. For instance, a DM with the ordering $GO \succ GOR \succ R \succ G$ might reveal that he would rather avoid R if he can get enough calories by combining the other two options, but would prefer R to having only G. In such case, the preference $GO \succ GOR$ could still reveal a temptation for R.

A.2.2. Strict Set Betweenness

In the paper, I study the structure of temptation by assessing the performance of the Set Betweenness axiom in predicting the choice data. To this end, I look at the 9 pairwise comparisons that involve two non-nested reimbursement options M and M' and consider five categories of preferences depending on how the individual ranks $M \cup M'$:

1- Standard (STD): $M \succ M'$ implies $M \sim M \cup M' \succ M'$

2- Flexibility-loving $(FLEX)$: $M \succ M'$ implies $M \cup M' \succ M \succ M'$

3- No Self-Control (NSC): $M \succ M'$ implies $M \succ M \cup M' \sim M'$

4- Self-Control (SC): $M \succ M'$ implies $M \succ M \cup M' \succ M'$

5- Global Commitment (GC): $M \succ M'$ implies $M \succ M' \succ M \cup M'$

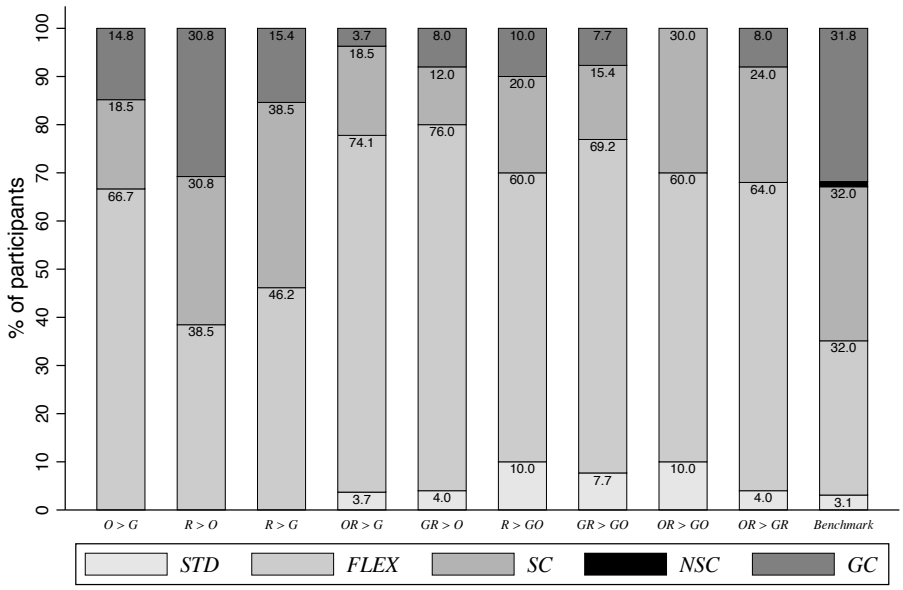
I show that category #4, referring to the strict form of Set Betweenness, is most prevalent in comparisons between one menu that does not contain R and one (less preferred) menu that contains R. On the other hand, preference for flexibility (category #2) is most prevalent in comparisons between two menus that either both contain R or both exclude R.

Figure **3** shows the percentage of each category for the least frequent binary choices (see main text for the most frequent choices). Table **6** presents a complete breakdown. For the least frequent choices, one should note that in none of the comparisons is an option that contains R dominated by an option that does not contain R. Consistent with the observations made above and in the main text, $FLEX$ is the modal category in all those comparisons, despite some variance in the actual frequency; by contrast, the fraction of SC preferences is below the random benchmark fraction in all but one of the 9 comparisons.

Table **5** presents a regression analysis of the propensity to exhibit preference SC across the 9 binary comparisons (M, M'), pooling all observations from Table **6** (last column); observations where participants were indifferent between the two options under comparison are excluded. All regressions study the effect of R belonging to the dominated option in the pair ($R \in M'$, $M \succ M'$) and R not belonging to the top option ($R \notin M$, $M \succ M'$). Columns 3 & 4 add an interaction term to consider both conditions jointly and Columns 2 & 4 control for respondent fixed effects. The proportion of SC preferences increases by 14-21 percentage points for comparisons in which R belongs to the dominated option and increases by 21-26 percentage points when R does not belong to the top option. The interaction term is positive and large: controlling for the two conditions separately, the joint restriction increases the proportion of SC preferences by more than 40 percentage points. Overall, SC is satisfied 57.6% of the time for comparisons in which both $R \notin M$ and $R \in M'$; the corresponding percentages for the other cases are 13.9% when $R \in M$ and $R \in M'$, 19.0% when $R \notin M$

and $R \notin M'$, and 19.9% when $R \in M$ and $R \in M'$. Controlling for these conditions explains about 12%-15% of the variance in the propensity to satisfy SC.

FIGURE 3: Distribution of menu preferences in bilaterial comparisons (least frequent cases)



Notes: Distribution of menu types for the least frequent preferences over two options (M, M'); see main text and Table **6** for the most frequent preferences. For instance, the number 66.7% on the first bar means that among those who ranked O strictly above G ($O \succ G$), 66.7% had the $FLEX$ ordering $GO \succ O \succ G$.

TABLE 5: Likelihood of satisfying SC

	(1)	(2)	(3)	(4)
$R \notin M$ *and* $M \succ M'$	0.263***	0.216***	-0.008	-0.122**
	(0.043)	(0.043)	(0.054)	(0.051)
$R \in M'$ *and* $M \succ M'$	0.207***	0.145***	-0.060	-0.189***
	(0.034)	(0.038)	(0.040)	(0.044)
$R \notin M$, $R \in M'$, *and* $M \succ M'$			0.445***	0.531***
			(0.061)	(0.069)
Respondent FE	No	Yes	No	Yes
Observations	979	979	979	979
R^2	0.120	0.426	0.149	0.467

Notes: Linear probability model where the dependent variable = 1 if the respondent satisfied SSB for a given comparison (SC category in Table **6**; $N = 979$ corresponds to the total numbers of observations in Table **6**, last column). Standard errors clustered at the subject level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

TABLE 6: Commitment vs. flexibility in pairwise comparisons

	% (N) of participants with preferences					
Binary choice	*STD*	*FLEX*	*SC*	*NSC*	*GC*	Total
Panel A: Most frequent binary choices						
$G \succ O$	6.4 (5)	66.7 (52)	19.2 (15)	0.0 (0)	7.7 (6)	100.0 (78)
$O \succ R$	0.0 (0)	23.4 (22)	70.2 (66)	2.1 (2)	4.3 (4)	100.0 (94)
$G \succ R$	1.0 (1)	23.5 (23)	69.4 (68)	1.0 (1)	5.1 (5)	100.0 (98)
$G \succ OR$	4.8 (4)	36.9 (31)	48.8 (41)	0.0 (0)	9.5 (8)	100.0 (84)
$O \succ GR$	3.4 (2)	36.2 (21)	48.3 (28)	0.0 (0)	12.1 (7)	100.0 (58)
$GO \succ R$	4.0 (4)	33.7 (34)	59.4 (60)	0.0 (0)	3.0 (3)	100.0 (101)
$GO \succ GR$	4.2 (4)	31.2 (30)	49.0 (47)	2.1 (2)	13.5 (13)	100.0 (96)
$GO \succ OR$	4.0 (4)	32.3 (32)	53.6 (53)	0.0 (0)	10.1 (10)	100.0 (99)
$GR \succ OR$	3.6 (3)	79.5 (66)	10.8 (9)	0.0 (0)	6.0 (5)	100.0 (83)
Total	3.4 (27)	39.3 (311)	48.9 (387)	0.6 (5)	7.7 (61)	100.0 (791)
Panel B: Least frequent binary choices						
$O \succ G$	0.0 (0)	66.7 (18)	18.5 (5)	0.0 (0)	14.8 (4)	100.0 (27)
$R \succ O$	0.0 (0)	38.5 (5)	30.8 (4)	0.0 (0)	30.8 (4)	100.0 (13)
$R \succ G$	0.0 (0)	46.2 (6)	38.5 (5)	0.0 (0)	15.4 (2)	100.0 (13)
$OR \succ G$	3.7 (1)	74.1 (20)	18.5 (5)	0.0 (0)	3.7 (1)	100.0 (27)
$GR \succ O$	4.0 (2)	76.0 (38)	12.0 (6)	0.0 (0)	8.0 (4)	100.0 (50)
$R \succ GO$	10.0 (1)	60.0 (6)	20.0 (2)	0.0 (0)	10.0 (1)	100.0 (10)
$GR \succ GO$	7.7 (1)	69.2 (9)	15.4 (2)	0.0 (0)	7.7 (1)	100.0 (13)
$OR \succ GO$	10.0 (1)	60.0 (6)	30.0 (3)	0.0 (0)	0.0 (0)	100.0 (10)
$OR \succ GR$	4.0 (1)	64.0 (16)	24.0 (6)	0.0 (0)	8.0 (2)	100.0 (25)
Total	3.7 (7)	66.0 (124)	20.2 (38)	0.0 (0)	10.1 (19)	100.0 (188)
Total sample	3.5	44.4	43.4	0.5	8.2	100.0
Benchmark	3.1	32.0	32.0	1.1	31.8	100.0

Notes: Distribution of preference patterns for the most and least frequent binary choices (Panel A and B). For instance, the first line of Panel A says that out of the 78 subjects who strictly preferred G to O, 19.2% (15 subjects) placed GO strictly in between. See main text for an explanation of the benchmark frequencies.

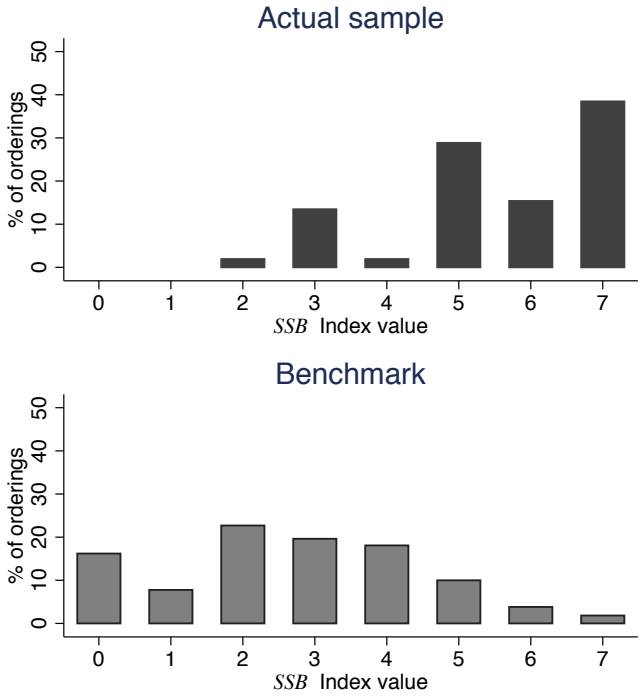
Proportion significantly higher than benchmark frequency (two-sided binomial test):
at 1% level; at 5% level; at 10% level

In the main text, I present an index SSB_{-x} for each food category $x \in \{G, O, R\}$ that computes the number of times a given respondent reveals a temptation for x of the Strict Set Betweenness form (type SC). This index takes into account all cases (M, M') in which x belongs to the dominated option M', but not to the top option M. For instance,

$$SSB_{-R} = \sum_{\mathcal{P}_R} \mathbb{1}_{\{M \succ M \cup M' \succ M'\}}$$

where $\mathcal{P}_R = \{(G, R), (O, R), (G, OR), (GO, R), (O, GR), (GO, GR), (GO, OR)\}$. In other words, $SSB_{-R} \in \{0, 1, ..., 7\}$ measures the number of times R is revealed to be a resistible temptation for a given respondent. I show that the index value for G and O is either 0 or 1 for most respondents. On the other hand, the index for R takes a value of 5 or higher for 40% of respondents and 83% of those for whom R is globally tempting ($GT_{-R} = 3$). Figure **4** presents the distribution of the SSB_{-R} index among the 52 respondents for whom R is globally tempting. About 40% exhibit the highest possible score, while the corresponding percentage for the random benchmark is less than 5%.

FIGURE 4: Distribution of the SSB_{-R} index for R globally tempting



Notes: "SSB Index value" refers to the value of the Strict Set Betweenness Index for option R, that is, $SSB_{-R} = \sum_{\mathcal{P}_R} \mathbb{1}_{\{M \succ M \cup M' \succ M'\}} \in \{0, 1, ..., 7\}$. The distribution of SSB_{-R} is shown for the subsample of 52 respondents for whom R is globally tempting ($GT_{-R} = 3$).

A.2.3. Typology of menu preferences

In the main text, I discuss the consistency of individual preference orderings with the following three axioms:

Positive Set Betweenness (PSB): $M \succeq M'$ implies $M \succeq M \cup M'$

Weak Set Betweenness (WSB): If $\{x\} \succeq \{y\}$ for all $x \in M$, $y \in M'$ then $M \succeq M \cup M' \succeq M'$

Monotonicity (MON): $M \subseteq M'$ implies $M' \succeq M$

The first two axioms proposed by Dekel et al. (2009) allow respectively for cumulative temptations (CT) and stochastic temptations (ST), while the last axiom refers to preference for flexibility as formulated in Kreps (1979). To construct the classification presented in Figure 4 of the main text, I count the number of times the preference ordering of a given participant violates each of the three axioms. To avoid double counting, I take the following two steps. First, I count only one restriction for each indifference. For instance, if $G \sim O$ and $GO \succ G$, then $GO \succ O$ automatically follows by transitivity; as a result, I count $(G \succeq O) \wedge \neg(G \succeq GO)$ and $(O \succeq G) \wedge \neg(O \succeq GO)$ as a single violation of PSB. Similarly, if $G \sim O \sim R$, then WSB implies both $G \succeq GOR \succeq OR$ and $OR \succeq GOR \succeq G$, that is, $G \sim GOR \sim OR$; in this case, $(G \sim O \sim R) \wedge \neg(G \sim GOR \sim OR)$ is counted as one violation of WSB. Second, to count violations of WSB, I restrict attention to the set of binary comparisons of any two non-nested menus (9 in total). To see why, assume $G \succ O \succ R$ and consider the binary comparison G vs. OR; by WSB, it follows that $G \succeq GOR \succeq OR$. Similarly, comparing the two nested options G and GOR, one obtains $G \succeq GOR \sim GOR$. Thus, if WSB is violated at (G, GOR), it is a fortiori violated at (G, OR). I therefore focus on the tightest restrictions and discard the comparisons involving two nested menus.

In total, the falsification tests involve 9 binary comparisons for PSB and WSB (the same as for SB), and 12 comparisons for MON.[2] In the appendix of the main text, I present the distribution of violations for each axiom, both for the actual sample and for the random benchmark. To see how permissive PSB and WSB are relative to SB, I also report the distribution of violations for SB. In order to get an idea of where the violations come from, Panel (a) of Figures **5**, **6** and **7** present a more complete breakdown, with the distribution of violations shown separately for (i) comparisons involving only singleton options; (ii) comparisons involving at least one doubleton. Note that WSB restricted to singleton options is simply SB. To understand how different these distributions are from the random benchmark, I perform 1,000 Kolmogorov-Smirnov tests of equality of distribution that compare the actual sample to each of the 1,000 permutations that make the random benchmark.

[2]The comparisons for PSB and WSB are $\{(G,O),(G,R),(O,R),(GO,R),(GR,O),(OR,G),(GO,GR),(GO,OR),(GR,OR)\}$. The comparisons for MON are $\{(G,GO),(G,GR),(O,GO),(O,OR),(R,GR),(R,OR),(G,GOR),(O,GOR),(R,GOR)\}$, (GO,GOR), (GR,GOR), $(OR,GOR)\}$.

The distribution of p-values is shown in Panel (b) of each figure. Except for PSB, the Kolmogorov-Smirnov tests reject the null of equality of distributions at the 5% level in more than 90% of the cases.

Based on these falsification tests, I classify participants into types (ST, CT, $FLEX$), assuming they violate the corresponding axiom (WSB, PSB, MON) at most once. The distribution of preferences is presented in Figure 4 of the main text. Table **7** shows how the typology changes if one allows for a stricter requirement (no violation) or a weaker requirement (two violations); the one violation case is also reported. For subjects who are consistent with both ST and CT (i.e., violate WSB and PSB no more than once), I report the proportion who also satisfy SB. Without allowing for any violation, only about a third of respondents can be classified; this proportion increases to 83% when allowing for two violations. The proportion of individuals with temptation-driven preferences ranges from 22% with no violation, to 51% with one violation, and 53% with two violations; the corresponding benchmark proportions are 14%, 36% and 57%. The proportion who satisfy SB is about 10 times higher in the actual sample than in the benchmark sample; however, it remains low if one allows for at most one violation.

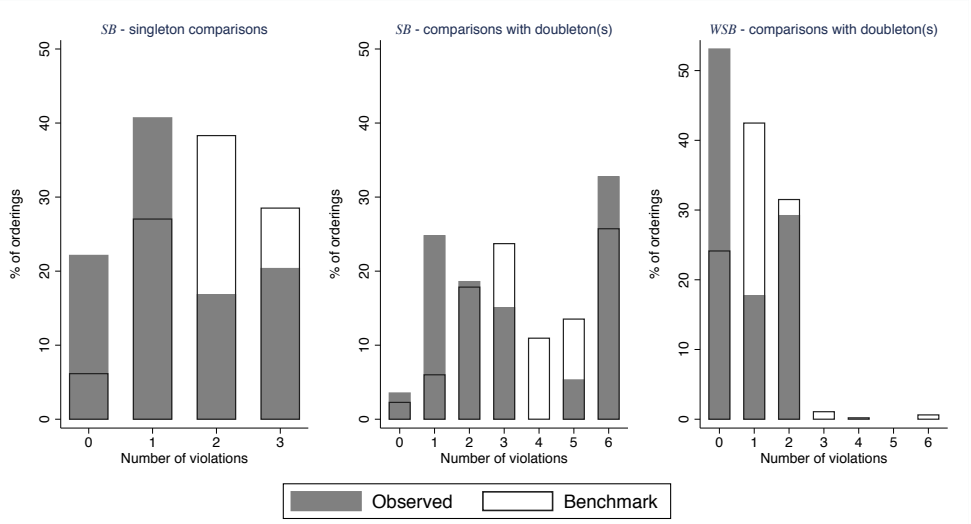
TABLE 7: Typology of participants

	% (N) of participants of type					
	ST	CT	ST or CT		$FLEX$	Other
			SB	$\neg SB$		
Panel A: No violations						
actual	15.9 (18)	3.5 (4)	1.8 (2)	0.9 (1)	11.5 (13)	66.4 (75)
benchmark	0.8	12.3	0.2	1.1	1.2	84.4
Panel B: At most one violation						
actual	24.8 (28)	8.9 (10)	10.6 (12)	7.1 (8)	20.3 (23)	28.3 (32)
benchmark	3.0	25.7	1.0	6.2	3.9	60.2
Panel C: At most two violations						
actual	10.6 (12)	4.4 (5)	30.1 (34)	8.0 (9)	30.1 (34)	16.8 (19)
benchmark	6.8	29.2	3.7	17.2	7.6	35.5

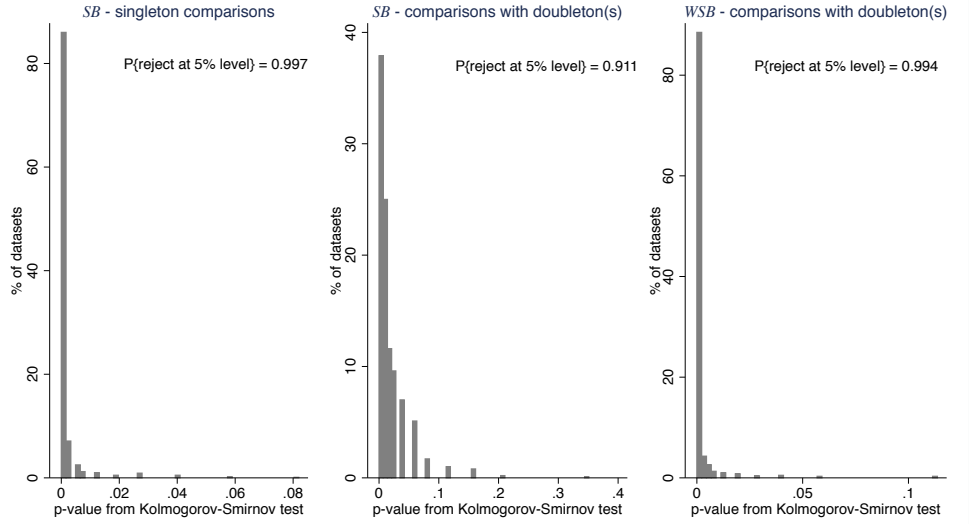
Notes: Participant classified as a ST type if s/he violates WSB at most n times where $n \in \{0, 1, 2\}$, and is consistent with at most one type; types CT and $FLEX$ for the PSB and MON axioms are similarly defined. Category "ST or CT" refers to participants who are consistent with both WSB and PSB in all (but possibly one) cases, and among them, $(\neg)$ SB refers to those (not) consistent with SB. Finally, "Other" refers to subjects who either could not be classified (more than one violation of WSB, PSB and MON), or were consistent with both $FLEX$ and some other type.

