Supplementary Information

1 Experimental Material

Stimuli

Physical Condition

Figure S1 shows the snack foods that were used in the physical condition. Snack foods were comprised of pairs of similar items (e.g. two chocolate bars or two licorice snacks). One snack food of each pair was presented in the weight condition, the other one in the condition without weight. This way, it was assured that there were the same kind of snack foods present in both conditions. Which item of each pair was presented in which weight condition was randomized across participants.

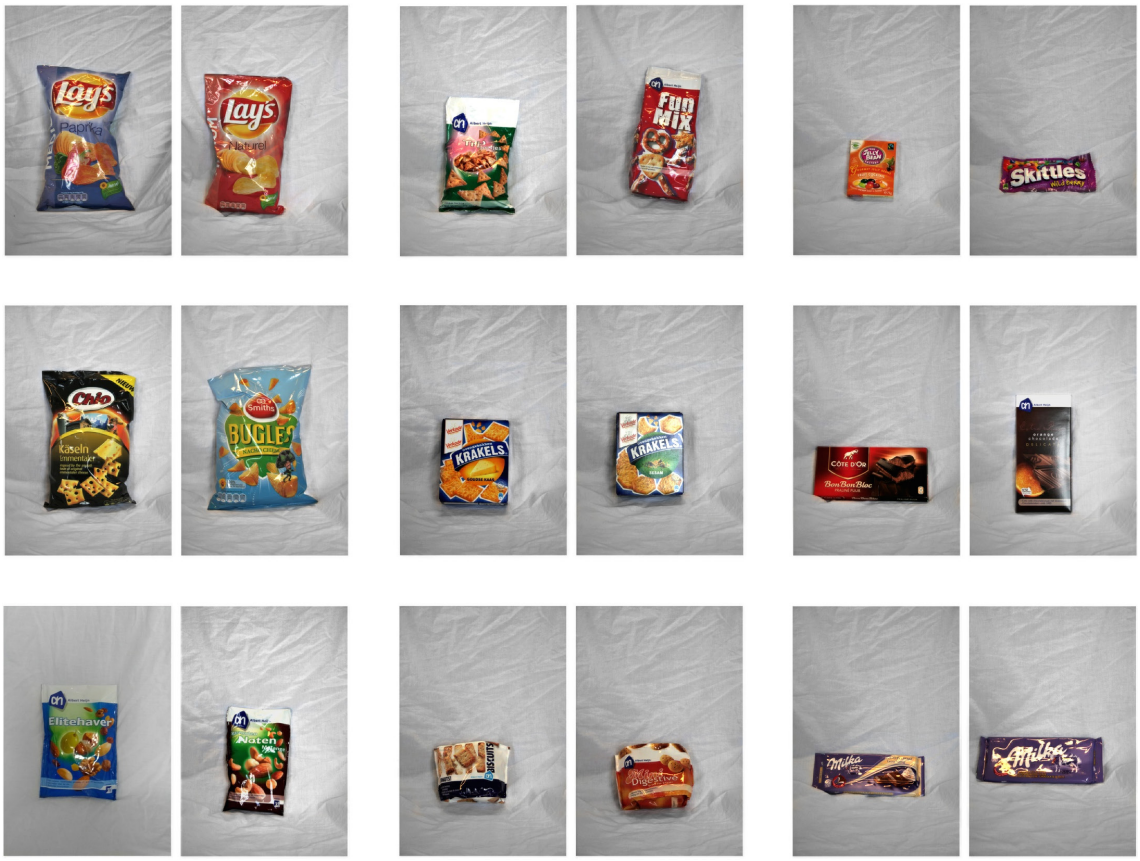

Figure S1. Pictures of the snack food items used in the physical condition. Snack foods near to each other comprise pairs. Elements of a pair were never presented in the same condition within one participant. One was presented in the weight, the other one in the no weight condition. Which item was presented in which condition was randomized across participants.