FIGURE 5: Violations of SB and WSB

(A) Distribution of violations



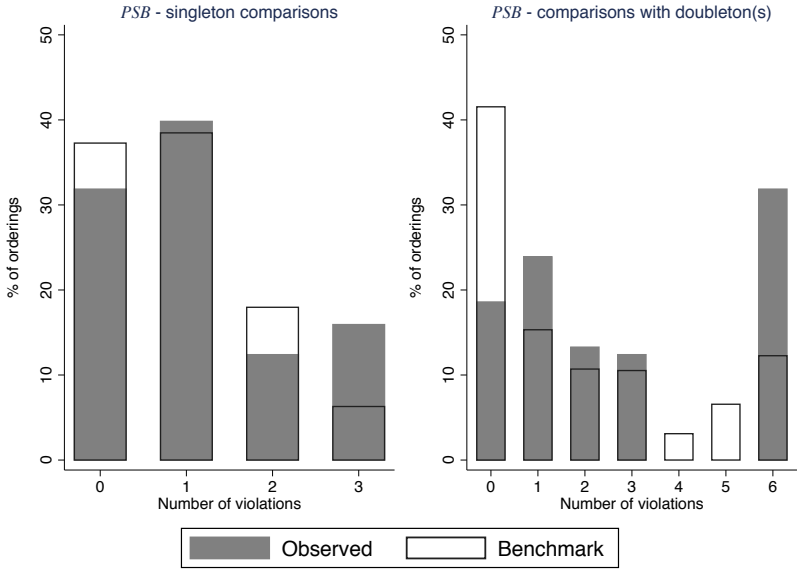
(B) p-values from tests of equality of distributions



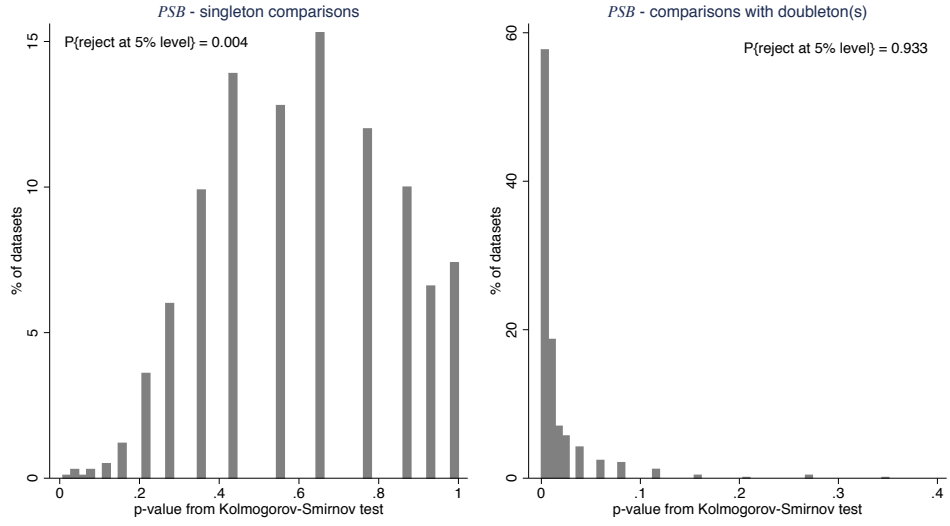
Notes: Panel (a) shows the distribution of violations of SB and WSB when comparing two singleton options (left) and singleton/doubleton vs. doubleton options (center and right). Panel (b) shows the corresponding distributions of p-values from Kolmogorov-Smirnov tests comparing violations in the actual sample to each random permutation in the benchmark sample ($N = 1,000$). P(reject at 5% level) refers to the proportion of times (out of 1,000) the p-value was below 0.05.

FIGURE 6: Violations of PSB

(A) Distribution of violations



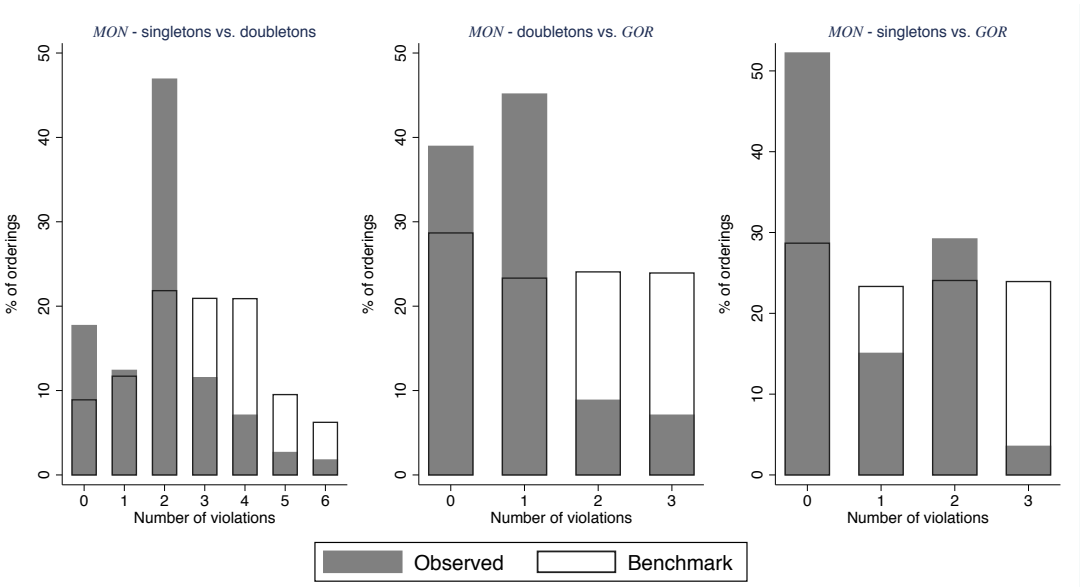
(B) p-values from tests of equality of distributions



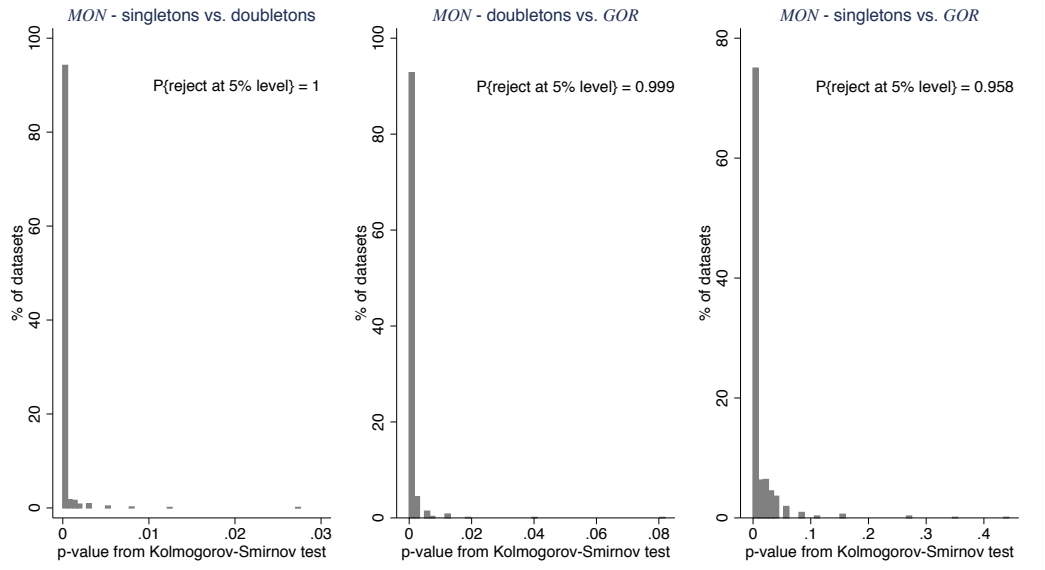
Notes: Panel (a) shows the distribution of violations of PSB when comparing two singleton options (left) and singleton/doubleton vs. doubleton options (right). Panel (b) shows the corresponding distributions of p-values from Kolmogorov-Smirnov tests comparing violations in the actual sample to each random permutation in the benchmark sample ($N = 1,000$). P(reject at 5% level) refers to the proportion of times (out of 1,000) the p-value was below 0.05.

FIGURE 7: Violations of *MON*

(A) Distribution of violations



(B) p-values from tests of equality of distributions



Notes: Panel (a) shows the distribution of violations of *MON* when comparing a singleton and a doubleton that contains it (left), a doubleton and *GOR* (center) and a singleton and *GOR* (right). Panel (b) shows the corresponding distributions of p-values from Kolmogorov-Smirnov tests comparing violations in the actual sample to each random permutation in the benchmark sample ($N = 1,000$). P(reject at 5% level) refers to the proportion of times (out of 1,000) the p-value was below 0.05.

Finally, bringing all findings together, Table 8 presents a correlation table of the various menu choice measures of temptation presented in this section. The correlation between G *top* (*GO top*) and the GT_{-R} and SSB_{-R} indices is around 0.3 (0.5). Unsurprisingly, the GT_{-j} and SSB_{-j} values are highly correlated, with a particularly strong correlation for option R. The temptation measures based on the entire ordering ($TEMPT$ types in Panel C) are also strongly correlated with the two indices. Furthermore, the correlation becomes stronger for all food categories if one allows for more violations in the definition of the $TEMPT$ measures. For instance, $TEMPT1$ and $TEMPT2$ are positively correlated with SSB_{-G}. As such, the $TEMPT$ measures likely capture broader concerns than temptation.

TABLE 8: Correlation between menu preference variables

Panel A: Correlation between top choice and temptation measures

	GT_{-G}	GT_{-O}	GT_{-R}	SSB_{-G}	SSB_{-O}	SSB_{-R}
G top	-0.003	0.60***	0.30***	0.07	0.56***	0.28***
GO top	-0.03	-0.05	0.45***	-0.12	-0.18*	0.45***
GOR top	-0.25***	-0.39***	-0.66***	-0.14	-0.22**	-0.60***

Panel B: Correlation between temptation measures

	GT_{-G}	GT_{-O}	GT_{-R}	SSB_{-G}	SSB_{-O}	SSB_{-R}
GT_{-G}	–					
GT_{-O}	0.48***	–				
GT_{-R}	-0.08	0.16*	–			
SSB_{-G}	0.62***	-0.003	-0.07	–		
SSB_{-O}	-0.06	0.58***	0.27***	0.09	–	
SSB_{-R}	-0.22**	-0.05	0.87***	0.10	0.37***	–

Panel C: Correlation between type and temptation measures

	GT_{-G}	GT_{-O}	GT_{-R}	SSB_{-G}	SSB_{-O}	SSB_{-R}
TEMPT0	0.03	0.23**	0.33***	0.02	0.34***	0.33***
TEMPT1	0.10	0.14	0.73***	0.21**	0.37***	0.74***
TEMPT2	0.08	0.29***	0.77***	0.20**	0.46***	0.80***

Notes: Coefficients in Panel A/C (B) are point biserial (Pearson) correlations. The indicator variables *TEMPT0*, *TEMPT1*, and *TEMPT2* are equal to 1 if the participant can be classified as having temptation-driven preferences (ST and/or CT) based on 0,1, and 2 violations of the corresponding axioms (WSB and/or PSB). See main text for a description of the other variables. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

A.3 Implementation procedure and actual assignments

Respondents were asked to assign a rank number between 1 and 7 to each of the 7 reimbursement options and could express indifferences by assigning the same rank to several options. A random implementation procedure was used in order to incentivize respondents to report their entire ordering truthfully. More specifically, participants were told that their reimbursement option would be determined through a lottery assigning higher odds to higher ranked options. The exact odds were (0.35, 0.3, 0.2, 0.1, 0.03, 0.02, 0) where $0.35 = \mathbb{P}\{\text{rank } 1\}$ and $0 = \mathbb{P}\{\text{rank } 7\}$. Options assigned the same rank received in expectation the same chances of being selected. For instance, if three options were assigned rank 1, then each option was equally likely to be drawn with probability 0.35, 0.3 or 0.2. Participants could verify the exact odds by clicking on a button "Learn more about the selection procedure"; 27 of them (24%) clicked on the button. Table **9** shows the distribution of assigned options. For instance, 20.4% of participants were assigned option G and 35.4% received (one of) their rank 1 option(s); the observed proportions are almost identical to the ex ante odds of assignment.[3]

TABLE 9: Distribution of assigned options

Option assigned	% (N)	Selected rank	% (N)
G	20.4 (23)	1	35.4 (40)
GO	23.9 (27)	2	31.9 (36)
GOR	24.5 (28)	3	20.3 (23)
GR	11.5 (13)	4	8.0 (9)
O	8.9 (10)	5	2.6 (3)
OR	7.1 (8)	6	1.8 (2)
R	3.5 (4)	7	0.0 (0)
Total	100 (113)	Total	100 (113)

[3]The randomization was performed manually. A number between 1 and 100 was drawn with replacement for each participant. If the random number (called `randomdraw` in the database) was between 1 and 35, the participant was assigned his top option; if the random number was between 36 and 65, he was assigned his second best option, etc (variable `selectedmenu` in the database). Due to an error in the allocation, 6 participants (ID 8, 22, 47, 55, 114 and 117) were assigned an option other than the one dictated by the random draw (discrepancy between `randomdraw` and `selectedmenu`); none of them were assigned their last option however.

A.4 Frequency, Health and Temptation ratings by food item

In Section 4.1 of the main text, I discuss data on participants' perception and relative consumption of the three lunch categories, G, O and R. Below I present more disaggregated data, looking separately at each food item in a given category. To study their subjective perceptions of the various food items, participants were asked to rate each item on a scale from 1 to 7 in terms of their health value and temptation value. The following two questions were asked during Survey 1, right after participants submitted their ranking of the various reimbursement options:

1. *"On a scale from 1 to 7, how healthy do you consider each of the following options? (1 = not healthy at all; 7 = extremely healthy)"*

2. *"On a scale from 1 to 7, how tempting do you consider each of the following options? (1 = not tempting at all; 7 = extremely tempting)"*

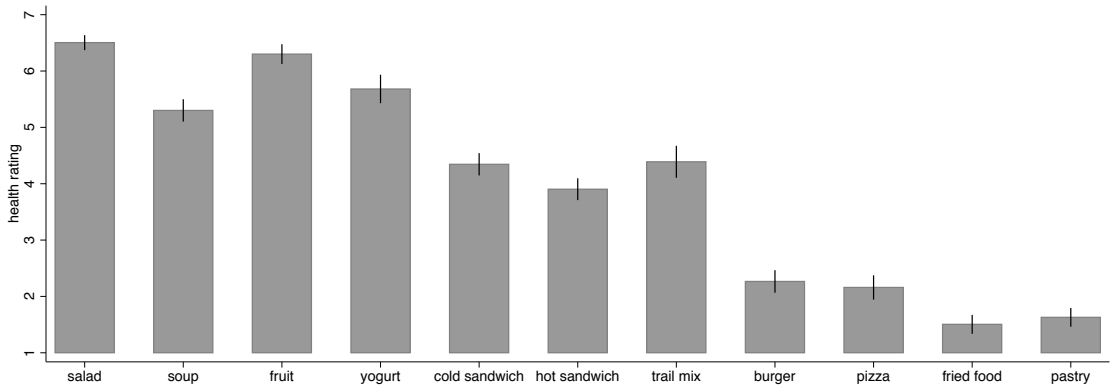
These questions were asked for each of 11 food items: salad, soup, yogurt and fruit (category G); cold sandwich, hot sandwich and cereal bar/trail mix (category O); burger, pizza, fried food and pastry (category R). The mean rating for each item is presented in Panel A (health rating) and Panel B (temptation rating) of Figure **8**. As can be seen, participants perceive a clear difference in health value between items belonging to different lunch categories: the mean health rating is above 5 for all items in G, between 3.9 and 4.4 for items in O, and below 2.5 for all items in R. On the other hand, perceived differences in temptation value appear relatively small, with a mean temptation rating between 3.5 and 5.1 for each of the 11 food items. Although the two most (least) tempting items belong to R (G and O), soup and fruit appear on average as tempting as burger and fried food. As a result, Question 2 might measure taste rather than temptation per se. To explore this hypothesis, I asked a follow-up question to the 87 respondents of Survey 2, which defined "tempting":

3. *"You may consider certain foods to be **tempting** in the sense that (1) you find them **hard to resist**; (2) but you know that **you should avoid them**. On a scale from 1 to 7, please indicate how tempting you find each of the following food items (1 = not tempting at all, 7 = extremely tempting)"*

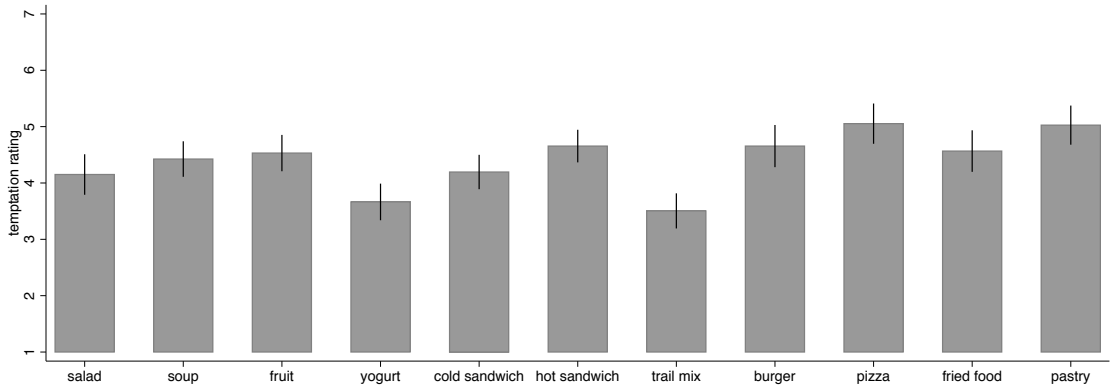
Figure **8** Panel C presents the mean response for each food item, while Table **10** compares the mean ratings in Questions 2 & 3 for the subsample of Survey 2 respondents. The mean temptation ratings are nearly identical in the two questions for all items in the R category. On the other hand, the mean temptation value of food items in G and O appears significantly lower if one uses Question 3, suggesting that answers to Question 2 are most likely an expression of respondents' tastes.

FIGURE 8: Perceived health and Temptation value of each food item

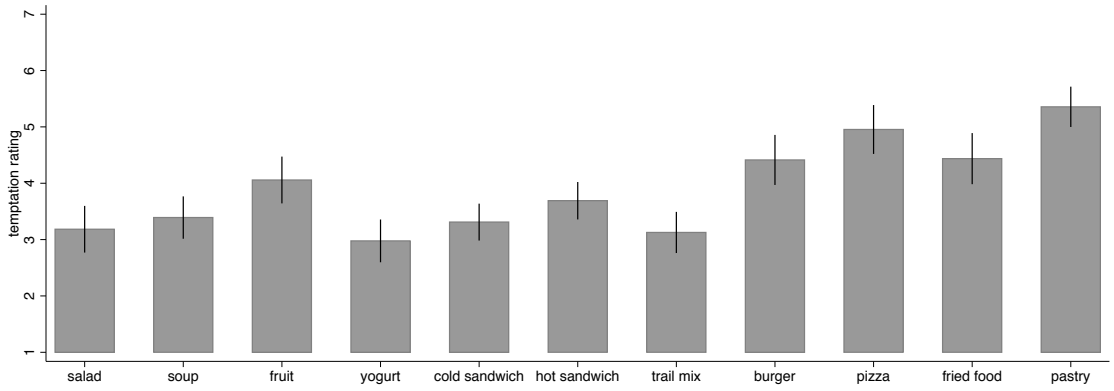
(A) Health rating



(B) Temptation rating (Survey 1)



(C) Temptation rating (Survey 2)



Notes: Mean response over all respondents ($N = 113$ for Panels A & B and $N = 87$ for C).

TABLE 10: Comparison of mean temptation ratings (Survey 1 vs. Survey 2 questions)

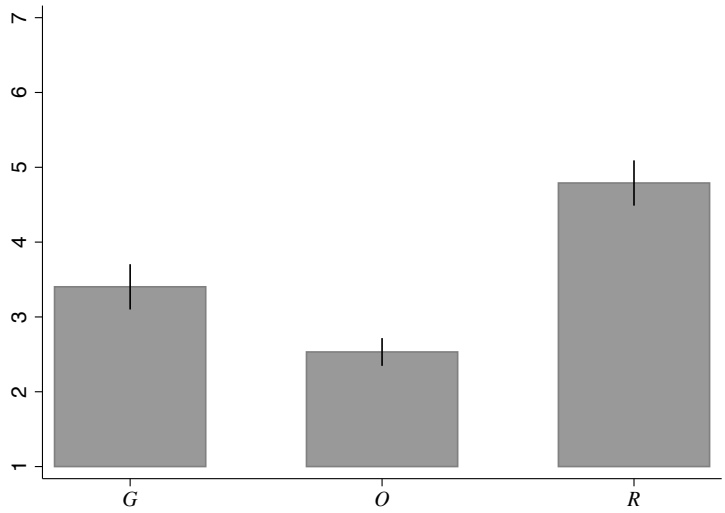
food item	Survey 1 measure	Survey 2 measure	t-stat	p-value
salad	4.11	3.18	4.78	< 0.001
	(0.21)	(0.21)		
soup	4.38	3.39	4.45	< 0.001
	(0.18)	(0.19)		
fruit	4.61	4.06	2.85	0.005
	(0.18)	(0.21)		
yogurt	3.64	2.98	3.99	< 0.001
	(0.19)	(0.19)		
cold sandwich	4.16	3.31	5.16	< 0.001
	(0.17)	(0.16)		
hot sandwich	4.57	3.69	5.48	< 0.001
	(0.17)	(0.17)		
cereal/trail mix	3.45	3.13	1.41	0.162
	(0.18)	(0.18)		
burger	4.54	4.41	0.71	0.482
	(0.22)	(0.22)		
pizza	4.99	4.95	0.19	0.849
	(0.20)	(0.22)		
fried food	4.46	4.44	0.13	0.898
	(0.21)	(0.23)		
pastry	5.08	5.36	-1.31	0.195
	(0.19)	(0.18)		
observations	87	87		

Notes: Survey 1 (2) measure refers to Question 2 (3) in the text.

Using the Survey 2 temptation ratings to define a new Temptation Score for each lunch category, Figure **9** shows clearer differences in temptation value between R and the other two categories, and G now appears significantly more tempting than O. Some differences also appear across menu preferences for a given category, as shown in Table **12**. In particular, red (green) foods appear less tempting for menu preference category G *top* ($GT_{-R} = 0$).

Table **11** shows that regardless of the measure, foods in the R category are perceived on average as significantly less healthy and more tempting than both G and O. While O is also perceived as less healthy than G, it is not perceived as more tempting. Confirming the above observations, belonging to a specific lunch category explains nearly 70% of the variance in respondents' perceptions of the health value of a given food item. On the other hand, lunch category has much less power to predict how respondents perceive the various food items on the temptation dimension, with an R^2 equal to 0.03 (0.12) for the Survey 1 (2) measure.

FIGURE 9: Mean Temptation Score (Survey 2 measure)



Notes: The error bars are 95% confidence intervals obtained from linear regressions of the Temptation score on dummies for the lunch category; standard errors clustered at the subject level ($N = 87$).

TABLE 11: Differences in health and temptation value across food categories G, O and R

	Health rating	Temptation rating 1	Temptation rating 1	Temptation rating 2
	(1)	(2)	(3)	(4)
Belongs to category O	-1.735***	-0.074	-0.125	-0.027
	(0.085)	(0.139)	(0.164)	(0.172)
Belongs to category R	-4.058***	0.633***	0.580**	1.388***
	(0.101)	(0.197)	(0.220)	(0.242)
constant	5.947***	4.192***	4.187***	3.402***
	(0.062)	(0.124)	(0.140)	(0.152)
F-stat: $O = R$	623.65***	26.17***	21.68***	69.15***
Sample	All	All	Survey 2	Survey 2
Observations	1,243	1,243	957	957
R^2	0.699	0.030	0.028	0.116

Notes: Linear regressions where the dependent variable is the health (temptation) rating assigned to a given food item. In columns 2 & 3 (4), the temptation rating refers to the Survey 1 (2) question; in column 3, the regression is estimated on the subset of Survey 2 respondents. The indicator variable *Belongs to category O* (*R*) is equal to 1 if the given food item belongs to category O (R); G is the reference category. Standard errors clustered at the subject level; ** $p < 0.05$ and *** $p < 0.01$.

TABLE 12: Temptation ratings of G, O, R by menu preference

	Survey 1 measure			Survey 2 measure		
Lunch category	G	O	R	G	O	R
	Panel A: By top choice					
G top	4.75	4.07	4.13	4.18	2.55	3.85
	(0.41)	(0.34)	(0.38)	(0.44)	(0.27)	(0.44)
GO top	4.19	3.97	4.53	3.51	2.59	4.74
	(0.26)	(0.21)	(0.23)	(0.28)	(0.17)	(0.27)
GOR top	4.06	4.31	5.07	3.14	2.67	5.13
	(0.23)	(0.19)	(0.21)	(0.24)	(0.15)	(0.24)
Other	4.10	3.72	4.92	3.29	2.19	4.76
	(0.31)	(0.26)	(0.28)	(0.33)	(0.20)	(0.33)
F-stat	0.75	1.22	2.13	1.47	1.22	2.21*
	Panel B: By value of the GT_{-R} Index					
$GT_{-R} = 0$	3.86	4.23	5.19	2.81	2.3	5.01
	(0.29)	(0.24)	(0.27)	(0.31)	(0.19)	(0.32)
$GT_{-R} = 1$	4.14	3.94	4.78	3.45	2.77	4.84
	(0.33)	(0.27)	(0.30)	(0.34)	(0.22)	(0.36)
$GT_{-R} = 2$	4.06	4.49	4.72	3.18	2.72	4.69
	(0.32)	(0.26)	(0.29)	(0.33)	(0.21)	(0.35)
$GT_{-R} = 3$	4.46	3.80	4.54	3.84	2.46	4.68
	(0.22)	(0.19)	(0.21)	(0.24)	(0.15)	(0.25)
F-stat	0.98	1.78	1.23	2.51*	1.22	0.26
Observations	87	87	87	87	87	87

Notes: Results from linear regressions of the temptation scores (Survey 1 & 2 measures) of lunch category $m \in \{G, O, R\}$ on dummies for top choice (Panel A) and level of GT_{-R} (Panel B); standard errors are in parentheses. Reported F-statistic corresponds to a test of the null hypothesis that all dummy coefficients are equal. * $p < 0.10$

Participants were also asked about their consumption habits, aspirations and desires:

4. *"Since the beginning of the year, how often did you have each of the following options for lunch? (0 = never; 50 = quite often; 100 = all the time)"*

5. *"Ideally, indicate how frequently you think you* ***should*** *consume each of the following food items: (0 = never; 50 = quite frequently; 100 = all the time)"*

6. *"Suppose you could eat anything you want without gaining a single pound and without any consequences for your health. How frequently would you eat each of the following food items? (0 = never; 50 = quite frequently; 100 = all the time)"*

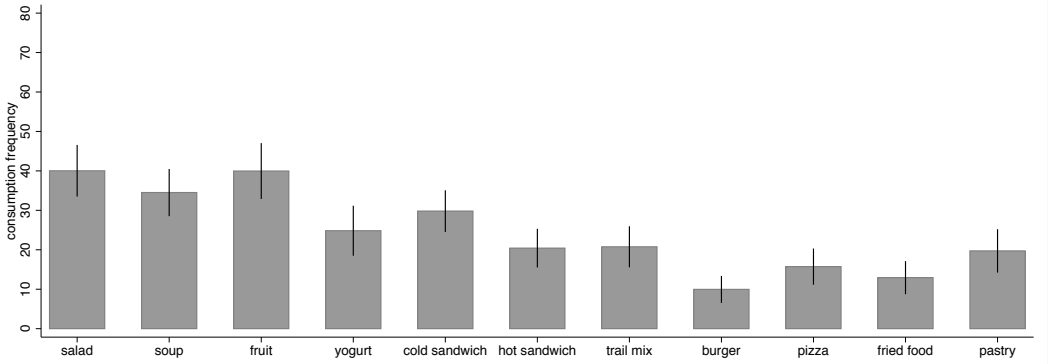
Figure **10** shows the mean response for each food item in each question. In the main text, I present two additional measures of temptation based on answers to these three questions: (1) the *Actual* − *Ideal* gap, which compares answers to Questions 4 & 5; (2) the *Unrestricted* − *Ideal* gap, which contrasts answers to Questions 5 & 6. In models of costly self-control à la GP, one can think of ideal consumption (Question 5) as maximizing u, unrestricted consumption (Question 6) as maximizing v, and actual consumption (Question 4) as maximizing $u+v$. Letting $s_w(j)$ be the share in total consumption of food category $j \in \{G, O, R\}$ that maximizes utility $w \in \{u, v, u+v\}$, I then define the *Actual* − *Ideal* gap for j as $\Delta_{A-I}(j) := s_{u+v}(j) - s_u(j)$ and the *Unrestricted* − *Ideal* gap as $\Delta_{U-I}(j) := s_v(j) - s_u(j)$. To construct those measures, I first compute for each of the three questions the consumption share of food category $j \in \{G, O, R\}$

$$s(j) = \frac{f(j)}{f(G) + f(O) + f(R)}$$

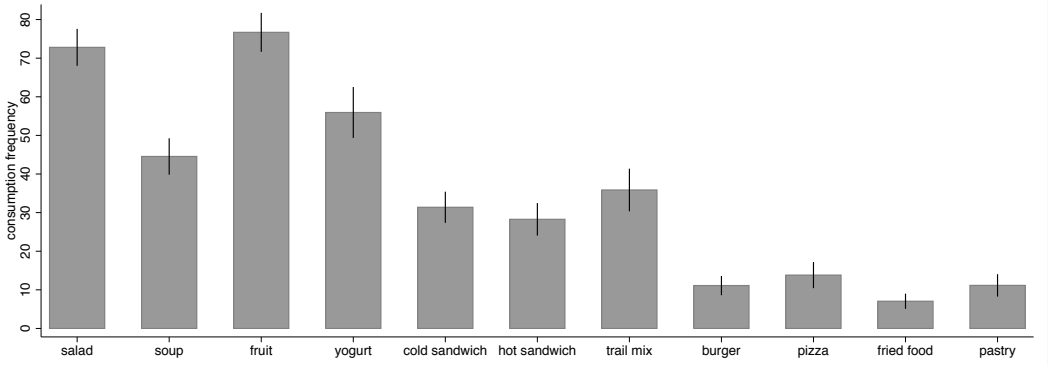
where $f(j)$ is the average of a respondent's answers to that question for all foods belonging to category j. Table **13** presents the actual, ideal and unrestricted consumption shares for G, O and R (Questions 4, 5 and 6) for each menu preference category. The ideal and unrestricted consumption shares are very similar across menu preferences; on the other hand, there are differences in actual consumption: participants with a stronger preference for removing R from the coverage report consuming green foods at a higher frequency. Differences in actual consumption also appear for foods in category O; however, actual consumption of R foods is fairly stable across menu preference categories.

FIGURE 10: Actual, ideal and unrestricted consumption frequencies by food item

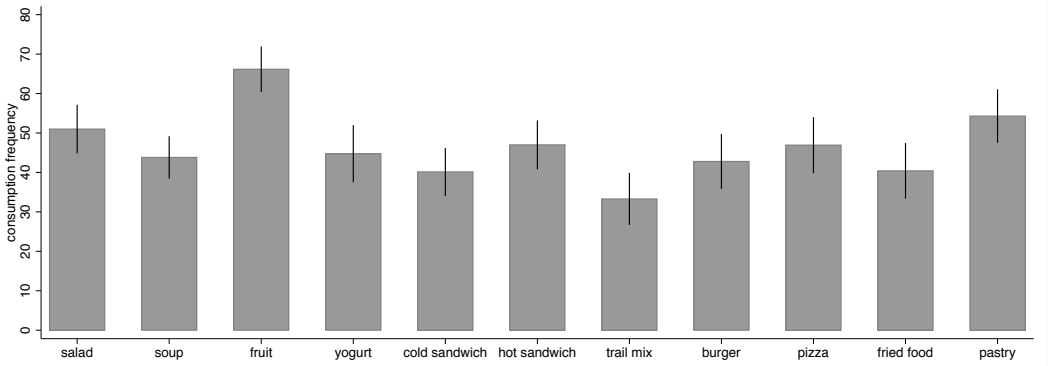
(A) Actual consumption



(B) Ideal consumption



(C) Unrestricted consumption



Notes: Mean response to health, temptation and consumption frequency questions presented in the text. For the actual consumption frequency question (asked in Survey 1), the mean is taken over the subsample of participants who responded to Survey 2 (therefore, $N = 87$ in all 3 figures).

TABLE 13: Actual, ideal and unrestricted consumption of G, O and R by menu preference

	Actual Consumption			Ideal Consumption			Unrestricted Consumption		
Lunch category	G	O	R	G	O	R	G	O	R
				Panel A: By top choice					
G top	0.65	0.20	0.15	0.66	0.26	0.08	0.45	0.27	0.27
	(0.07)	(0.06)	(0.05)	(0.04)	(0.03)	(0.03)	(0.05)	(0.03)	(0.05)
GO top	0.51	0.30	0.19	0.61	0.28	0.11	0.40	0.29	0.31
	(0.04)	(0.04)	(0.03)	(0.03)	(0.02)	(0.02)	(0.03)	(0.02)	(0.03)
GOR top	0.49	0.32	0.20	0.62	0.30	0.09	0.38	0.28	0.34
	(0.04)	(0.03)	(0.03)	(0.02)	(0.02)	(0.01)	(0.03)	(0.02)	(0.03)
Other	0.38	0.45	0.17	0.60	0.29	0.11	0.39	0.26	0.35
	(0.05)	(0.04)	(0.04)	(0.03)	(0.02)	(0.02)	(0.04)	(0.03)	(0.04)
F-stat	3.64**	4.45***	0.32	0.53	0.29	1.02	0.59	0.29	0.77
				Panel B: By value of the GT_{-R} Index					
$GT_{-R} = 0$	0.36	0.44	0.21	0.58	0.32	0.10	0.36	0.27	0.37
	(0.05)	(0.04)	(0.03)	(0.03)	(0.02)	(0.02)	(0.04)	(0.02)	(0.03)
$GT_{-R} = 1$	0.46	0.33	0.21	0.62	0.29	0.09	0.37	0.27	0.35
	(0.05)	(0.05)	(0.04)	(0.03)	(0.03)	(0.02)	(0.04)	(0.03)	(0.04)
$GT_{-R} = 2$	0.57	0.31	0.12	0.62	0.28	0.10	0.39	0.31	0.30
	(0.05)	(0.04)	(0.04)	(0.03)	(0.03)	(0.02)	(0.04)	(0.03)	(0.04)
$GT_{-R} = 3$	0.54	0.27	0.18	0.63	0.27	0.10	0.43	0.27	0.30
	(0.04)	(0.03)	(0.03)	(0.02)	(0.02)	(0.01)	(0.03)	(0.02)	(0.03)
F-stat	4.51***	3.39**	1.24	0.63	0.92	0.06	1.11	0.47	1.31
Observations	87	87	87	87	87	87	87	87	87

Notes: Results from linear regressions of actual, ideal and unrestricted consumption share of lunch category $m \in \{G, O, R\}$ on dummies for top choice (Panel A) and level of GT_{-R} (Panel B); standard errors are in parentheses. Reported F-statistic corresponds to a test of the null hypothesis that all dummy coefficients are equal. ** $p < 0.05$ and *** $p < 0.01$

A.5 Motivations for choosing commitment versus flexibility

In Survey 2, participants were asked to explain their ranking of the various reimbursement options and in particular, why they chose or did not choose to assign rank 1 to the most flexible option *GOR*. Here are the motivations given by 72 of the 87 survey respondents.[4] Comments are ordered by ID number and divided into two categories according to whether the respondent ranked *GOR* first or not.

Among the respondents who did not assign rank 1 to *GOR* (45/48 usable comments), 47% explicitly mentioned that they wanted to be motivated to eat healthier/discouraged to eat unhealthy. Another 42% explained that they ranked reimbursement options based on their current consumption habits and did not see the interest of being reimbursed for food categories they did not consume.[5] The remaining 11% mentioned a desire not to be reimbursed for unhealthy foods, without being more specific about their regular food habits and/or intentions to improve their diet.

Among the respondents who did assign rank 1 to *GOR* (27/39 usable comments), the main motivations were maximizing reimbursement, having a diversified and/or realistic plan, or being able to deal with the uncertainty of food and work schedules (about 67%). Some respondents simply mentioned a desire to retain flexibility, keep options open and/or avoid restrictions (about 26%). The various motives described by participants are summarized in **Table 14** at the end of this section; an explanation of the coding scheme is also provided.

A.5.1. Comments of participants who did not assign rank 1 to *GOR*

1. "I always eat salad for lunch. I was indifferent between all the options that included green." (ID 3 - ranked *G* first)

2. "I had just come off a cleanse in which I was limited to fruits and vegetables and I did not want to shock my system by suddenly introducing a lot of breads, juices and trail mix into my diet. (Plus, I liked the diet I'd had on my cleanse and did not want to ruin it by going to sandwiches and juices.)" (ID 4 - ranked *G* first)

3. "My purchases fall almost exclusively in the G category except perhaps the trail mix part as I would do nut mixes. As long as G was in it, I would have been fine. [...] I will say that I am not generally tempted to buy from the red category and bring my lunch." (ID 5 - ranked *G* and *GO* first)

[4] The 15 remaining respondents either did not respond to the question, mentioned that they did not remember the ranking procedure/were not sure of their choice, or answered something unrelated to the question. The entire set of comments is available from the author upon request.

[5] Except for two respondents, all specified that they essentially consume green foods and/or rarely eat red foods.

4. "More likely to get fast food for lunch, which was the only meal that could be reimbursed" (ID 6 - ranked R first)

5. "I love eating meat, chicken, and fried food but I tried to make sandwiches, cereal and fruit bars, trail mix, and juices my top ranking to encourage me to eat that." (ID 11 - ranked OR first, O second)

6. "Because, if I was going to put the effort into this challenge, I have to set standards that I must follow in order to reach my goal." (ID 16 - ranked GO first)

7. "Most of the food covered under the Green category is food that I usually make myself and take to work for lunch, so it didn't make sense for me to choose it. I wanted to pick something that made more sense with my habits." (ID 18 - ranked GO first, GOR second)

8. "If I decided to do the program I would have less incentive to eat the stuff I knew would get me in trouble. No money back meant less likely to eat it." (ID 19 - ranked GO first)

9. "I anticipated that if I had R as an option I would be less motivated to eat in a healthy manner." (ID 20 - ranked GO first)

10. "Because I didn't want to give myself the option of choosing something unhealthy." (ID 22 - ranked GO first)

11. "it would dissuade me from unhealthier choices." (ID 23 - ranked G first)

12. "I felt that by assigning only G as rank 1 that maybe I would have more incentive to eat healthier (that was the healthiest option)." (ID 24 - ranked O first, G second, GO third)

13. "It is more aligned with the way I currently eat." (ID 27 - ranked GO first)

14. "I tend to not eat burgers and pizza, so anything with R was going to be ranked lower." (ID 30 - ranked GO first)

15. "I do not eat fast foods/ soda often so I prefer foods in the Green category." (ID 34 - ranked G,O and GO first)

16. "I do not purchase lunch. I also typically do not eat any of the red categories for lunch." (ID 35 - ranked G first)

17. "I wanted to discourage myself as much as possible from eating R foods" (ID 38 - ranked GO first)

18. "I didn't want to encourage eating EVERYTHING because I knew that would be bad." (ID 40 - ranked *GO* first)

19. "I wanted to force myself to eat healthy" (ID 44 - ranked *GO* first)

20. "I very rarely eat R category foods." (ID 46 - ranked *G* first)

21. "I didn't want to be tempted by the R category." (ID 52 - ranked *GO* first)

22. "I wanted to be motivated to eat better." (ID 53 -ranked *GO* first)

23. "Because this allows me not to indulge in a tendency of R." (ID 54 - ranked *G* first)

24. "I prefer not to eat any of the choices of Red" (ID 55 - ranked *G* first)

25. "The top three were based on the items that I eat the most during breakfast and lunch." (ID 57 - ranked *GO* first, *GOR* second, *G* third)

26. "I chose my options based on whatever was healthiest for me." (ID 59 - ranked *GO* first)

27. "What determined my first 3 options had to do with how greens and orange choices played a role in it. They went in that order for me. I was indifferent to option OR and R since it was the least desirable for me. [...] The reason was that I do not eat beef, chicken and pork. The only thing was fish and i'm trying to cut that as well, but it is a slower process for that one. During this 8 week period i was trying to also not have it as well, so i chose the O" (ID 67 - ranked *GO* first)

28. "I wanted to eat healthy, by adding more salads and fruits to my diet." (ID 68 - ranked *GO* first)

29. "Because I didn't want to be tempted to eat anything in the red category." (ID 75 - ranked *G* and *GO* first)

30. "Did not want to eat unhealthy" (ID 83 - ranked *GO* first)

31. "I wanted to discourage myself from choosing junk food options at lunch." (ID 84 - ranked *GO* first)

32. "I chose Greens as my number one because I planned to eat healthier. [...] Because really group R I wasn't going to eat any of that stuff, and I didn't" (ID 86 - ranked *G* first)

33. "I don't eat anything in the red category except sweet potato fries and only rarely." (ID 90 - ranked *GO* first)

34. "G should be reimbursed most, as one should have them more frequently. Next comes GO for occasional allowance of O. Finally GOR comes to allow rare inclusion of R. [...] to make sure that one eats G more." (ID 94 - ranked *G* first)

35. "I mostly eat salads and soups and yogurt parfait. I rarely eat the options in the orange and red categories" (ID 98 - ranked *G* first)

36. "I based my ranking on what would be healthiest but also on my own preferences." (ID 99 - ranked *GO* first)

37. "I assigned the ranks based on the foods I eat - I consume red items very sporadically. I do not buy lunch often and if I would then it would be one of the fewer categories." (ID 100 - ranked *GO* first)

38. "I didn't want to be reimbursed for R." (ID 101 - ranked *GO* first)

39. "I based it off having a healthy variety. I wanted the option of sandwiches and salads, so chose GO as my first choice. From there I chose both of those separately for my second choice. As a third I opted for all three categories. It would allow for a range of food, which I find helpful when dieting. Restricting or limiting what I'm allowed to eat always stresses me out. For me, it's more about portion control. I'll order a burger if I'm craving it, but only eat half or order it without the bun for example. [...] I chose based on what I tend to eat for lunch. It's generally in the categories of G and O." (ID 104 - ranked *GO* first)

40. "I wanted to motivate myself to eat better." (ID 111 - ranked *GO* first)

41. "Because I don't eat burgers and fried food and I do not drink sodas. Thus, I need not be reimbursed for what I do not eat." (ID 118 - ranked *G* and *GO* first)

42. "Based on my normal choices." (ID 119 - ranked *GO* first)

43. "Having more options gives more flexibility to eat unhealthy foods." (ID 120 - ranked *O* first, *GO* second)

44. "I usually bring my own lunch and I don't consume the categories that are unhealthy, I am a nutritionist." (ID 122 - ranked *GO* first)

45. "I don't eat green category too often" (ID 123 - ranked *OR* first)

A.5.2. Comments of participants who assigned rank 1 to *GOR*

1. "I wanted the most variables. I started the challenge on a carb-free diet, so I chose the GOR option as my top choice, not to include fried food and pizza, but for the additional protein option (turkey burgers, etc.)" (ID 1 - only *GOR* first)

2. "I wanted to choose options that offered the most flexibility, which is why the combined categories was my first choice." (ID 7 - only *GOR* first)

3. "Because that gives me the most flexibility." (ID 8 - only *GOR* first)

4. "My thought process was I didn't know if I would be good or bad, so I decided to put all of them in for the first few, just to be safe. I thought I would rather cover everything since in my regular life I have all the options and have to try to be good knowing all options are available to me." (ID 10 - only *GOR* first)

5. "I know that I eat something different everyday, and sometimes I do not have time to pack a lunch, so I run out to a place nearby to grab something quick. I did not want to limit myself." (ID 25 - only *GOR* first)