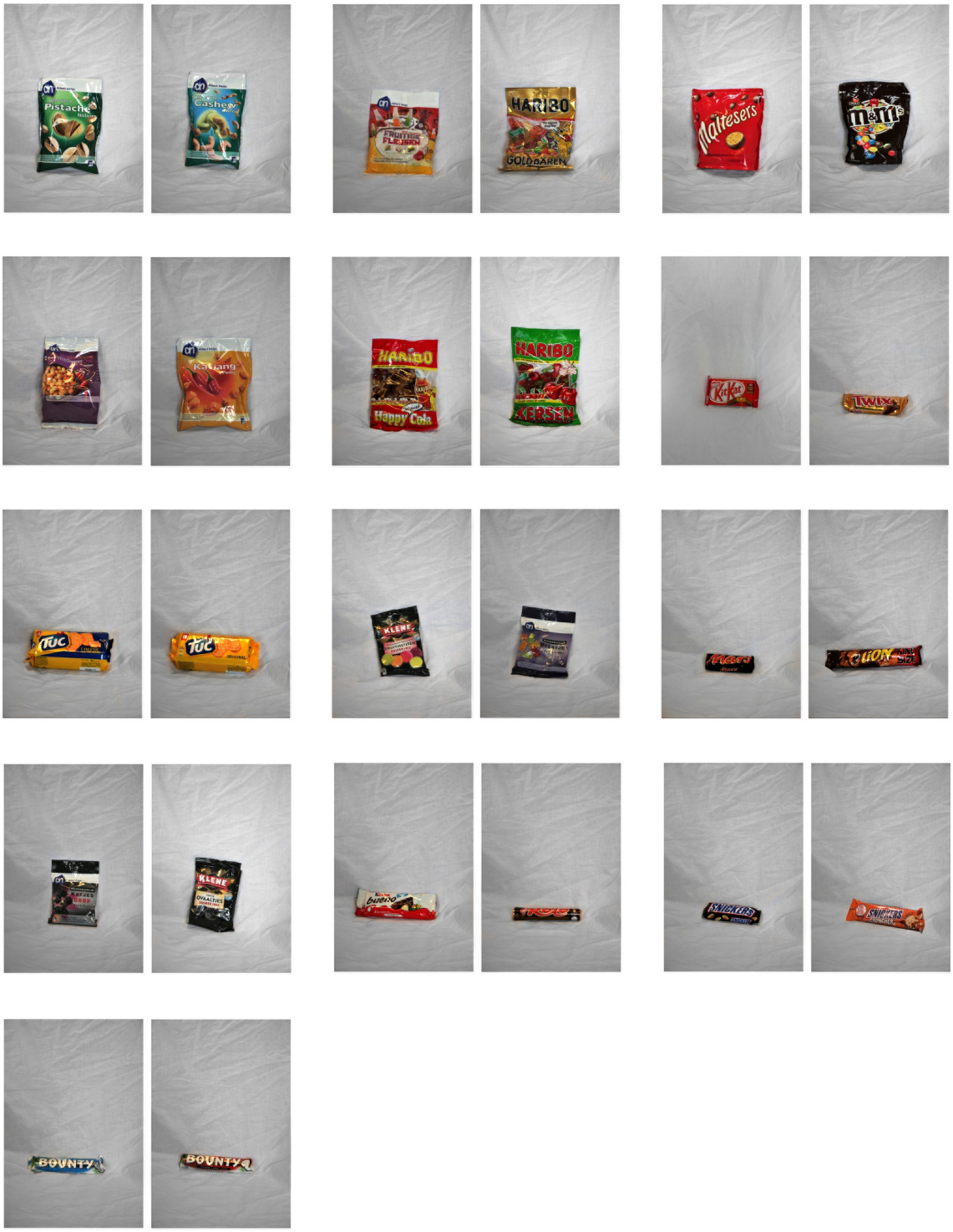

Figure S1 (continued). Pictures of the snack food items used in the physical condition. Snack foods near to each other comprise pairs. Elements of a pair were never presented in the same condition within one participant. One was presented in the weight, the other one in the no weight condition. Which item was presented in which condition was randomized across participants.

Computer Condition

Table S1 shows the snack foods presented in the computer condition. Each item was presented as text on the computer screen and accompanied by a short description shown in parentheses in Table S1.

Table S1

Stimuli used in the computer condition. Names and descriptions (in parentheses) were presented together on the computer screen. Snack foods in the same row comprise pairs.