6. "Because it did not restrict the types of foods I could eat." (ID 29 - only *GOR* first)

7. "I chose it because I thought it would enable more variety. However, I ended up not doing the reimbursement program because I do not buy lunch due to the high cost." (ID 31 - only *GOR* first)

8. "Maximize reimbursement" (ID 33 - only *GOR* first)

9. "I have difficulties in planning in advance what I will have for lunch" (ID 39 - only *GOR* first)

10. "Most of the time I bring salad for lunch, so if I do buy lunch it is occasionally a splurge like pizza or something under the R category. Other times it would be a salad or sandwich, so I wanted to keep all options open for reimbursement if possible." (ID 41 - only *GOR* first)

11. "Because I figured that this option would allow me to be reimbursed for more than the other options." (ID 42 - only *GOR* first)

12. "I gave the highest rankings to the broadest choices. The only food restriction I hold myself to is to try to eliminate added sugars and foods with added sugars. None of the categories were relevant to this." (ID 43 - ranked first *O*, *R*, *OR* and *GOR*)

13. "I wanted to choose all foods to be reimbursed, most times I eat healthy but I wanted to cover unhealthy as well, I felt I would be upset if I didn't achieve eating only healthy. I wanted to cover my bases" (ID 56 - only *GOR* first)

14. "Looking back, I wanted to have a balanced and realistic plan so I ranked the options with multiple categories first. [...]" (ID 65 - ranked *GO* and *GOR* first)

15. "I like to eat healthy" (ID 66 - ranked *GOR* first and *GO* second)

16. "I just was honest. I knew that I would eat Red food categories during the work week." (ID 72 - ranked GOR first and OR second)

17. "I wanted to eat healthy." (ID 73 - all options ranked first)

18. "I don't eat fast food or really at restaurants, but I kept my category pretty broad to make sure it covered everything that I normally eat, just in case I happened to eat at a restaurant once." (ID 76 - only GOR first)

19. "[...] maximize reimbursement potential" (ID 81 - only GOR first)

20. "I wanted to have more choices; even though I usually stay within the Orange group, I sometimes like to treat myself or cheat during difficult days when all I need is a burger." (ID 82 - only GOR first)

21. "so I chose the categories that would allow me the greatest chances of being reimbursed no matter what food I purchased. GOR, GO offered me the best chance of getting reimbursed since I would normally choose a sandwich and a seltzer for lunch." (ID 88 - G, O, GO and GOR first)

22. "I wanted flexibility in what I could eat. I expected that there would be cheat days and I ranked according to the plan that would allow me to be reimbursed for occasional cheat days." (ID 89 - only GOR first)

23. "I wanted to be self-motivated and not be motivated by that. [...] I did not want to limit my reimbursement" (ID 91 - only GOR first)

24. "Although I mostly eat salads, soups, fruit and other healthy options; I thought it would be better to have more choice when being reimbursed for the few days when I decided to eat something other than the most limited "healthy"option" (ID 92 - only GOR first)

25. "The options chosen were relatively based on eating desires. I like to mix it up" (ID 103 - ranked G, GR and GOR first)

26. "For the financial reward of having more items reimbursable." (ID 112 - only GOR first)

27. "Well, it would give the widest variety of potential foods I could be reimbursed for, especially if I decided to get a burger one day or something." (ID 121 - only GOR first)

TABLE 14: Classification and coding Scheme

Main reason for ranking	ID	% (N)
Panel A: Did not give rank 1 to *GOR*		
Desire to restrict diet choices	4, 11, 16, 19, 20, 22, 23, 24, 38, 40, 44 52, 53, 54, 68, 75, 84, 86, 94, 111, 120	47% (21)
Based on consumption habits	3, 5, 6, 18, 27, 30, 34, 35, 46, 57, 67 90, 98, 100, 104, 118, 119, 122, 123	42% (19)
Desire to eat healthy	55, 59, 83, 99, 101	11% (5)
Total		100% (45)
Panel B: Assigned rank 1 to *GOR*		
Having a balanced/diversified/realistic diet	1, 31, 65, 72, 103	19% (5)
Accommodating uncertainty/cheat days	10, 39, 56, 76, 82, 89, 92	26% (7)
Maximizing reimbursement	33, 42, 81, 88, 112, 121	22% (6)
Desire to eat healthy	66, 73	7% (2)
Preference for flexibility (general)	7, 8, 25, 29, 41, 43, 91	26% (7)
Total		100% (27)

Notes: Distribution of reasons given by respondents to Survey 2 in order to explain their decision to (not) assign rank 1 to *GOR*. For Panel A: "Desire to restrict diet choices" refers to all participants who mentioned an incentive to eat healthier and/or avoid tempting foods; "Based on consumption habits" refers to those who picked their preferred options based on what they usually eat. For Panel B: "Having a balanced/diversified/realistic diet" refers to participants who expressed a preference for variety and/or the desire to have an honest/realistic plan; "Accommodating uncertainty/cheat days" refers to those who mentioned facing some uncertainty in their meals and/or the possibility for occasional cheat days; "Maximizing reimbursement" corresponds to those who chose the widest coverage so as to maximize the amount reimbursed; "Preference for flexibilty (general)" refers to those who expressed a desire for flexibility, not to restrict their choice set, to keep options open etc... without more precision. Both panels: "Desire to eat healthy" refers to those who mentioned an intent to eat healthy/avoid unhealthy foods, without providing more details.

B. Goal setting contract and attendance

B.1 Goal setting and updating

During Survey 1, participants were offered to receive their $20 payment for completing the study only if they achieved self-set attendance goals belonging to 1, 2, or each of the following 3 categories: number of gym visits during one month, number of follow-up weigh-ins (out of 3) and number of wellness events (out of 4). Participants were instructed to enter a target number in 3 text boxes corresponding to each goal category and were explicitly asked to enter 0 if they did not want to set a goal. They were also given an example to explain how the contract works. Unfortunately, due to a mistake in the programming of the survey, the software initially allowed participants to move to the next section without having entered a number for one or several of the categories. Non-response was recorded as 0 and thus treated as having chosen not to commit. The programming error was rectified within the first 24 hours the survey was sent out, meaning that all participants after the first day were required to enter a number in all fields in order to move on. The table below shows contract take-up by completion date, as well as the distribution of the goals set.[6] As Survey 1 was sent out late on the night of March 4th, I grouped March 4th and March 5th under Day 1.

TABLE 15: Contract take-up by day

	Survey completion date T			
	Day 1 March 4th and 5th	Day 2 March 6th	$\geq$ Day 3 March 7th and after	All
# goals set	% of date T respondents (N)			
0 goal	52.6% (20)	25.5% (12)	25.0% (7)	34.5% (39)
1 goal	15.8% (6)	19.1% (9)	17.9% (5)	17.7% (20)
2 goals	23.7% (9)	27.7% (13)	39.3% (11)	29.2% (33)
3 goals	7.9% (3)	27.7% (13)	17.9% (5)	18.6% (21)
% of total respondents (N)	33.6% (38)	41.6% (47)	24.8% (28)	100% (113)
Take-up (any goal)	**47.4%**	**74.5%**	**75.0%**	**65.5%**
F-test	Day 1 = Day 2: $F(1,110) = 7.17$		$p = 0.009$	
	Day 1 = $\geq$ Day 3: $F(1,110) = 5.72$		$p = 0.018$	

Notes: Take-up refers to the percentage of participants who selected a positive target number in at least one of the three goal categories specified in the goal setting contract. The category "March 7th and after" includes all participants who completed the survey between March 7th and March 11th, as well as 4 participants for whom the goals were updated. The F-tests are from linear regressions of contract take-up on day dummies.

[6] Findings are nearly identical when using the date at which a participant started (but not necessarily completed) the survey; 11 participants started and finished Survey 1 on different days.

Of the 113 respondents, about a third completed the survey within the first 24 hours, another 42% completed it on Day 2 (March 6th), and 25% after Day 2.[7] The take-up rate increases from 47% on Day 1 to about 75% after. In the econometric analysis of contract take-up and default on goals, I therefore control for whether or not the survey was completed on Day 1.

Two other points should be noted about the goal setting decisions. First, although there were only three *follow-up* weigh-ins, four respondents entered a goal of 4, corresponding to the total number of weigh-ins. The weigh-in goal for those respondents was adjusted to 3. Second, for the gym goal, some participants mistakenly thought that the free membership would be valid for the entire duration of the challenge, although the instructions stipulated a validity of one month. Due to the initial confusion, all respondents were sent a summary of their goals within 10 days of having completed Survey 1; the reports were all sent at the same time, after the survey was closed. Participants were asked to verify their goals and were given some wiggle room to change their goals if needed. In total, six participants made a change to their goals. Among them, three participants had determined their gym goal based on 8 weeks, one participant originally missed the option to commit and asked to set goals, and two participants asked to lower down their goals due to sickness/absence. All changes occurred before the second weigh-in at the exception of one participant for whom the change was made the day following the second weigh-in.

To limit the chances that participants forget their goals, a reminder was sent before each weigh-in. The reminder sent before the second weigh-in included information about all three goal categories and listed the dates of the next weigh-ins and wellness events. The reminders preceeding the third and fourth weigh-ins only reminded participants about their weigh-in goal, since all wellness events had already occurred and the gym goal was only for one month. To see whether participants remembered their goal setting decisions by the end of the challenge, those who replied to Survey 2 were asked whether they had chosen to receive the $20 payment conditional on achieving a goal. The vast majority of respondents correctly replied to the question and the likelihood of correctly remembering did not depend on the actual decision: of the 87 respondents, 92% (80) correctly remembered their decision to commit or not to a goal; this was the case of 32 of the 35 participants who set no goal and 48 of the 52 participants who did set a goal (91.4% vs. 92.3%, $z = 0.15$, $p = 0.88$). Furthermore, there appears to be no relationship between the menu preferences expressed in the reimbursement program and participants' likelihood of correctly remembering their goal setting decision. Finally, Survey 2 respondents who default on their goal setting contract were not more likely to have forgotten their decision to commit or not to a goal.

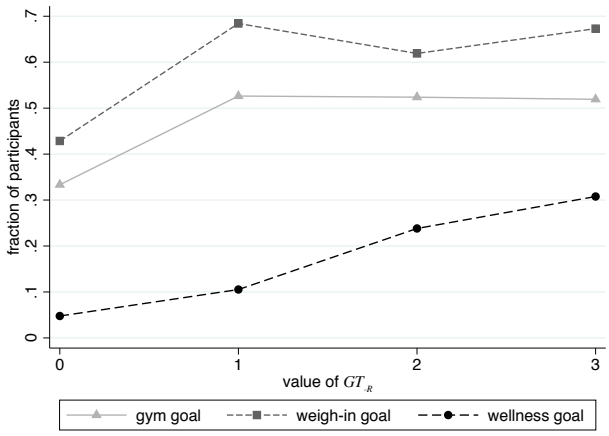
[7] In the group "March 7th and after", 18 completed the survey on Day 3 (March 7th) and 6 completed the survey between Day 4 and Day 7 (March 8th to 11th); the last official day to complete the survey was March 11th. In addition, the group "March 7th and after" includes 4 participants who asked to modify their goals and for whom the survey completion date corresponds to the modification date.

B.2 Goal setting and menu preferences

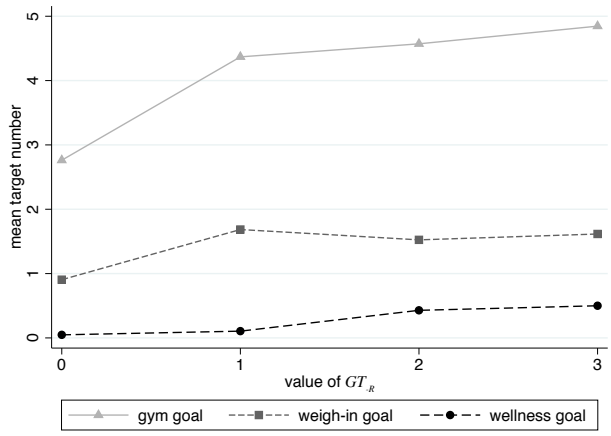
Figure **11** shows that there is a positive relationship between goal setting and temptation by R for all 3 goal categories. In particular, those for whom R did not appear tempting at all (i.e., $GT_{-R} = 0$) were less likely than the other participants to set goals of any type.

FIGURE 11: Goal setting by value of the GT_{-R} index

(A) Proportion who commit to a goal



(B) Level of the goal



Notes: Panel (a) refers to the proportion of participants who entered a positive target number for a given goal category as a function of their GT_{-R} score. Panel (b) is the mean target number chosen by participants with a given GT_{-R} score.

In Table 6 of the main text, I present linear regressions of contract take-up on various menu preference measures and a set of individual controls. I find that participants who ranked GO as their unique top (variable *GO top*) were about 20 percentage points more likely to commit to a goal compared to those who strictly preferred GOR; furthermore, an increase in the GT_{-R} (SSB_{-R}) index by one point corresponded to a 7-10 (3-5) percentage point increase in the likelihood of taking up the contract. Table **16** repeats the analysis on the subset of subjects for whom $G \succ R$ and $O \succ R$. The relationship between the GT_{-R} (SSB_{-R}) index and contract take-up is substantially weaker and loses statistical significance; the effect of *GO top* remains significant and of a similar size after (but not before) including all the individual controls. Thus, part of the relationship between contract take-up and the menu preference measures studied in this paper appears to be explained by individual heterogeneity in participants' ordering of the singletons. Tables **17** (all subjects) and **18** (subset with $G \succ R$ and $O \succ R$) present similar regressions for the goal level chosen in each goal category (controlling for intentions to attend). The positive effect of *GO top* is robust across nearly all specifications, while the effect of GT_{-R} (SSB_{-R}) appears less robust.

TABLE 16: Determinants of contract take-up
(subsample of participants with preferences $G \succ R$ and $O \succ R$)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
G top	0.041	0.114						
	(0.146)	(0.151)						
GO top	0.140	0.225**						
	(0.116)	(0.112)						
Other top	0.106	0.171						
	(0.155)	(0.152)						
GT_{-R}			0.007	0.050	0.054			
			(0.045)	(0.045)	(0.048)			
GT_{-G}					0.121			
					(0.119)			
GT_{-O}					-0.096			
					(0.096)			
SSB_{-R}						0.009	0.023	0.038
						(0.020)	(0.020)	(0.026)
SSB_{-G}								-0.001
								(0.090)
SSB_{-O}								-0.078
								(0.055)
female		-0.105		-0.096	-0.122		-0.100	-0.137
		(0.118)		(0.115)	(0.118)		(0.115)	(0.118)
single		0.075		0.085	0.075		0.085	0.069
		(0.097)		(0.095)	(0.097)		(0.095)	(0.098)
age		0.004		0.003	0.003		0.004	0.004
		(0.005)		(0.005)	(0.005)		(0.005)	(0.005)
years of educ		-0.102***		-0.098***	-0.096***		-0.097***	-0.098***
		(0.032)		(0.033)	(0.033)		(0.033)	(0.033)
prior participant		-0.063		-0.034	-0.022		-0.045	-0.032
		(0.100)		(0.098)	(0.099)		(0.098)	(0.099)
weight loss goal		0.006		0.007	0.007		0.006	0.006
		(0.004)		(0.004)	(0.004)		(0.004)	(0.004)
goal confidence		0.753***		0.742***	0.750***		0.749***	0.780***
		(0.250)		(0.248)	(0.252)		(0.248)	(0.252)
$(goal\ confidence)^2$		-0.083***		-0.082***	-0.083***		-0.082***	-0.085***
		(0.026)		(0.025)	(0.026)		(0.025)	(0.026)
diets attempted		0.008		0.008	0.008		0.009	0.009
		(0.010)		(0.010)	(0.010)		(0.010)	(0.010)
Day 1 decision	-0.207*	-0.226**	-0.223**	-0.259***	-0.249**	-0.218**	-0.241**	-0.228**
	(0.105)	(0.097)	(0.104)	(0.096)	(0.098)	(0.104)	(0.096)	(0.097)
N	90	90	90	90	90	90	90	90
adj. R^2	0.024	0.193	0.029	0.184	0.176	0.031	0.186	0.186

Notes: Linear probability models where the dependent variable is equal to 1 if the respondent committed to at least one goal; regressions run on the subset of participants for whom $G \succ R$ and $O \succ R$. *Day 1 decision* is an indicator variable equal to 1 if the respondent completed Survey 1 within the first 24 hours the survey was administered; see Section 3.3 and Table 4 of the main text for a description of the other control variables. Standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 17: Determinants of chosen goal

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
G top	0.361	0.240	0.305								
	(0.632)	(0.628)	(0.623)								
GO top	0.994*	0.888**	1.068**								
	(0.506)	(0.443)	(0.460)								
Other top	0.323	0.058	0.032								
	(0.570)	(0.550)	(0.490)								
GT_{-R}				0.306*	0.250*	0.318**	0.301**				
				(0.165)	(0.149)	(0.137)	(0.135)				
GT_{-G}							-0.046				
							(0.309)				
GT_{-O}							0.165				
							(0.274)				
SSB_{-R}								0.117	0.056	0.095	0.128*
								(0.079)	(0.073)	(0.069)	(0.067)
SSB_{-G}											-0.264*
											(0.147)
SSB_{-O}											-0.128
											(0.144)
weigh-in category	-2.850***	-0.044	0.072	-2.850***	-0.066	0.048	0.064	-2.850***	-0.070	0.019	0.034
	(0.494)	(0.514)	(0.520)	(0.493)	(0.516)	(0.524)	(0.517)	(0.493)	(0.515)	(0.525)	(0.525)
wellness category	-3.991***	-0.736	-0.601	-3.991***	-0.762	-0.630	-0.611	-3.991***	-0.767	-0.663	-0.646
	(0.513)	(0.591)	(0.592)	(0.511)	(0.592)	(0.596)	(0.590)	(0.511)	(0.591)	(0.599)	(0.598)
Day 1 decision	-0.804*	-0.921**	-0.766*	-0.892**	-0.989**	-0.840**	-0.884**	-0.843*	-0.969**	-0.800**	-0.759*
	(0.448)	(0.409)	(0.387)	(0.429)	(0.391)	(0.378)	(0.402)	(0.433)	(0.396)	(0.381)	(0.401)
planned attendance		0.280***	0.292***		0.278***	0.289***	0.291***		0.278***	0.287***	0.288***
		(0.063)	(0.063)		(0.063)	(0.064)	(0.064)		(0.063)	(0.064)	(0.064)
female			-0.115			-0.175	-0.153			-0.104	-0.258
			(0.389)			(0.384)	(0.398)			(0.395)	(0.393)
single			0.118			0.259	0.275			0.233	0.180
			(0.390)			(0.382)	(0.381)			(0.388)	(0.392)
age			0.058***			0.060***	0.056**			0.061***	0.064***
			(0.021)			(0.022)	(0.024)			(0.022)	(0.023)
years of educ			-0.408***			-0.407***	-0.400***			-0.393***	-0.403***
			(0.118)			(0.131)	(0.133)			(0.136)	(0.134)
prior participant			-0.687*			-0.604*	-0.612*			-0.631*	-0.674*
			(0.357)			(0.361)	(0.363)			(0.365)	(0.355)
weight loss goal			0.005			0.004	0.005			0.002	0.002
			(0.013)			(0.013)	(0.013)			(0.013)	(0.014)
goal confidence			0.764			0.670	0.635			0.785	0.814
			(0.933)			(0.893)	(0.896)			(0.881)	(0.915)
$(goal\ confidence)^2$			-0.099			-0.093	-0.091			-0.101	-0.104
			(0.095)			(0.091)	(0.091)			(0.090)	(0.093)
diets attempted			0.053			0.045	0.043			0.044	0.045
			(0.043)			(0.042)	(0.042)			(0.043)	(0.043)
N	339	339	339	339	339	339	339	339	339	339	339
adj. R^2	0.210	0.367	0.402	0.212	0.366	0.399	0.396	0.209	0.362	0.394	0.397

Notes: Linear regressions of the goal number chosen by a respondent for goal category j. The variables *weigh-in* (*wellness*) category are indicators for the goal category; *planned attendance* refers to the number of times a respondent planned to perform the activity for goal category j. *Day 1 decision* is an indicator equal to 1 if the respondent completed Survey 1 within the first 24 hours the survey was administered; see Section 3.3 and Table 4 of the main text for a description of the other control variables. Standard errors in parentheses clustered at the subject level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

TABLE 18: Determinants of chosen goal
(subsample of participants with preferences $G \succ R$ and $O \succ R$)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
G top	0.510	0.599	0.791								
	(0.706)	(0.690)	(0.687)								
GO top	0.901	0.952**	1.284***								
	(0.544)	(0.471)	(0.471)								
Other top	0.154	-0.006	0.240								
	(0.585)	(0.568)	(0.479)								
GT_{-R}				0.158	0.148	0.272	0.290				
				(0.208)	(0.186)	(0.183)	(0.185)				
GT_{-G}							-0.168				
							(0.472)				
GT_{-O}							-0.008				
							(0.371)				
SSB_{-R}								0.099	0.046	0.094	0.205**
								(0.090)	(0.084)	(0.081)	(0.098)
SSB_{-G}											-0.528
											(0.337)
SSB_{-O}											-0.212
											(0.204)
weigh-in category	-2.878***	0.202	0.443	-2.878***	0.175	0.404	0.388	-2.878***	0.160	0.367	0.387
	(0.565)	(0.535)	(0.532)	(0.563)	(0.538)	(0.544)	(0.543)	(0.563)	(0.541)	(0.549)	(0.541)
wellness category	-4.133***	-0.557	-0.278	-4.133***	-0.589	-0.323	-0.341	-4.133***	-0.606	-0.365	-0.342
	(0.579)	(0.624)	(0.614)	(0.577)	(0.627)	(0.626)	(0.626)	(0.577)	(0.630)	(0.633)	(0.623)
Day 1 decision	-0.831*	-0.950**	-0.977**	-0.894*	-0.994**	-1.066**	-1.050**	-0.827*	-0.952**	-0.979**	-0.990**
	(0.497)	(0.432)	(0.416)	(0.499)	(0.442)	(0.432)	(0.459)	(0.492)	(0.438)	(0.424)	(0.428)
planned attendance		0.296***	0.319***		0.293***	0.315***	0.314***		0.292***	0.312***	0.314***
		(0.067)	(0.067)		(0.068)	(0.069)	(0.068)		(0.068)	(0.069)	(0.068)
female			-0.083			-0.159	-0.158			-0.139	-0.265
			(0.467)			(0.444)	(0.434)			(0.447)	(0.456)
single			-0.059			0.006	-0.006			0.009	-0.162
			(0.406)			(0.408)	(0.406)			(0.410)	(0.411)
age			0.048**			0.051**	0.053*			0.053**	0.056**
			(0.022)			(0.024)	(0.028)			(0.025)	(0.025)
years of educ			-0.421***			-0.399**	-0.403**			-0.385**	-0.419**
			(0.146)			(0.162)	(0.166)			(0.167)	(0.164)
prior participant			-0.695*			-0.580	-0.587			-0.622	-0.640
			(0.389)			(0.397)	(0.402)			(0.398)	(0.390)
weight loss goal			0.011			0.012	0.011			0.009	0.011
			(0.014)			(0.014)	(0.014)			(0.015)	(0.015)
goal confidence			2.196**			2.201**	2.248**			2.235***	2.523***
			(0.925)			(0.853)	(0.876)			(0.846)	(0.893)
$(goal\ confidence)^2$			-0.249**			-0.250***	-0.254***			-0.251***	-0.284***
			(0.097)			(0.091)	(0.093)			(0.090)	(0.095)
diets attempted			0.072**			0.062**	0.063**			0.062**	0.073**
			(0.033)			(0.031)	(0.032)			(0.031)	(0.034)
N	270	270	270	270	270	270	270	270	270	270	270
adj. R^2	0.214	0.406	0.439	0.211	0.399	0.428	0.424	0.213	0.398	0.426	0.428

Notes: Linear regressions of the goal number chosen by a respondent for goal category j; regressions run on the subset of participants for whom $G \succ R$ and $O \succ R$. The variables *weigh-in* (*wellness*) category are indicators for the goal category; *planned attendance* refers to the number of times a respondent planned to perform the activity for goal category j. *Day 1 decision* is an indicator equal to 1 if the respondent completed Survey 1 within the first 24 hours the survey was administered; see Section 3.3 and Table 4 of the main text for a description of the other control variables. Standard errors in parentheses clustered at the subject level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

B.3 Goal setting, actual attendance and contract default

Figure **12** shows actual attendance for each goal category (gym visits, weigh-ins, wellness events) comparing participants did or did not set a goal for that category. Regardless of the category, participants who set a goal had a higher attendance than those who did not (p-value < 0.05 for all categories on a Kolmogorov-Smirnov test). However, since participants were not randomly assigned to different treatments with or without a goal setting option, one cannot conclude that goal setting per se increases attendance. Rather, these results suggest that goal setters are more likely to follow up on their intentions, possibly because they care more about achieving their goals.

FIGURE 12: Goal setting and attendance

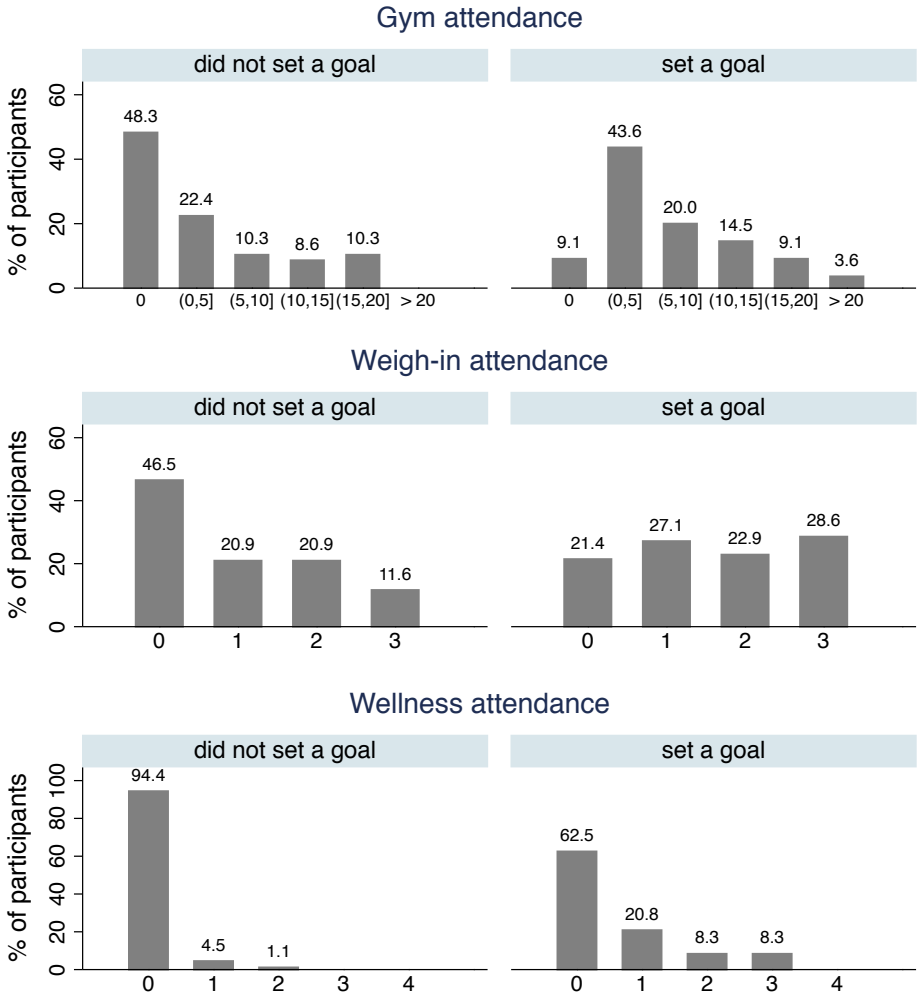
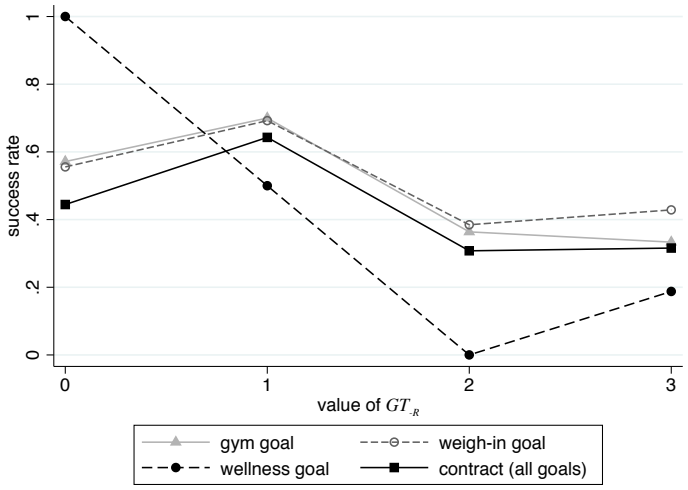


Figure **13** shows the fraction of goal setters who met their goals for each goal category and overall, as a function of their GT_R score. Participants who appeared more tempted by red foods ($GT\_R \in \{2,3\}$) were less likely to reach their goals than those who showed little to no temptation ($GT\_R \in \{0,1\}$). A similar pattern is observed for each goal category and for overall success. For instance, among those who committed to at least one goal ($N = 74$), the 51 participants for whom $GT\_R \in \{2,3\}$ were 25 percentage points more likely to default on their contract than the 23 participants with $GT\_R \in \{0,1\}$ (68.6% vs. 43.5%, $z = 2.05$, $p = 0.040$). Similarly, among those who committed to a weigh-in goal ($N = 70$), the 48 participants with $GT\_R \in \{2,3\}$ were 22 percentage points less likely to achieve their goal than the 22 participants with $GT\_R \in \{0,1\}$ (41.7% vs. 63.6%, $z = 1.71$, $p = 0.088$).[8]

FIGURE 13: Success rates on goals by value of the GT_R index



Notes: For "contract (all goals)", participants succeeded on their goal setting contract provided they met or surpassed all their goals. Participant with missing gym attendance assumed to have failed his gym goal.

In Table 7 of the main text, I present linear regressions of goal achievement on menu preferences and individual controls; Table **19** presents the same regressions for the subset of participants with $G \succ R$ and $O \succ R$. The main findings appear robust to restricting attention to this subsample. Tables **20** and **21** present a breakdown by goal, pooling the small number of participants who set a wellness goal with those who set a weigh-in goal; the same patterns are observed irrespective of the goal category. Finally, Table **22** looks at contract default; the relationships of interest appear weaker, but the general message is preserved.

[8] Among those who set a gym goal ($N = 54$, excluding the participant with missing gym attendance), the 38 participants with $GT\_R \in \{2,3\}$ were 30 percentage points less likely to achieve their goal than the 17 participants with $GT\_R \in \{0,1\}$ (34.2% vs. 64.7%, $z = 2.11$, $p = 0.035$). Finally, among those who set a wellness goal ($N = 24$), the 21 participants with $GT\_R \in \{2,3\}$ were 52 percentage points less likely to achieve their goal than the 3 participants with $GT\_R \in \{0,1\}$ (14.3% vs. 66.7%, $z = 2.09$, $p = 0.037$).

TABLE 19: Determinants of goal achievement
(subsample of participants with preferences $G \succ R$ and $O \succ R$)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
G top	-0.380**	-0.375**						
	(0.156)	(0.154)						
GO top	-0.223	-0.260*						
	(0.150)	(0.149)						
Other top	-0.204	-0.169						
	(0.168)	(0.151)						
GT_{-R}			-0.099*	-0.102*	-0.100*			
			(0.054)	(0.055)	(0.058)			
GT_{-G}					-0.062			
					(0.132)			
GT_{-O}					0.015			
					(0.116)			
SSB_{-R}						-0.034	-0.036	-0.077**
						(0.022)	(0.022)	(0.031)
SSB_{-G}								0.091
								(0.085)
SSB_{-O}								0.142
								(0.088)
female		0.017		0.046	0.048		0.062	0.100
		(0.110)		(0.123)	(0.124)		(0.129)	(0.131)
single		0.074		0.087	0.076		0.100	0.123
		(0.114)		(0.113)	(0.112)		(0.111)	(0.111)
age		-0.003		-0.005	-0.004		-0.005	-0.006
		(0.007)		(0.006)	(0.007)		(0.006)	(0.005)
years of educ		0.023		0.030	0.030		0.018	0.047
		(0.042)		(0.038)	(0.039)		(0.037)	(0.038)
prior participant		0.196		0.171	0.166		0.186	0.224*
		(0.120)		(0.113)	(0.117)		(0.117)	(0.118)
weight loss goal		-0.005		-0.004	-0.004		-0.003	-0.004
		(0.003)		(0.003)	(0.003)		(0.003)	(0.003)
goal confidence		0.026		0.025	0.042		0.027	-0.045
		(0.300)		(0.289)	(0.295)		(0.279)	(0.303)
$(goal\ confidence)^2$		0.001		0.001	-0.001		-0.000	0.009
		(0.031)		(0.030)	(0.031)		(0.030)	(0.032)
diets attempted		-0.024*		-0.025**	-0.025**		-0.024*	-0.031**
		(0.012)		(0.012)	(0.011)		(0.012)	(0.013)
Day 1 decision	0.149	0.157	0.183	0.186	0.189	0.155	0.152	0.175
	(0.139)	(0.125)	(0.141)	(0.125)	(0.125)	(0.140)	(0.126)	(0.129)
N	128	128	128	128	128	128	128	128
adj. R^2	0.092	0.130	0.090	0.128	0.114	0.076	0.117	0.135

Notes: Linear probability models where the dependent variable is equal to 1 if the respondent achieved his goal for goal category $j \in$ {gym, weigh-ins, wellness}; regressions run on the subset of participants for whom $G \succ R$ and $O \succ R$. Participant with missing gym attendance assumed to have failed his gym goal. All regressions include a control for selected target numbers and dummies for goal category. *Day 1 decision* is equal to 1 if the participant completed Survey 1 within the first 24 hours the survey was administered; see Section 3.3 and Table 4 of the main text for a description of the other control variables. Standard errors in parentheses clustered at the subject level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

TABLE 20: Determinants of goal achievement - weigh-in and wellness goals

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
G top	-0.268*	-0.330*						
	(0.153)	(0.168)						
GO top	-0.167	-0.196						
	(0.134)	(0.154)						
Other top	-0.034	0.005						
	(0.132)	(0.135)						
GT_{-R}			-0.078	-0.102*	-0.095			
			(0.051)	(0.057)	(0.058)			
GT_{-G}					0.122			
					(0.087)			
GT_{-O}					-0.068			
					(0.089)			
SSB_{-R}						-0.028	-0.037*	-0.056**
						(0.018)	(0.020)	(0.023)
SSB_{-G}								0.085
								(0.056)
SSB_{-O}								0.059
								(0.083)
goal target	-0.250***	-0.217***	-0.242***	-0.204**	-0.226***	-0.254***	-0.227***	-0.229***
	(0.060)	(0.077)	(0.062)	(0.078)	(0.081)	(0.063)	(0.080)	(0.081)
wellness category	-0.439***	-0.439***	-0.420***	-0.405***	-0.428***	-0.449***	-0.442***	-0.424***
	(0.094)	(0.112)	(0.098)	(0.108)	(0.111)	(0.099)	(0.110)	(0.107)
Day 1 decision	0.206*	0.228**	0.220*	0.241**	0.252**	0.204*	0.218**	0.200*
	(0.115)	(0.103)	(0.114)	(0.103)	(0.104)	(0.111)	(0.102)	(0.109)
female		-0.023		0.003	0.015		0.013	0.055
		(0.115)		(0.123)	(0.127)		(0.127)	(0.133)
single		-0.075		-0.073	-0.053		-0.063	-0.029
		(0.109)		(0.106)	(0.107)		(0.106)	(0.110)
age		-0.004		-0.007	-0.007		-0.008	-0.008*
		(0.006)		(0.005)	(0.006)		(0.005)	(0.005)
years of educ		0.028		0.037	0.034		0.028	0.042
		(0.031)		(0.030)	(0.032)		(0.030)	(0.030)
prior participant		0.263**		0.239*	0.236*		0.236*	0.269**
		(0.119)		(0.122)	(0.121)		(0.122)	(0.125)
weight loss goal		-0.004		-0.002	-0.002		-0.001	-0.001
		(0.003)		(0.003)	(0.003)		(0.003)	(0.003)
goal confidence		-0.736*		-0.700*	-0.727**		-0.699**	-0.751**
		(0.381)		(0.352)	(0.359)		(0.346)	(0.354)
$(goal\ confidence)^2$		0.079**		0.075**	0.077**		0.073**	0.079**
		(0.039)		(0.036)	(0.037)		(0.036)	(0.036)
diets attempted		-0.011		-0.010	-0.009		-0.010	-0.012
		(0.013)		(0.012)	(0.011)		(0.012)	(0.013)
N	94	94	94	94	94	94	94	94
adj. R^2	0.202	0.224	0.210	0.226	0.218	0.205	0.218	0.223

Notes: Linear probability models where the dependent variable is equal to 1 if the respondent achieved the goal he set for goal category $j \in$ {weigh-in, wellness}. The variable *goal target* is the selected target number and *wellness category* is an indicator equal to 1 if the goal is for the wellness events. *Day 1 decision* is an indicator equal to 1 if the respondent completed Survey 1 within the first 24 hours the survey was administered; see Section 3.3 and Table 4 of the main text for a description of the other control variables. Standard errors in parentheses clustered at the subject level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

TABLE 21: Determinants of goal achievement - gym goal

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
G top	-0.412*	-0.469*						
	(0.237)	(0.251)						
GO top	-0.146	-0.279						
	(0.168)	(0.189)						
Other top	-0.256	-0.252						
	(0.198)	(0.198)						
GT_{-R}			-0.105*	-0.149**	-0.141*			
			(0.061)	(0.067)	(0.072)			
GT_{-G}					0.037			
					(0.171)			
GT_{-O}					-0.085			
					(0.170)			
SSB_{-R}						-0.034	-0.049*	-0.080**
						(0.026)	(0.028)	(0.034)
SSB_{-G}								0.032
								(0.091)
SSB_{-O}								0.150
								(0.110)
goal target	-0.020	-0.018	-0.021	-0.022*	-0.021	-0.021	-0.022*	-0.024*
	(0.014)	(0.014)	(0.013)	(0.013)	(0.014)	(0.014)	(0.013)	(0.014)
Day 1 decision	0.116	0.111	0.134	0.141	0.149	0.109	0.113	0.072
	(0.165)	(0.161)	(0.161)	(0.154)	(0.158)	(0.162)	(0.156)	(0.160)
gym visits recorded	-0.030	-0.025	-0.080	-0.045	-0.030	-0.095	-0.070	-0.068
	(0.147)	(0.168)	(0.138)	(0.157)	(0.166)	(0.139)	(0.160)	(0.163)
female		0.032		0.095	0.080		0.115	0.196
		(0.184)		(0.175)	(0.182)		(0.181)	(0.189)
single		0.136		0.124	0.131		0.149	0.139
		(0.152)		(0.147)	(0.166)		(0.149)	(0.168)
age		-0.002		-0.006	-0.004		-0.007	-0.009
		(0.007)		(0.007)	(0.008)		(0.007)	(0.007)
years of educ		0.030		0.056	0.051		0.039	0.065
		(0.053)		(0.049)	(0.051)		(0.048)	(0.051)
prior participant		0.076		0.065	0.069		0.053	0.118
		(0.189)		(0.183)	(0.188)		(0.186)	(0.194)
weight loss goal		-0.004		-0.002	-0.003		-0.001	-0.001
		(0.006)		(0.006)	(0.006)		(0.006)	(0.006)
goal confidence		1.135**		1.133**	1.130**		1.113**	1.105**
		(0.444)		(0.427)	(0.437)		(0.435)	(0.433)
$(goal\ confidence)^2$		-0.111**		-0.110**	-0.109**		-0.110**	-0.108**
		(0.046)		(0.044)	(0.045)		(0.045)	(0.045)
diets attempted		-0.048***		-0.049***	-0.048***		-0.049***	-0.052***
		(0.017)		(0.016)	(0.016)		(0.016)	(0.016)
N	55	55	55	55	55	55	55	55
adj. R^2	0.024	0.130	0.050	0.179	0.143	0.027	0.146	0.154

Notes: Linear probability models where the dependent variable is equal to 1 if the respondent achieved his gym goal. The variable *goal target* is the selected target number and *gym visits recorded* is equal to 1 if the respondent's gym attendance was recorded by the system. *Day 1 decision* is an indicator equal to 1 if the respondent completed Survey 1 within the first 24 hours the survey was administered; see Section 3.3 and Table 4 of the main text for a description of the other control variables. Standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 22: Determinants of contract default

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
G top	0.325*	0.419**						
	(0.168)	(0.189)						
GO top	0.165	0.225						
	(0.127)	(0.142)						
Other top	0.012	0.049						
	(0.149)	(0.159)						
GT_{-R}			0.061	0.083	0.070			
			(0.049)	(0.055)	(0.057)			
GT_{-G}					-0.109			
					(0.126)			
GT_{-O}					0.108			
					(0.116)			
SSB_{-R}						0.019	0.028	0.042
						(0.020)	(0.023)	(0.027)
SSB_{-G}								-0.060
								(0.074)
SSB_{-O}								-0.052
								(0.082)
gym goal	0.008	0.003	0.010	0.007	0.005	0.009	0.006	0.006
	(0.009)	(0.010)	(0.009)	(0.010)	(0.011)	(0.009)	(0.010)	(0.010)
weigh-in goal	0.142**	0.151**	0.148**	0.145*	0.161**	0.149**	0.150**	0.150*
	(0.062)	(0.073)	(0.062)	(0.073)	(0.075)	(0.062)	(0.073)	(0.076)
wellness goal	0.169***	0.231***	0.149**	0.191**	0.201**	0.165***	0.216***	0.197**
	(0.060)	(0.075)	(0.063)	(0.077)	(0.079)	(0.061)	(0.076)	(0.079)
Day 1 decision	-0.206*	-0.242*	-0.215*	-0.249*	-0.253*	-0.201*	-0.229*	-0.213
	(0.117)	(0.126)	(0.119)	(0.130)	(0.132)	(0.119)	(0.129)	(0.131)
gym visits recorded		0.138		0.146	0.146		0.155	0.150
		(0.134)		(0.136)	(0.137)		(0.136)	(0.137)
female		-0.086		-0.135	-0.130		-0.148	-0.173
		(0.139)		(0.140)	(0.142)		(0.141)	(0.144)
single		0.195		0.151	0.159		0.154	0.113
		(0.128)		(0.124)	(0.128)		(0.124)	(0.131)
age		0.002		0.005	0.003		0.005	0.006
		(0.006)		(0.006)	(0.007)		(0.006)	(0.007)
years of educ		-0.038		-0.038	-0.033		-0.032	-0.040
		(0.045)		(0.046)	(0.048)		(0.046)	(0.047)
prior participant		-0.069		-0.014	-0.036		-0.013	-0.034
		(0.150)		(0.149)	(0.152)		(0.150)	(0.154)
weight loss goal		-0.001		-0.002	-0.002		-0.003	-0.003
		(0.005)		(0.005)	(0.006)		(0.005)	(0.006)
goal confidence		-0.154		-0.258	-0.228		-0.216	-0.214
		(0.404)		(0.406)	(0.412)		(0.407)	(0.411)
$(goal\ confidence)^2$		0.013		0.024	0.021		0.020	0.020
		(0.042)		(0.042)	(0.043)		(0.043)	(0.043)
diets attempted		0.015		0.015	0.013		0.015	0.016
		(0.013)		(0.013)	(0.013)		(0.013)	(0.013)
N	74	74	74	74	74	74	74	74
adj. R^2	0.236	0.213	0.220	0.184	0.169	0.213	0.174	0.162

Notes: Linear probability models where the dependent variable is equal to 1 if the participant defaulted on his goal-setting contract (i.e., failed to achieve one of his goals). The variables *gym goal*, *weigh-in goal* and *wellness goal* refer to the selected target numbers; *Day 1 decision* is an indicator equal to 1 if the respondent completed Survey 1 within the first 24 hours the survey was administered; *gym visits recorded* is equal to 1 if the respondent's gym attendance was recorded by the system. See Section 3.3 and Table 4 of the main text for a description of the other control variables. Standard errors in parentheses.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

B.4 Weigh-in attendance, gym attendance, and other measures of participation

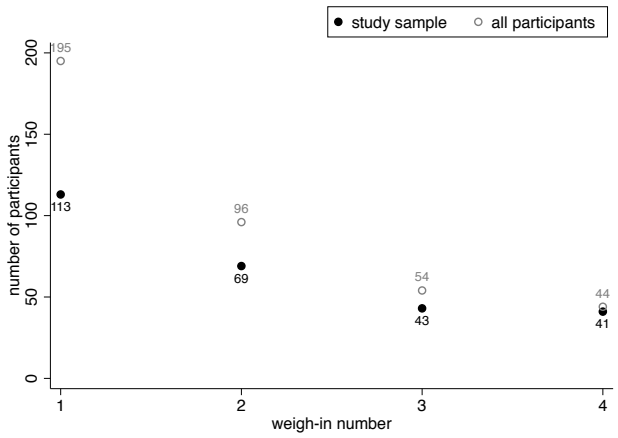
In this section, I discuss several points related to attendance and participation in the challenge. First, I look at weigh-in attendance over time and likelihood of attending the final weigh-in (i.e., of completing the challenge). I then discuss issues around the measurement of gym attendance, as well as its relationship with menu preferences. Finally, I discuss the relationship between menu preferences and two other outcomes: (*i*) likelihood of completing Survey 2 and (*ii*) likelihood of submitting lunch receipts for reimbursement.

B.4.1. Weigh-in attendance

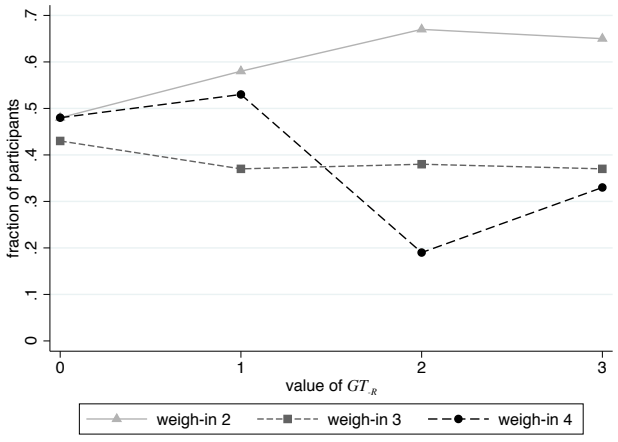
As Figure **14** shows, participants who enrolled in the study were more likely to stay in the challenge than those who did not (as expected given the enrollment criteria; see instructions). Among the study participants, those who were more tempted by R were less likely to complete the challenge, despite showing higher rates of attendance initially. Thus, commitment to removing R from the coverage is unlikely to be simply due to a demand effect; rather, it appears to reflect an initial, albeit difficult to sustain, motivation to lose weight.

FIGURE 14: Attendance of the weigh-ins

(A) Attendance numbers over time



(B) Attendance rate by value of the GT_{-R} index



In Tables **23** (entire sample) and **24** (subsample with $G \succ R$ and $O \succ R$), I further test the relationship between challenge completion and temptation by R in regressions that control for attendance of the intermediate weigh-ins. The negative relationship shown in Figure **14** between the GT_{-R} score and attendance of weigh-in 4 is confirmed; the effect of SSB_{-R} is also negative, but less robust, and the effect of *G top* or *GO top* is not significant.

TABLE 23: Determinants of the likelihood of completing the challenge

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
G top	-0.099	-0.100						
	(0.126)	(0.134)						
GO top	-0.096	-0.125						
	(0.103)	(0.108)						
Other top	-0.158	-0.155						
	(0.115)	(0.118)						
GT_{-R}			-0.069**	-0.077**	-0.070*			
			(0.035)	(0.037)	(0.038)			
GT_{-G}					0.115*			
					(0.068)			
GT_{-O}					-0.014			
					(0.066)			
SSB_{-R}						-0.022	-0.028*	-0.041**
						(0.016)	(0.017)	(0.018)
SSB_{-G}								0.090**
								(0.039)
SSB_{-O}								0.042
								(0.042)
attended weigh-in 2	0.168*	0.167*	0.188**	0.185*	0.196**	0.180*	0.181*	0.186*
	(0.094)	(0.097)	(0.092)	(0.095)	(0.094)	(0.093)	(0.096)	(0.094)
attended weigh-in 3	0.385***	0.375***	0.369***	0.367***	0.353***	0.374***	0.366***	0.357***
	(0.095)	(0.096)	(0.093)	(0.093)	(0.093)	(0.094)	(0.094)	(0.092)
set weigh-in goal	0.017	0.118	0.028	0.134	0.132	0.018	0.123	0.156*
	(0.088)	(0.095)	(0.086)	(0.093)	(0.092)	(0.086)	(0.093)	(0.093)
female		-0.076		-0.035	-0.015		-0.043	0.010
		(0.105)		(0.104)	(0.104)		(0.104)	(0.105)
single		-0.013		-0.030	-0.018		-0.023	-0.006
		(0.089)		(0.086)	(0.086)		(0.087)	(0.085)
age		-0.009*		-0.010**	-0.011**		-0.010**	-0.011**
		(0.005)		(0.005)	(0.005)		(0.005)	(0.005)
years of educ		0.043		0.051*	0.051*		0.049	0.055*
		(0.030)		(0.030)	(0.030)		(0.030)	(0.030)
prior participant		0.133		0.118	0.124		0.125	0.141
		(0.091)		(0.088)	(0.088)		(0.089)	(0.087)
weight loss goal		-0.000		-0.001	0.000		0.000	0.000
		(0.004)		(0.004)	(0.004)		(0.004)	(0.004)
goal confidence		-0.218		-0.209	-0.207		-0.227	-0.246
		(0.198)		(0.193)	(0.192)		(0.194)	(0.190)
$(goal\ confidence)^2$		0.025		0.024	0.023		0.025	0.027
		(0.021)		(0.020)	(0.020)		(0.020)	(0.020)
diets attempted		-0.012		-0.012	-0.013		-0.013	-0.013
		(0.009)		(0.009)	(0.009)		(0.009)	(0.008)
N	113	113	113	113	113	113	113	113
adj. R^2	0.215	0.235	0.242	0.267	0.276	0.229	0.256	0.288

Notes: Linear probability models where the dependent variable is equal to 1 if the respondent completed the challenge by attending the final weigh-in; the variable *attended weigh-in 2* (*weigh-in 3*) is equal to 1 if the participant attended the intermediate (non compulsory) weigh-in #2 (3), and *set weigh-in goal* is equal to 1 if the participant committed to a weigh-in goal. See Section 3.3 and Table 4 of the main text for a description of the other control variables. Standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

TABLE 24: Determinants of the likelihood of completing the challenge (subsample of participants with preferences $G \succ R$ and $O \succ R$)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
G top	-0.081	-0.175						
	(0.142)	(0.157)						
GO top	-0.123	-0.194						
	(0.115)	(0.120)						
Other top	-0.263*	-0.375**						
	(0.150)	(0.158)						
GT_{-R}			-0.067	-0.083*	-0.103**			
			(0.045)	(0.048)	(0.051)			
GT_{-G}					0.109			
					(0.129)			
GT_{-O}					0.030			
					(0.101)			
SSB_{-R}						-0.022	-0.032	-0.057**
						(0.020)	(0.021)	(0.028)
SSB_{-G}								0.107
								(0.095)
SSB_{-O}								0.059
								(0.059)
attended weigh-in 2	0.159	0.169	0.165	0.166	0.196*	0.158	0.167	0.165
	(0.106)	(0.110)	(0.105)	(0.111)	(0.115)	(0.106)	(0.112)	(0.114)
attended weigh-in 3	0.352***	0.366***	0.340***	0.355***	0.338***	0.343***	0.349***	0.352***
	(0.105)	(0.107)	(0.104)	(0.107)	(0.108)	(0.106)	(0.109)	(0.109)
set weigh-in goal	0.050	0.151	0.022	0.114	0.106	0.029	0.119	0.129
	(0.101)	(0.109)	(0.100)	(0.108)	(0.110)	(0.100)	(0.108)	(0.111)
female		-0.040		-0.044	-0.040		-0.046	-0.013
		(0.123)		(0.120)	(0.124)		(0.121)	(0.125)
single		-0.099		-0.085	-0.070		-0.085	-0.048
		(0.102)		(0.100)	(0.101)		(0.100)	(0.104)
age		-0.012**		-0.011**	-0.013**		-0.011**	-0.012**
		(0.005)		(0.005)	(0.006)		(0.005)	(0.005)
years of educ		0.067*		0.062*	0.067*		0.060*	0.068*
		(0.036)		(0.036)	(0.036)		(0.036)	(0.036)
prior participant		0.204*		0.161	0.164		0.175*	0.175*
		(0.105)		(0.103)	(0.104)		(0.104)	(0.105)
weight loss goal		0.001		-0.000	0.001		0.001	0.000
		(0.005)		(0.004)	(0.005)		(0.005)	(0.005)
goal confidence		-0.256		-0.221	-0.256		-0.239	-0.311
		(0.271)		(0.268)	(0.272)		(0.269)	(0.274)
$(goal\ confidence)^2$		0.032		0.027	0.030		0.028	0.036
		(0.028)		(0.028)	(0.028)		(0.028)	(0.028)
diets attempted		-0.011		-0.013	-0.014		-0.014	-0.016
		(0.010)		(0.010)	(0.010)		(0.011)	(0.011)
N	90	90	90	90	90	90	90	90
adj. R^2	0.181	0.216	0.191	0.207	0.204	0.182	0.200	0.200

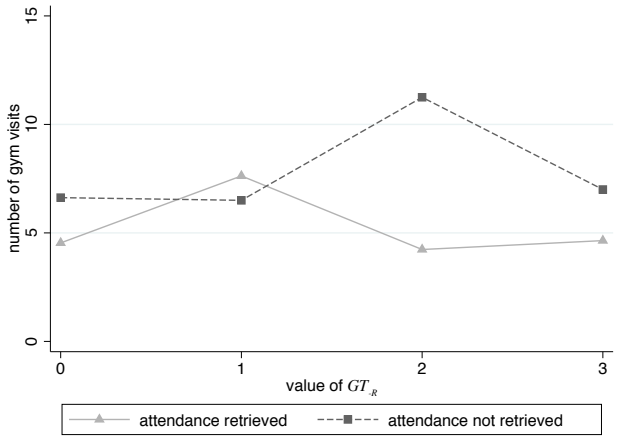
Notes: Linear probability models where the dependent variable is equal to 1 if the respondent completed the challenge by attending the final weigh-in; regressions run on the subset of participants for whom $G \succ R$ and $O \succ R$. The variable *attended weigh-in 2* (*weigh-in 3*) is equal to 1 if the participant attended the intermediate (non compulsory) weigh-in #2 (3), and *set weigh-in goal* is equal to 1 if the participant committed to a weigh-in goal. See Section 3.3 and Table 4 of the main text for a description of the other control variables. Standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

B.4.2. Gym attendance

As explained in the main text, the gym attendance data comes from two sources: 1) gym records; 2) self-reports. Only 69 of the 113 names were retrieved by the computer system of the gym. Figure **15** Panel (a) shows that the estimated number of gym visits tends to be higher for participants whose attendance was not recorded by the gym system; however, differences are only significant for participants with $GT\_R = 2$.

FIGURE 15: Gym visits by value of the GT_R index

(A) Attendance retrieved vs. not retrieved



(B) Goal setters vs. non goal setters

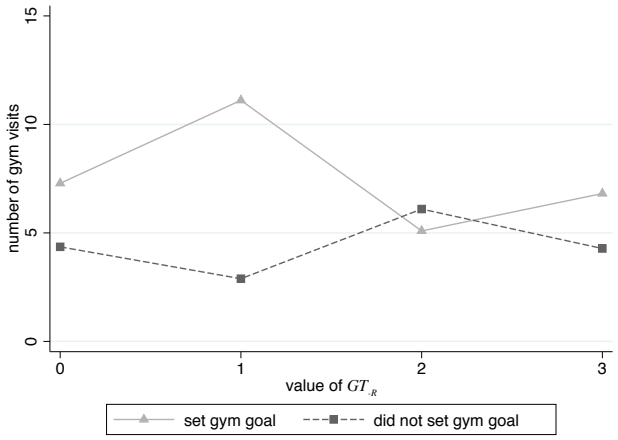


Table **25** compares the individual characteristics of those whose attendance was retrieved vs. not retrieved. Participants differ on essentially one dimension i.e., their past participation in the challenge: those retrieved by the system were less likely to have participated in a previous edition of the challenge than those who were not (13% vs. 59%, $p < 0.001$).[9] All regressions that look at goal achievement control for past participation in the challenge. Tables **20** and **21** in the previous section also present regressions by goal category (exercise vs. weigh-ins/wellness events). For the exercise category, all regressions include an indicator for whether attendance was recorded by the gym system; the coefficient for this indicator is negative but small and insignificant. Most importantly, the relationship between goal achievement and menu preferences does not appear to be driven exclusively by the exercise category, as similar patterns are observed for those who set a weigh-in and/or wellness goal.

[9]One possible reason (which remains a speculation) pertains to the way attendance was recorded in the gym system. Free month memberships had to be manually entered in the system for each participant. Due to the large number of memberships distributed, the gym staff may have accumulated some delays in coding this information and prioritized the entry of first-time users.

As shown in Table **25**, the menu preferences and goal setting decisions of participants whose attendance was retrieved vs. not retrieved do not significantly differ. Figure **15** Panel (b) reveals an interesting interaction effect between goal setting and menu preferences: the higher gym attendance observed among those who set a gym goal compared to those who did not (see B.3) comes from participants who scored low on the GT_{-R} index ($GT_{-R} \in \{0, 1\}$) (9.4 vs. 3.8 visits, $t = -2.83$, $p = 0.008$); by contrast, among those who revealed high temptation by R ($GT_{-R} \in \{2, 3\}$), goal setters did not attend the gym significantly more often than those who did not set a gym goal (6.3 vs. 4.8 visits, $t = 1.0$, $p = 0.322$).[10]

TABLE 25: Comparison of participants retrieved vs. not retrieved by the gym system

Variable	Retrieved	Not retrieved	p-value	Variable	Retrieved	Not retrieved	p-value
female	0.75	0.84	0.273	*G top*	0.14	0.16	0.839
	(0.05)	(0.06)			(0.04)	(0.06)	
single	0.64	0.57	0.465	*GO top*	0.32	0.34	0.810
	(0.06)	(0.08)			(0.06)	(0.07)	
age	34.72	35.32	0.757	*GOR top*	0.30	0.34	0.687
	(1.31)	(1.29)			(0.06)	(0.07)	
years of educ	4.75	5.20	0.144	*Other top*	0.23	0.16	0.353
	(0.18)	(0.27)			(0.05)	(0.06)	
prior participant	0.13	0.59	< 0.001	GT_{-G}	0.42	0.27	0.264
	(0.04)	(0.07)			(0.09)	(0.08)	
weight loss goal	14.49	14.14	0.513	GT_{-O}	0.46	0.52	0.679
	(1.51)	(1.09)			(0.09)	(0.11)	
goal confidence	4.90	4.77	0.578	GT_{-R}	1.96	1.86	0.684
	(0.13)	(0.19)			(0.14)	(0.18)	
diets attempted	3.74	5.00	0.167	SSB_{-G}	0.61	0.27	0.087
	(0.45)	(0.89)			(0.14)	(0.09)	
set gym goal	0.51	0.45	0.589	SSB_{-O}	0.65	0.77	0.548
	(0.06)	(0.08)			(0.13)	(0.16)	
gym goal	4.49	4.07	0.699	SSB_{-R}	3.23	3.18	0.921
	(0.70)	(0.82)			(0.30)	(0.41)	
Observations (Total)	69	44	(113)	Observations (Total)	69	44	(113)

Notes: Reported p-values are from two-sided t-tests; standard errors in brackets. (Not) Retrieved corresponds to participants whose attendance was (not) recorded by the gym system. The variable *gym goal* refers to the number of gym visits a respondent committed to. See Section 3.3 and Table 4 of the main text for a description of the other variables.

B.4.3. Response to Survey 2

In order to complete the study and receive the $20 gift card, participants were required to complete a follow-up survey administered at the end of the challenge. The survey was announced to last 10 to 15 minutes and participants were given one week to complete it. For logistical reasons, the survey was sent one week later than the date announced at the time of enrollment. Of the 113 participants who responded to Survey 1, 87 also completed

[10]The interaction effect $GT_{-R} \geq 2 \times$ *set gym goal* is however not significant in a regression framework. More information is available upon request.

Survey 2 (one additional respondent only partially completed Survey 2 so his response was excluded from the analysis). As shown in Table **26**, the attrition from Survey 1 to Survey 2 is clearly non random: participants who did not respond to Survey 2 were more likely to take up the goal setting contract and set higher goals overall; on the other hand, they were less likely to attend each of the follow up weigh-ins, use the gym and claim reimbursement. Non response is also positively correlated with temptation by R. Tables **27** (entire sample) and **28** (subsample with $G \succ R$ and $O \succ R$) show that menu preferences predict Survey 2 response even after controlling for other individual differences.

TABLE 26: Summary statistics for Survey 2 respondents vs. non respondents

Variable	Responded	Did not respond	p-value	Variable	Responded	Did not respond	p-value
female	0.80	0.73	0.424	*G top*	0.11	0.27	0.054
	(0.04)	(0.09)			(0.03)	(0.09)	
single	0.62	0.50	0.276	*GO top*	0.30	0.42	0.240
	(0.05)	(0.10)			(0.05)	(0.10)	
age	34.43	38.08	0.107	*GOR top*	0.38	0.12	0.011
	(0.99)	(2.47)			(0.05)	(0.06)	
years of educ	4.92	4.96	0.907	*Other top*	0.21	0.19	0.873
	(0.16)	(0.38)			(0.04)	(0.08)	
prior participant	0.34	0.19	0.143	GT_{-G}	0.31	0.54	0.135
	(0.05)	(0.08)			(0.07)	(0.14)	
weight loss goal	12.99	18.92	0.013	GT_{-O}	0.41	0.73	0.053
	(0.70)	(3.61)			(0.08)	(0.15)	
goal confidence	4.84	4.88	0.862	GT_{-R}	1.75	2.50	0.004
	(0.13)	(0.19)			(0.13)	(0.17)	
diets attempted	3.94	5.19	0.237	SSB_{-G}	0.43	0.65	0.318
	(0.50)	(0.94)			(0.12)	(0.16)	
set goal	0.60	0.85	0.019	SSB_{-O}	0.67	0.81	0.544
	(0.05)	(0.07)			(0.12)	(0.18)	
gym goal	3.86	5.88	0.110	SSB_{-R}	2.89	4.31	0.013
	(0.64)	(0.83)			(0.28)	(0.46)	
weigh-in goal	1.32	2.00	0.017	*returned receipts*	0.22	0.00	0.009
	(0.13)	(0.24)			(0.04)	(0.00)	
wellness goal	0.22	0.73	0.002	*attended weigh-in 2*	0.68	0.38	0.007
	(0.06)	(0.22)			(0.30)	(0.41)	
planned gym visits	13.54	10.46	0.134	*attended weigh-in 3*	0.44	0.19	0.024
	(1.02)	(1.50)			(0.05)	(0.08)	
planned weigh-ins	2.84	2.77	0.551	*attended weigh-in 4*	0.46	0.04	0.0001
	(0.05)	(0.14)			(0.05)	(0.04)	
planned wellness events	1.15	1.46	0.228	*actual gym attendance*	6.71	2.65	0.005
	(0.12)	(0.25)			(0.75)	(0.62)	
Observations (Total)	69	44	(113)	Observations (Total)	69	44	(113)

Notes: Reported p-values are from two-sided t-tests; standard errors in brackets. The variable *set goal* is equal to 1 if the respondent took up the goal setting contract, while the variables *gym goal*, *weigh-in goal* and *wellness goal* refer to the number of gym visits, weigh-ins and wellness events a respondent committed to; *planned gym visits*, *planned weigh-ins* and *planned wellness events* refer to the number of times a respondent planned to attend the gym, weigh-ins and wellness events at the start of the challenge (Survey 1 responses). The variable *returned receipts* is equal to 1 if a participant submitted lunch receipts for reimbursement; *actual gym attendance* refers to the number of gym visits a participant made. See Section 3.3 and Table 4 of the main text for a description of the other variables.

TABLE 27: Determinants of the likelihood of completing Survey 2

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
G top	-0.225**	-0.203*						
	(0.107)	(0.115)						
GO top	-0.098	-0.134						
	(0.086)	(0.092)						
Other top	-0.075	-0.070						
	(0.096)	(0.100)						
GT_{-R}			-0.056*	-0.069**	-0.072**			
			(0.030)	(0.032)	(0.033)			
GT_{-G}					-0.096			
					(0.058)			
GT_{-O}					-0.024			
					(0.056)			
SSB_{-R}						-0.023*	-0.029**	-0.025
						(0.013)	(0.014)	(0.016)
SSB_{-G}								-0.041
								(0.035)
SSB_{-O}								-0.006
								(0.036)
achieved goals	0.345***	0.345***	0.364***	0.360***	0.341***	0.372***	0.372***	0.371***
	(0.100)	(0.102)	(0.098)	(0.099)	(0.099)	(0.098)	(0.099)	(0.100)
set no goals	0.337***	0.289***	0.326***	0.270***	0.265***	0.334***	0.281***	0.295***
	(0.082)	(0.091)	(0.081)	(0.089)	(0.089)	(0.080)	(0.089)	(0.091)
attended weigh-in 4	0.183**	0.151*	0.169**	0.134	0.164*	0.172**	0.134	0.154*
	(0.081)	(0.085)	(0.081)	(0.084)	(0.085)	(0.081)	(0.085)	(0.087)
female		0.116		0.165*	0.142		0.163*	0.144
		(0.089)		(0.088)	(0.088)		(0.089)	(0.091)
single		0.067		0.062	0.048		0.068	0.061
		(0.076)		(0.074)	(0.073)		(0.074)	(0.074)
age		0.000		-0.001	0.001		-0.001	-0.001
		(0.004)		(0.004)	(0.004)		(0.004)	(0.004)
years of educ		-0.004		0.001	-0.000		0.001	-0.002
		(0.026)		(0.026)	(0.025)		(0.026)	(0.026)
prior participant		0.091		0.078	0.071		0.082	0.071
		(0.078)		(0.076)	(0.075)		(0.076)	(0.077)
weight loss goal		-0.005		-0.005	-0.005		-0.004	-0.004
		(0.003)		(0.003)	(0.003)		(0.003)	(0.003)
goal confidence		-0.187		-0.206	-0.190		-0.224	-0.211
		(0.169)		(0.166)	(0.164)		(0.166)	(0.167)
$(goal\ confidence)^2$		0.022		0.024	0.024		0.025	0.024
		(0.018)		(0.017)	(0.017)		(0.017)	(0.017)
diets attempted		-0.009		-0.010	-0.009		-0.011	-0.011
		(0.008)		(0.007)	(0.007)		(0.007)	(0.007)
N	113	113	113	113	113	113	113	113
adj. R^2	0.284	0.283	0.290	0.303	0.319	0.287	0.299	0.294

Notes: Linear probability models where the dependent variable is equal to 1 if the participant responded to Survey 2. The variable *achieved goals* is equal to 1 if the participant set goals and achieved all of them, while *set no goals* is equal to 1 if the participant chose not to commit to any goal; the reference category corresponds to participants who failed to achieve at least one of their goals, thus forfeiting their $20 payment. See Section 3.3 and Table 4 of the main text for a description of the other control variables. Standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

TABLE 28: Determinants of the likelihood of completing Survey 2
(subsample of participants with preferences $G \succ R$ and $O \succ R$)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
G top	-0.318**	-0.342**						
	(0.124)	(0.136)						
GO top	-0.165*	-0.257**						
	(0.099)	(0.104)						
Other top	-0.245*	-0.347**						
	(0.129)	(0.137)						
GT_{-R}			-0.050	-0.073*	-0.058			
			(0.039)	(0.042)	(0.044)			
GT_{-G}					-0.123			
					(0.108)			
GT_{-O}					-0.007			
					(0.087)			
SSB_{-R}						-0.023	-0.035*	-0.007
						(0.017)	(0.018)	(0.024)
SSB_{-G}								-0.161**
								(0.081)
SSB_{-O}								-0.027
								(0.050)
achieved goals	0.331***	0.325***	0.361***	0.337***	0.318***	0.369***	0.352***	0.361***
	(0.115)	(0.115)	(0.115)	(0.117)	(0.118)	(0.115)	(0.116)	(0.115)
set no goals	0.276***	0.132	0.307***	0.188*	0.181*	0.302***	0.180*	0.188*
	(0.097)	(0.106)	(0.098)	(0.106)	(0.107)	(0.098)	(0.106)	(0.107)
attended weigh-in 4	0.199**	0.121	0.200**	0.147	0.171	0.198**	0.138	0.153
	(0.097)	(0.099)	(0.099)	(0.101)	(0.103)	(0.099)	(0.102)	(0.101)
female		0.177*		0.199*	0.201*		0.204*	0.183*
		(0.105)		(0.105)	(0.108)		(0.105)	(0.107)
single		0.023		0.057	0.049		0.055	0.012
		(0.087)		(0.087)	(0.088)		(0.087)	(0.088)
age		-0.003		-0.003	-0.001		-0.004	-0.003
		(0.005)		(0.005)	(0.005)		(0.005)	(0.005)
years of educ		0.031		0.021	0.017		0.022	0.012
		(0.031)		(0.031)	(0.031)		(0.031)	(0.031)
prior participant		0.179*		0.111	0.104		0.126	0.109
		(0.091)		(0.091)	(0.091)		(0.091)	(0.091)
weight loss goal		-0.005		-0.006	-0.006		-0.005	-0.004
		(0.004)		(0.004)	(0.004)		(0.004)	(0.004)
goal confidence		-0.416*		-0.420*	-0.376		-0.449*	-0.376
		(0.236)		(0.238)	(0.242)		(0.238)	(0.240)
$(goal\ confidence)^2$		0.049**		0.047*	0.043*		0.050**	0.041*
		(0.024)		(0.025)	(0.025)		(0.025)	(0.025)
diets attempted		-0.013		-0.013	-0.011		-0.014	-0.011
		(0.009)		(0.009)	(0.009)		(0.009)	(0.009)
N	90	90	90	90	90	90	90	90
adj. R^2	0.293	0.335	0.259	0.295	0.296	0.261	0.301	0.319

Notes: Linear probability models where the dependent variable is equal to 1 if the participant responded to Survey 2; regressions run on the subset of participants for whom $G \succ R$ and $O \succ R$. The variable *achieved goals* is equal to 1 if the participant set goals and achieved all of them, while *set no goals* is equal to 1 if the participant chose not to commit to any goal; the reference category corresponds to participants who failed to achieve at least one of their goals, thus forfeiting their $20 payment. See Section 3.3 and Table 4 of the main text for a description of the other control variables. Standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

B.4.4. Submission of lunch receipts

As an additional dimension of participation, below I examine participants' decisions to submit their lunch receipts for reimbursement. Although participants seemed initially interested in the reimbursement program, only 17% (19) ended up submitting receipts. One likely reason for the lack of participation is the logistical burden implied by the reimbursement process. As a reminder, participants were required to submit itemized lunch receipts with their name or credit card number on it. As only originals could be accepted, participants were asked to bring their receipts at the last weigh-in, although they could provide their receipts at a later date if they missed it.[11] Participating in the program thus required a high level of organization, motivation and effort. Furthermore, the benefits were uncertain since only 10% of study participants were randomly selected to be reimbursed and the winners were announced only after the reimbursement period was over (i.e., at the end of the challenge). To investigate whether the high logistical costs/low returns of participating in the program were the main reason for the low participation, respondents to Survey 2 were asked to explain their reason(s) for submitting no or few receipts:

If you brought less than 20 receipts, please indicate the main reason(s):

☐ *I forgot to ask for the receipt on one or several occasions.*
☐ *I lost one or several receipts.*
☐ *I ordered at places which were not giving out detailed receipts.*
☐ *I was frustrated with eating the only things I could be reimbursed for.*
☐ *I thought it was not worth the effort or the investment given that only 10% of respondents would be reimbursed.*
☐ *Other. Indicate:*
[ ]

Among the 87 participants who responded to the above question, 24% (21) mentioned that they forgot to ask for a receipt, 21% (18) that they could not get detailed receipts and 10% (9) that they lost receipts. In total, over 40% mentioned at least one of those 3 logistical reasons for not submitting receipts. Relatedly, 34% (30) mentioned that the costs were not worth the effort given the small chances of being reimbursed. As another important reason, 29% (25) of respondents mentioned that they usually bring their own food and/or rarely eat out.[12] Finally, 13% (11) appeared to have found the reimbursement coverage

[11] Of the 19 participants who submitted receipts, 6 missed the final weigh-in.

[12] In Survey 1, participants were also asked to provide information about their lunch habits; 56% (63) reported bringing their own lunch (either exclusively or at least a few days a week) and another 33% (37) reported buying take-out (either exclusively or at least a few days a week). A minority reported having a meal plan/eating at university cafeterias, or skipping lunch altogether. Those who selected "I bring my own food" were however only slightly less likely to claim reimbursement than others (13.8% vs. 20.0%, $z = 0.88$,

too restrictive, some mentioning that they were not interested in the food options included in their coverage or that they consumed foods not covered by the reimbursement (e.g., sushis).[13] Since participants were randomly assigned a reimbursement option based on their rank ordering, Table **29** tests whether the likelihood of claiming reimbursement depends on the option assigned. Participants were more likely to submit lunch receipts when they received option G, whether G was their preferred option or not; in fact, receiving an option ranked higher does not seem to have influenced the decision to submit receipts.

TABLE 29: Distribution of assigned options for participants who claimed vs. did not claim reimbursement

Option assigned	claimed % (N)	did not claim % (N)	p-value	Selected rank	claimed % (N)	did not claim % (N)	p-value
G	36.8 (7)	17.0 (16)	0.050	1	36.8 (7)	35.1 (33)	0.887
GO	21.05 (4)	24.5 (23)	0.753	2	21.0 (4)	34.0 (32)	0.272
GOR	21.05 (4)	25.5 (24)	0.683	3	31.6 (6)	18.1 (17)	0.186
GR	15.8 (3)	10.6 (10)	0.525	4	5.3 (1)	8.5 (8)	0.637
O	0.0 (0)	10.6 (10)	0.139	5	5.3 (1)	2.1 (2)	0.443
OR	5.3 (1)	7.5 (7)	0.738	6	0.0 (0)	2.1 (2)	0.526
R	0.0 (0)	4.3 (4)	0.364	7	0.0 (0)	0.0 (0)	-
Total	100 (19)	100 (94)		Total	100 (19)	100 (94)	

Notes: Reported p-values are from two-sided t-tests; standard errors in brackets.

The regressions presented in Tables **30** (full sample) and **31** (subsample with $G \succ R$ and $O \succ R$) test whether temptation by R predicts the likelihood of returning receipts, after controlling for participants' individual characteristics, the option they were assigned, and whether they attended the final weigh-in. Participants who ranked GO as their top option were about 20 percentage points less likely to return their receipts; the effect is robust to restricting attention to the subsample of participants who ranked both G and O above R. The $GT\text{-}_R$ index also negatively predicts the likelihood of submitting receipts, although the effect is less robust; the effect of the $SSB\text{-}_R$ index is much weaker and insignificant in most regressions. The main findings are also relatively robust to controlling for the rank number of the option assigned or for whether the participant reported bringing his own lunch (those controls have no predictive power).

$p = 0.378$). The program did not allow participants to be reimbursed for the ingredients of their home-made lunches. An email was sent at the start of the reimbursement period (early April) to clarify this point. Reimbursing home-made food would have likely increased participation, at the cost of making verification more difficult.

[13] In Survey 1, respondents were asked to report any dietary restrictions; about a third listed at least one such restriction.

TABLE 30: Determinants of the likelihood of submitting lunch receipts

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
G top	-0.166	-0.124						
	(0.104)	(0.111)						
GO top	-0.201**	-0.213**						
	(0.083)	(0.087)						
Other top	-0.187*	-0.158						
	(0.094)	(0.097)						
GT_{-R}			-0.071**	-0.063**	-0.069**			
			(0.030)	(0.032)	(0.032)			
GT_{-G}					-0.097			
					(0.058)			
GT_{-O}					0.033			
					(0.056)			
SSB_{-R}						-0.024*	-0.025*	-0.024
						(0.013)	(0.014)	(0.016)
SSB_{-G}								-0.036
								(0.035)
SSB_{-O}								0.012
								(0.036)
assigned G	0.221**	0.220**	0.265***	0.248**	0.237**	0.235**	0.227**	0.211**
	(0.095)	(0.099)	(0.098)	(0.102)	(0.102)	(0.097)	(0.100)	(0.105)
assigned GO	0.074	0.113	0.104	0.125	0.098	0.094	0.118	0.097
	(0.092)	(0.099)	(0.092)	(0.098)	(0.101)	(0.092)	(0.099)	(0.104)
assigned GOR	0.071	0.123	0.080	0.117	0.099	0.078	0.119	0.105
	(0.090)	(0.097)	(0.090)	(0.096)	(0.098)	(0.091)	(0.097)	(0.101)
attended weigh-in 4	0.216***	0.186**	0.211***	0.183**	0.202***	0.222***	0.188**	0.199***
	(0.070)	(0.072)	(0.070)	(0.072)	(0.073)	(0.070)	(0.073)	(0.074)
female		0.018		0.037	0.023		0.035	0.025
		(0.086)		(0.087)	(0.087)		(0.088)	(0.090)
single		0.025		0.002	-0.006		0.006	-0.002
		(0.075)		(0.073)	(0.074)		(0.074)	(0.075)
age		0.002		0.001	0.002		0.001	0.001
		(0.004)		(0.004)	(0.004)		(0.004)	(0.004)
years of educ		-0.015		-0.017	-0.015		-0.016	-0.015
		(0.024)		(0.024)	(0.024)		(0.024)	(0.024)
prior participant		0.117		0.096	0.088		0.102	0.095
		(0.075)		(0.075)	(0.075)		(0.075)	(0.076)
weight loss goal		0.000		-0.000	-0.001		0.000	0.000
		(0.003)		(0.003)	(0.003)		(0.003)	(0.003)
goal confidence		-0.206		-0.213	-0.211		-0.226	-0.218
		(0.160)		(0.159)	(0.159)		(0.160)	(0.161)
$(goal\ confidence)^2$		0.016		0.016	0.017		0.017	0.016
		(0.017)		(0.017)	(0.017)		(0.017)	(0.017)
diets attempted		-0.011		-0.010	-0.010		-0.011	-0.011
		(0.007)		(0.007)	(0.007)		(0.008)	(0.008)
N	113	113	113	113	113	113	113	113
adj. R^2	0.126	0.139	0.131	0.137	0.143	0.112	0.129	0.122

Notes: Linear probability models where the dependent variable is equal to 1 if the participant submitted lunch receipts for reimbursement. The variable *assigned G* (*GO*, *GOR*) is equal to 1 if the participant was assigned reimbursement option *G* (*GO*, *GOR*); the reference category corresponds to participants who received either *O*, *R*, *GR* or *OR*; *attended weigh-in 4* is equal to 1 if the participant attended the last weigh-in. See Section 3.3 and Table 4 of the main text for a description of the other control variables. Standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

TABLE 31: Determinants of the likelihood of submitting lunch receipts (subsample of participants with preferences $G \succ R$ and $O \succ R$)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
G top	-0.153	-0.154						
	(0.114)	(0.132)						
GO top	-0.192**	-0.188*						
	(0.091)	(0.100)						
Other top	-0.202*	-0.167						
	(0.121)	(0.133)						
GT_{-R}			-0.074**	-0.062	-0.048			
			(0.036)	(0.040)	(0.043)			
GT_{-G}					-0.086			
					(0.103)			
GT_{-O}					-0.020			
					(0.082)			
SSB_{-R}						-0.023	-0.020	-0.020
						(0.015)	(0.017)	(0.024)
SSB_{-G}								-0.015
								(0.082)
SSB_{-O}								0.010
								(0.049)
assigned G	0.225**	0.228*	0.232**	0.228*	0.226*	0.207*	0.209*	0.209*
	(0.107)	(0.114)	(0.108)	(0.115)	(0.115)	(0.108)	(0.114)	(0.118)
assigned GO	0.111	0.135	0.095	0.118	0.096	0.093	0.113	0.110
	(0.107)	(0.117)	(0.105)	(0.115)	(0.116)	(0.106)	(0.115)	(0.122)
assigned GOR	0.127	0.167	0.090	0.129	0.112	0.100	0.136	0.134
	(0.104)	(0.115)	(0.101)	(0.112)	(0.113)	(0.103)	(0.113)	(0.117)
attended weigh-in 4	0.248***	0.237***	0.243***	0.238***	0.247***	0.256***	0.245***	0.244***
	(0.078)	(0.084)	(0.078)	(0.084)	(0.084)	(0.079)	(0.084)	(0.085)
female		-0.010		0.003	-0.002		-0.001	0.002
		(0.102)		(0.099)	(0.101)		(0.099)	(0.104)
single		-0.032		-0.032	-0.042		-0.031	-0.033
		(0.085)		(0.082)	(0.083)		(0.083)	(0.087)
age		0.004		0.003	0.005		0.003	0.003
		(0.005)		(0.005)	(0.005)		(0.005)	(0.005)
years of educ		-0.038		-0.039	-0.041		-0.041	-0.042
		(0.029)		(0.029)	(0.029)		(0.029)	(0.029)
prior participant		0.055		0.027	0.024		0.034	0.031
		(0.089)		(0.086)	(0.087)		(0.087)	(0.089)
weight loss goal		-0.001		-0.002	-0.002		-0.002	-0.002
		(0.004)		(0.004)	(0.004)		(0.004)	(0.004)
goal confidence		-0.007		-0.020	0.009		-0.029	-0.028
		(0.220)		(0.216)	(0.218)		(0.217)	(0.222)
$(goal\ confidence)^2$		-0.004		-0.003	-0.006		-0.003	-0.003
		(0.022)		(0.022)	(0.022)		(0.022)	(0.023)
diets attempted		-0.010		-0.010	-0.008		-0.010	-0.010
		(0.009)		(0.009)	(0.009)		(0.009)	(0.009)
N	90	90	90	90	90	90	90	90
adj. R^2	0.137	0.105	0.147	0.112	0.106	0.127	0.101	0.078

Notes: Linear probability models where the dependent variable is equal to 1 if the participant submitted lunch receipts for reimbursement; regressions run on the subset of participants for whom $G \succ R$ and $O \succ R$. The indicator *assigned G* (*GO*, *GOR*) is equal to 1 if the participant was assigned reimbursement option *G* (*GO*, *GOR*); the reference category corresponds to participants who received either *O*, *R*, *GR* or *OR*; *attended weigh-in 4* is equal to 1 if the participant attended the last weigh-in. See Section 3.3 and Table 4 of the main text for a description of the other control variables. Standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

C. Link between commitment and Survey 2 measures of self-control

In this section, I present results from a previous version of the paper, which looked at the relationship between menu preference variables, goal setting and standard measures of self-control used in the literature: (i) present bias over time-dated monetary rewards measured through a Multiple Price List mechanism; (ii) the self-control measures developed by Ameriks et al. (2007), which rely on survey answers to an hypothetical intertemporal consumption problem. Data on (i) & (ii) was collected on the sample of participants who replied to Survey 2 ($N = 87$). In this smaller and selected sample, the link between standard measures of self-control and commitment demand in the experiment is at best weak and is presented here for the sake of completeness and transparency.

C.1 Present bias in monetary choices

One popular measure is whether an individual exhibits preference reversals in intertemporal choices, preferring earlier (later) rewards as the early date is moved closer to (away from) the present. Models of dynamically inconsistent time preferences predict a positive relationship between sophisticated present bias and desire for commitment. Yet, there is so far limited evidence of a positive relationship between present bias and commitment.[14] Since alternative models such as Gul and Pesendorfer (2001) predict that commitment choices may arise even in the absence of preference reversals, it is important to further investigate this relationship.

In Survey 2, time preferences were measured through intertemporal choices over gift cards used for purchases at a food store. Respondents were presented with two series of 5 choices between a \$100 gift card at time t and a $\$X$ gift card at time t + one week where t was either today or in 4 weeks and $X \in \{98, 102, 105, 108, 110\}$. Let X_1 and X_2 be the minimum gift card amounts such that a respondent prefers X_1 in one week (resp. X_2 in 5 weeks) to \$100 today (in 4 weeks). A respondent is classified as *present-biased* if $X_1 > X_2$, implying a higher patience in the future compared to now. Similarly, a respondent is said to be *future-biased* if $X_1 < X_2$ (higher patience today than in the future) and time consistent if $X_1 = X_2$. Finally, a time consistent respondent is classified as *patient* if he prefers the later payment whenever

[14] In the context of intertemporal choices over money, Casari (2009) finds that when commitment is free, 61.5% of subjects who exhibit choice reversals prefer commitment. However, the paper does not report the take-up rate among individuals who do not reverse their choice; therefore, one cannot conclude from this study that present-biased agents are more likely to commit. Ashraf et al. (2006) find a positive and significant relationship between time inconsistency in hypothetical monetary choices and take-up of their commitment savings product, but only among women. The most convincing piece of evidence comes from Augenblick et al. (2015) in the context of effort choices. They show that individuals who prefer to be committed to early effort are also more likely to be present-biased in effort; however, they find that present bias in monetary choices is not a significant predictor of commitment in effort choices.

$X > 100$, and *impatient* if he always prefers the earlier \$100 payment to the later payment X; the category *other* includes all respondents who either switch between the earlier and later payment at some $X \neq 100$, or always prefer the late payment.[15] Table **32** presents the distribution of time preferences among the population of respondents who replied to Survey 2. All but two respondents switched no more than once between the sooner and the later dates in both price lists; these non monotonic respondents are excluded from the analysis.

TABLE 32: Distribution of time preferences

	% of participants	(N)
Time inconsistent	**24.7**	**(21)**
Present-biased	10.6	(9)
Future-biased	14.1	(12)
Time consistent	**75.3**	**(64)**
Impatient	16.5	(14)
Patient	48.2	(41)
other	10.6	(9)
Total	100.0	(85)

# times the early reward was chosen	mean (s.d.)	min-max
	4.3 (3.2)	0-10

Overall, 10.6% of respondents are classified as present-biased; with a slightly larger category of future-biased respondents, about 25% of the sample exhibit time inconsistent choices. Suggested explanations for future bias include survey noise, anticipatory utility or future uncertainty (Ashraf et al. (2006), Loewenstein (1987), Ameriks et al. (2007), Takeuchi (2011)). However, the vast majority of respondents exhibit time consistent behavior, with 48% of individuals classified as patient. The proportion of time inconsistent individuals is somewhat below what is commonly found in the literature using similar elicitation techniques. For instance, Ashraf et al. (2006) find 27.5% of present-biased and 19.8% of future-biased individuals, Meier and Sprenger (2010) find 36% (9%) of present (future) -biased individuals; the most comparable estimates are those of Andreoni and Sprenger (2012) who find that 16.7% (10.7%) exhibit present (future) bias with their Multiple Price List procedure.[16]

[15] For the exact breakdown of the *other* category, 4 respondents always chose the later payment, 3 switched from early to late at $X = 108$, while the last two respondents switched at $X = 105$ and $X = 110$.

[16] Two additional studies using a similar elicitation procedure find a low fraction of present-biased individuals: John (2019) finds that 16.6% of the individuals in her sample are present-biased and 18.9% are

Four reasons might account for these low time inconsistency estimates. First, although the two multiple price lists were separated from each other by a set of unrelated questions, the temporal separation between the two sets of questions was short and respondents might have anchored their answers in the second block on their choices in the first block. Second, the temporal distance between the earlier and later rewards was set to one week, while it is typically one month in the literature. This difference could have translated into more patience and less present bias if respondents treated one week delays as the near present. Third, several papers show that once controlling for transaction costs and payment reliability (as this study did), present bias tends to disappear (Andreoni and Sprenger (2012), Giné et al. (2016), Andersen et al. (2014)). Finally, as discussed in Section **B.4.3** (see Table **26**), the population of participants who replied to Survey 2 is not random: non respondents were more likely to have set a goal in Survey 1 (84% vs. 60%, $p = 0.02$), and less likely to have attended the gym (2.6 visits vs. 6.9 visits, $p < 0.01$) or the last weigh-in (4% vs. 45%, $p < 0.01$).[17] They were also more likely to have ranked G as their unique top option (28% vs. 11%, $p = 0.04$) and less likely to have selected GOR as their unique top option (12% vs. 38%, $p = 0.02$). If present bias is positively correlated with commitment demand in the reimbursement program and/or the ability to stick to one's goals, then sample selection will lead to a downward bias on the estimated proportion of present-biased participants in the overall sample. Although the importance of selection bias relative to the other three factors remains unclear, it is important to stress that the next set of results concern a selected sample of respondents.

With this caveat in mind, I now examine the relationship between time preference and commitment demand among Survey 2 respondents. For this purpose, I consider 5 commitment variables: whether the respondent selected G or GO as his unique top option in the reimbursement program, his GT_{-R} and SSB_{-R} scores, and whether he took up the goal setting contract. Time preference is measured through preference reversals, treating separately present and futur bias, and through the number of times a respondent preferred the early reward (out of 10 intertemporal trade-offs). As shown in Table **33**, being classified as present-biased is associated with a 24.1 percentage point increase in the likelihood of preferring commitment to G :on a base of 11.8%. Present bias is not a significant predictor of any other measure of commitment demand. As expected, no relationship exists between demand for commitment and preference reversals in the form of a future bias. While goal setting and strict preference for GO are not linked to preference reversals, they are correlated with preference for the early reward: those who choose the early reward more frequently are less likely to prefer GO and more likely to have set a goal.

future-biased, while Kaur et al. (2015) find that 17% of their surveyed workers exhibit preference reversals in the direction of a present bias (they do not report the fraction of future-biased respondents).

[17] One consequence of these lower attendance rates is that they were less likely to have reached their gym goal (17% vs. 62%, $p < 0.01$) or their weigh-in goal (11% vs. 63%, $p < 0.01$).

TABLE 33: Relationship between time preference and commitment decisions

	Dependent variable				
	G first	*GO first*	GT_{-R}	SSB_{-R}	*set a goal*
Panel A: *Present-biased* (=1)	0.241**	-0.094	0.237	0.798	0.088
	(0.112)	(0.164)	(0.420)	(0.910)	(0.175)
R^2	0.053	0.004	0.004	0.009	0.003
Panel B: *Future-biased* (=1)	-0.040	0.032	0.441	0.055	-0.006
	(0.101)	(0.145)	(0.369)	(0.808)	(0.155)
R^2	0.002	0.001	0.017	0.0001	0.000
Panel C: *# times preferred early reward* (/10)	0.011	-0.031**	-0.021	0.012	0.030*
	(0.068)	(0.016)	(0.041)	(0.089)	(0.017)
R^2	0.013	0.046	0.003	0.0002	0.038
Observations	85	85	85	85	85
Mean dependent variable	0.118	0.306	1.788	2.953	0.588

Notes: Results from *OLS* regressions of commitment variables on an indicator =1 if a respondent is present-biased (Panel A), future-biased (Panel B) and on a variable for the number of times a respondent preferred the early reward (Panel C). Standard errors are in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

While there is no clear explanation for these results, it is worth noting that several motives could lead respondents to favor the early reward: for instance, they could be truly impatient to redeem their card, but they could also perceive the card as an additional incentive to eat healthy while they are still motivated to lose weight. In any case, more research is needed to understand the link between intertemporal choices and commitment decisions.

C.2 The self-control measures of Ameriks et al. (2007)

As an alternative way of measuring self-control problems, Ameriks et al. (2007) (referred to as ACLT subsequently) develop a survey instrument inspired by the model of Gul and Pesendorfer (2001), which is based on a hypothetical choice scenario. In a sample of highly educated adults, they find that their instrument is strongly correlated with wealth accumulation as well as with standard measures of conscientiousness developed in the psychology literature. One natural question to ask is whether their survey instrument predicts commitment decisions in the reimbursement program or through goal setting. At the end of Survey 2, respondents were presented with the hypothetical problem of allocating 10 restaurant certificates between the current and the following year and were asked:

1. How many of the 10 certificates they would ideally use in year 1 as opposed to year 2.
2. Whether they would be tempted to depart from their ideal allocation and in which direction.
3. How many certificates they would use in year 1 if they were to give in to their temptation.
4. How many certificates they think they would end up using in year 1.

In addition, participants who replied being tempted to use more certificates in year 1 (2) were asked how many certificates they would restrict for use in year 1 (2) if they could. This final question is intended to measure respondents' desire to constrain their future choices i.e., to exercise commitment. Let X_i be their answer to Question $i \in \{1, 3, 4\}$. ACLT measured *Perceived Self-Control* with the Expected-Ideal (EI) gap $X_4 - X_1$ and *Perceived Temptation* with the Temptation-Ideal (TI) gap $X_3 - X_1$. To see how the two measures relate to the model of Gul and Pesendorfer (2001), let x_{U+V} denote the solution to the problem

$$\max_{x_1 \in \{0,1,\ldots,10\}} u(x_1, x_2) + v(x_1, x_2) \text{ subject to } x_1 + x_2 = 10$$

where u and v refer to the commitment and temptation utilities in their model, and x_i denotes the number of certificates allocated to year i. Similarly define x_U and x_V. Perceived Self-Control is then measured as $x_{U+V} - x_U$ and Perceived Temptation as $x_V - x_U$.

Consistent with ACLT, the EI gap is typically small: for close to 90% of respondents, the EI gap is less than two in absolute value and is equal to zero for about 69% of respondents who report no self-control problems. The corresponding numbers in ACLT are 95% and 65%. However the ideal allocations differ between the two studies: while the equal split was the ideal allocation for almost 60% of respondents in ACLT, the modal response in this

TABLE 34: Distribution of the EI and TI gaps

	EI gap				TI gap			
	All		Restricted		All		Restricted	
	%	*N*	%	*N*	%	*N*	%	*N*
-6	1.1	(1)	1.5	(1)	0.0	(0)	0.0	(0)
-5	3.5	(3)	3.0	(2)	2.3	(2)	1.5	(1)
-4	2.3	(2)	0.0	(0)	1.1	(1)	0.0	(0)
-3	1.1	(1)	0.0	(0)	1.1	(1)	1.5	(1)
-2	9.2	(8)	9.1	(6)	6.9	(6)	1.5	(1)
-1	5.8	(5)	7.6	(5)	0.0	(0)	0.0	(0)
0	69.0	(60)	69.7	(46)	83.9	(73)	89.4	(59)
1	5.8	(5)	6.1	(4)	0.0	(0)	0.0	(0)
2	0.0	(0)	0.0	(0)	1.1	(1)	1.5	(1)
3	1.1	(1)	1.5	(1)	3.5	(3)	4.6	(3)
6	1.1	(1)	1.5	(1)	0.0	(0)	0.0	(0)
Total	100.0	(87)	100.0	(66)	100.0	(87)	100.0	(66)

Notes: The restricted sample corresponds to the subset of respondents who were consistent in their answers to Questions 2 and 3 above.

study was allocating all certificates in year 1 (44% of the sample), followed by the equal split (about 30% of respondents). The measure of the TI gap is quite noisy, as about 24% of respondents were inconsistent in their answers to Questions 2 and 3. Among those who provided consistent answers (66 respondents), the correlation between the EI and TI gaps is 0.46, a number almost identical to that of ACLT; however the average TI gap is smaller with about 89% of respondents reporting a TI gap of zero. Table **34** summarizes these findings both for the full sample and for the subsample of respondents who were consistent in their answers to Questions 2 and 3.

Table **35** shows the relationship between the ACLT self-control measures and commitment (versus flexibility) decisions in the experiment. The EI and TI gaps are not predictive of commitment demand, but the TI gap predicts preference for flexibility in the reimbursement program. In particular, an increase by one unit of the TI gap leads to a significant 12 percentage point decrease in the likelihood of strictly preferring the unrestricted coverage *GOR* (Panel B).

Finally 28 participants - those who replied that they would be tempted to deviate from their ideal allocation - were asked whether they would restrict the number of certificates for use in year 1 (or 2). More than 80% of them chose to restrict a positive amount of certificates, with an average restriction of 5 certificates (see Table **36** below). Interestingly, those who restricted a higher number of certificates were also more (less) likely to exhibit a preference

TABLE 35: Relationship between commitment, flexibility and the ACLT measures

	Dependent variable					
	G first	*GO first*	*GOR first*	GT_{-R}	SSB_{-R}	*set goal*
Panel A: *Expected - Ideal gap*	-0.020	0.021	-0.051	0.050	0.088	-0.039
	(0.022)	(0.031)	(0.033)	(0.081)	(0.174)	(0.033)
Observations	87	87	87	87	87	87
R^2	0.010	0.005	0.028	0.005	0.003	0.016
Mean dependent variable	0.115	0.299	0.379	1.747	2.885	0.598
Panel B: *Temptation - Ideal gap*	0.042	0.039	-0.121**	0.089	0.420	-0.052
	(0.037)	(0.057)	(0.056)	(0.144)	(0.307)	(0.059)
Observations	66	66	66	66	66	66
R^2	0.019	0.007	0.069	0.006	0.028	0.012
Mean dependent variable	0.106	0.333	0.333	1.864	3.045	0.606
Panel C: *# certificates restricted*	0.053***	0.016	-0.082***	0.064	0.248	-0.004
	(0.018)	(0.023)	(0.026)	(0.066)	(0.148)	(0.030)
Observations	28	28	28	28	28	28
R^2	0.245	0.018	0.283	0.035	0.098	0.0006
Mean dependent variable	0.143	0.179	0.536	1.536	2.571	0.536

Notes: Results from *OLS* regressions of commitment (flexibility) decisions on the temptation and self-control measures constructed from Ameriks et al. (2007). Standard errors are in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

for commitment (flexibility) in the reimbursement program. In particular, the restriction of one additional certificate corresponds to a significant 8.2 percentage point decrease in the probability of strictly preferring the unrestricted coverage GOR and to a 5.3 percentage point increase in the probability of strictly preferring full commitment to G (Panel C of Table **35**). Although it is not clear how to interpret decisions to restrict certificates for use during a specific year, these findings may capture respondents' willingness (aversion) to impose restrictions on themselves.

TABLE 36: Distribution of answers to the ACLT commitment question

# certificates restricted (year 1 and 2 combined)	% respondents	N
0	17.9	(5)
1	0.0	(0)
2	0.0	(0)
3	7.1	(2)
4	17.9	(5)
5	25.0	(7)
6	0.0	(0)
7	3.6	(1)
8	10.7	(3)
9	0.0	(0)
10	17.9	(5)
Total	100.0	(28)

Notes: Distribution of answers to the question taken from Ameriks et al. (2007): "Suppose you had the option to restrict the number of certificates for use in year 1 [year 2]. How many of the certificates would you restrict? Please enter a number between 0 (no restriction) and 10 (all certificates used in year 1 [year 2])." The question was asked to the 28 respondents who replied that they would be tempted to consume more [less] than their ideal allocation of restaurant certificates in year 1.

D. More on the weight loss challenge

D.1 Details about the rules, support system and incentives

In this section, I provide more details regarding the rules, incentives and support system surrounding the yearly weight loss challenge organized at New York University.

Enrollment criteria This challenge called "Lose Smart" was first launched by the NYU wellness services in 2011; the 2014 challenge was therefore the fourth edition of this weight loss contest. Only NYU faculty and staff members could participate in this challenge. Pre-registration happened through a secured website. There were no other enrollment criteria besides those of belonging to the NYU professional community and having pre-registered online; in particular, enrollment was not made conditional on any prior medical information that would document a need to lose weight. Instead, participants were required to sign a waiver to acknowledge the risks of entering a weight loss contest.

Rules and prizes The challenge rules and prizes were posted on the website of the wellness services at the time of registration and were recorded on the weigh-in card. The grand prize winners were those who lost on the last weigh-in day of April 29th the highest percentage of their original reference weight as of March 4th. Participants were told that in case of a tie, the person who lost the most weight would be deemed the winner. Winners of the contest were announced at an award ceremony one week after the end of the challenge (May 6th). The grand prize for each gender category was a fitness wristband. To be able to compete for the grand prize, participants were only required to attend the first and the last weigh-in. In addition to the grand prize, the participant who had lost the most weight (in lbs) between any two weigh-in dates received a gift certificate at an organic food store.

Weight records The weights were recorded by a staff member of the private gym club that was partner of the challenge. The weigh-ins took place in the basement of the gym club, out of the sight of the regular members. The weight records were kept by the gym in order to maintain strict confidentiality of the data. After each weigh-in, the gym staff transmitted the list of the top 10 losers to the NYU wellness services. This list was emailed to all participants but only the first name and initial of the last name appeared on the list; no other information (such as the amount of weight lost by the top 10 losers) was conveyed to participants. At the conclusion of the challenge, the wellness services only requested the names of the two winners (one male, one female) as well as aggregate weight loss data.

Weigh-ins The list of weigh-in dates was announced at the time of registration and recorded on a weigh-in card that participants could use to keep track of their progress. The weigh-in card was distributed at the first weigh-in (see card below). Due to Spring break, the weigh-ins were scheduled either two or three weeks apart (March 4th, March 25th, April

15th and April 29th). Otherwise, they were all scheduled on the same week day (Tuesday) and at the same time of the day (11 am - 1 pm). Participants were not allowed to weigh in outside of those times and the schedule was tighly followed. Reminders were sent before each weigh-in day.

Support system Participants benefited from an extensive support system during the challenge. First, the gym club partner of the challenge offered a complimentary one month membership to all participants who signed up for it at the first weigh-in. The membership included free access to the basic gym facilities as well as to a large set of classes. Participants could also purchase personal training sessions at a highly discounted price. Secondly, participants could sign up for up to four wellness events scheduled by the wellness services during the challenge (two nutrition seminars and two cardio classes); these events were advertised on the weigh-in card distributed to participants. Finally, participants were invited to join informal yoga classes or weight loss meetings with a group leader.

Prior participants Because the challenge was conducted every year since 2011, some participants in the 2014 edition had already participated in at least one of the previous editions. Among the 113 study participants, 31% (35) had previously taken part in the challenge: 18% (20) once, 10% (11) twice and 3% (4) in all 3 editions. The 2013 edition took place at the same gym club than in 2014, but the first two editions (2011-2012) were conducted in partnership with a different health club.

Previous editions Prior editions of the challenge differed in a few rules. As participants in prior editions might have been influenced by the former set of rules, I highlight the changes that were made in 2014. Before 2014, weigh-ins occurred on a weekly basis and participants were required to attend at least 7 of the 8 scheduled weigh-ins in order to stay in the challenge. Due to the low retention rates of the previous years, the 2014 version of the challenge only required participants to attend the first and last weigh-ins in order to stay in the contest and fewer weigh-ins were scheduled.[18] Secondly, the structure of the competition was modified by having 2014 participants compete for the grand prize only against their own gender. Other minor differences with previous editions include a change in the weekly and final prizes and the change in gym partnership between 2011-2012 and 2013-2014.

[18] In 2011, 2012 and 2013, the proportion of initial participants who were still in the challenge after 8 weeks was 20%, 36% and 18% respectively. The corresponding number for 2014 was 23% (44 of 193 participants overall) and 36% (41/113) of the study participants.

LiveSmart

NYU "LOSE SMART" CHALLENGE WEIGH-IN CARD

Congratulations on taking the first step to a healthier you in 2014! NYU's "Lose Smart" Challenge is a special two-month long weight-loss challenge offered through NYU's LiveSmart wellness program. Similar to NBC's "Biggest Loser" program, NYU's "Lose Smart" Challenge will help motivate and encourage participants to achieve their weight loss goals with the support – and competition – from their colleagues.

Challenge Rules

1. All registrants will receive a complimentary one month membership to the Astor Place David Barton Gym for the month of March. Registration for the free month must be done by Friday, March 7.
2. Weigh-in at David Barton Gym, 4 Astor Place, with initial weigh-in on March 4
   - Dates: March 4, March 25, April 15, April 29
   - Time: 11:00 am—1:00 pm

You do not need to be a member of the David Barton Gym to weigh in or participate in the challenge.

- The Grand Prize winners of the "Lose Smart" Challenge will be the participants who loses the highest percentage of their original "weigh-in weight" as of March 4, 2014.
- In the case of a tie, the person who has lost the most weight will be deemed the winner.
- One grand prize of fitness bands will be awarded to the top male and the top female participants.

- Top weigh-in date winner will receive a Whole Foods gift certificate.
- Discounted personal training sessions will be available for purchase.

You can use this weigh-in card for your personal source of information to keep track of your progress.

Weigh-In	*Weight*
Initial Weight (Mar. 4th)	
March 25th	
April 15th	
April 29th	

To help you stay on track, take advantage of the following exercise classes:

- **3/7**
  Total Body Blast
  (registration is available through iLearn)
- **4/4**
  Cardio Bollywood
  (registration is available through iLearn)
- **March & April**
  Friday Yoga (12:30 pm; 238 Thompson St, Room 472)

To learn more about nutrition, attend one of the following workshops or webinars:

- **3/11**
  Aging Gracefully
- **3/13**
  Current Food Trends
  (register through iLearn)
- **Fridays**
  Weight Watchers at Work
  (12 pm;194 Mercer Street, 4th Fl Conference Room)
- **4/3**
  Build the Perfect Shopping List for Weight Mgmt
  (register through iLearn)

Links to healthy eating websites:

- http://www.eatingwell.com
- http://www.cookinglight.com
- http://www.foodnetwork.com/healthy-eating/index.html
- http://www.mayoclinic.org/healthy-lifestyle/recipes
- http://www.eatright.org/NNM/#

Research Study

Join a study on improving health through exercise and nutrition and receive a $20 Whole Foods gift card plus the chance to be reimbursed for your lunch meals in April! ($300 value).

Link to the study: https://nyu.qualtrics.com/SE/?SID=SV_abl1vlXK9x4jOrr

D.3 Flyer distributed at the first weigh-in

Join a study on improving health through exercise and nutrition and receive a $20 Whole Foods gift card plus the chance to be reimbursed for your lunch meals in April!

This study is conducted in collaboration with the NYU Department of Economics and the program LiveSmart.

Interested?

Sign up at:
https://nyu.qualtrics.com/SE/?SID=SV_abl1vIXK9x4jOrr

Any Questions?

Contact: Séverine Toussaert at st1445@nyu.edu

References

Ameriks, J., A. Caplin, J. Leahy, and T. Tyler (2007): "Measuring Self-Control Problems," *American Economic Review*, 97, 966–972.

Andersen, S., G. W. Harrison, M. I. Lau, and E. E. Rutström (2014): "Discounting behavior: A reconsideration," *European Economic Review*, 71, 15 – 33.

Andreoni, J. and C. Sprenger (2012): "Estimating Time Preferences from Convex Budgets," *American Economic Review*, 102, 3333–56.

Ashraf, N., D. Karlan, and W. Yin (2006): "Tying Odysseus to the Mast: Evidence From a Commitment Savings Product in the Philippines," *Quarterly Journal of Economics*, 121, 635–672.

Augenblick, N., M. Niederle, and C. Sprenger (2015): "Working Over Time: Dynamic Inconsistency in Real Effort Tasks*," *Quarterly Journal of Economics*.

Casari, M. (2009): "Pre-commitment and flexibility in a time decision experiment," *Journal of Risk and Uncertainty*, 38, 117–141.

Giné, X., J. Goldberg, D. Silverman, and D. Yang (2016): "Revising Commitments: Field Evidence on the Adjustment of Prior Choices," *Economic Journal*, 128, 159–188.

Gul, F. and W. Pesendorfer (2001): "Temptation and Self-Control," *Econometrica*, 69, 1403–1435.

John, A. (2019): "When Commitment Fails - Evidence from a Field Experiment," *Management Science*, forthcoming.

Kaur, S., M. Kremer, and S. Mullainathan (2015): "Self-Control at Work," *Journal of Political Economy*, 123, 1227–1277.

Loewenstein, G. (1987): "Anticipation and the Valuation of Delayed Consumption," *Economic Journal*, 97, 666–684.

Meier, S. and C. Sprenger (2010): "Present-Biased Preferences and Credit Card Borrowing," *American Economic Journal: Applied Economics*, 2, 193–210.

Takeuchi, K. (2011): "Non-parametric Test of Time Consistency: Present Bias and Future Bias," *Games and Economic Behavior*, 71, 456–478.