Name	**Name**
Digestive Mini's Milk Chocolate* (Cookies)	Choc Chip Mini's* (Cookies)
Salty Liquorice (Liquorice)	Sweet Liquorice (Liquorice)
Soft Fruitbears (Fruitgum)	Fruity Winegums (Fruitgum)
Haribo Happy Cola (Fruitgum)	Haribo Happy Cherries (Fruitgum)
Haribo Gummibears (Fruitgum)	Fruity Bottles (Fruitgum)
Jelly Beans (Fruit Candy)	Skittles (Fruit Candy)
Bounty (Chocolate Bar)	Bounty Dark (Chocolate Bar)
Cote d'or BonBonBloc Praline (Chocolate)	Bitter Chocolate with Orange (Chocolate)
Kinder Bueno (Chocolate Bar)	Rolo (Chocolates with Caramel)
KitKat (Chocolate Bar)	KitKat Chunky* (Chocolate Bar)
Choco M&Ms (Chocolate Candy)	Maltesers (Chocolate Candy)

Note. *These items were not identical but close substitutes to items in the physical condition, all other items were identical across the two conditions.

Table S1 (continued)

Stimuli used in the computer condition. Names and descriptions (in parentheses) were presented together on the computer screen. Snack foods in the same row comprise pairs.

Name	**Name**
Mars (Chocolate Bar)	Snickers (Chocolate Bar)
Milka Milk (Chocolate)	Milka Cream (Chocolate)
Twix (Chocolate Bar)	Lion (Chocolate Bar)
Bugles Cheese (Nacho Chips)	Chio Cheese (Potato Chips)
Lay's Paprika (Potato Chips)	Lay's Natural (Potato Chips)
Fun Mix (Salty Crackers)	Trio Zoutjes (Salty Crackers)
Kaas Zoutjes* (Cheese Crackers)	Snack Zoutjes* (Salty Crackers)
Tuc Cheese (Crackers)	Tuc Naturel (Crackers)
Thai Sweet Chili Nuts (Spicy Peanuts)	Katjang Pedis (Spicy Peanuts)
Cashews (Nuts)	Pistachios (Nuts)
Nut and Raisins Mix (Nuts)	Nut Mix (Nuts)

Note. *These items were not identical but close substitutes to items in the physical condition, all other items were identical across the two conditions.

Experimental Instructions

Figure S2 shows the instructions and comprehension questions given to the participants. After reading the instructions and answering the comprehension questions and before entering the main experiment, the participants further finished five training trials with the experimenter to ensure that the Becker, DeGroot and Marschak auction was well understood. In each training trial the participant had to determine a hypothetical willingness to pay for a (non-food) good, roll the dice to determine a hypothetical price and explain whether, given the stated willingness to pay and the randomly determined price, he would buy the good and, if so, at what price. The instructions in the computer condition were identical to those in the physical condition, except for page 2 shown in Figure S3.

Welcome,

You are participating in an experiment on decision making.
The study will last about 90 minutes.

The experiment is structured as follows:

1. You read the instructions for the experiment.
2. You answer some comprehension questions.
3. You can practice the task (5 min).
4. You do the task (35 min).
5. Your earnings are determined
6. You answer a short questionnaire (10 min).
7. You will stay in an adjacent room for further 30 min.

Please proceed to reading the instructions on the next pages.
Whenever you have a question, please ask the experimenter for clarification.

Figure S2. Instructions given to the participants in the physical condition. Page 1/6.

Instructions

Task
In this experiment, you will have an opportunity to buy a snack food from our store using €14.50 that you receive from us. You will receive the €14.50 once you have read the instructions and answered the comprehension questions correctly.

At the end of the experiment, you will be asked to stay in an adjacent room for 30 minutes. During this time, the only food that you will be allowed to eat is whatever snack you bought from us during the experiment.

The snack food items that are available in this experiment are real and recently purchased for the purpose of this experiment. You will see all of them during the experiment.

Your task in this experiment is to decide and tell us the maximum amount that you would currently be willing to pay for each of these items.

Round structure
The experiment consists of many rounds, all of which have a similar structure. In each round:

1. First, you will see which item is on offer in this round
2. Then you will enter the maximum amount that you are willing to pay for this item (in €). You will need to enter a number between €0 and €4. You can also enter decimal numbers like €3.47 using the dot as a decimal point.
3. You will answer a few questions about the item.

There is no strict time limit for giving an answer. Nevertheless, try to answer spontaneously, without thinking too much. After 22 of these rounds there will be a break, then there will be another 22 rounds. You will see each item only once. The items will be right in front of you so you can see them well. Please do not touch the items.

During this task a wristband will be attached to each of your arms.

Figure S2 (continued). Instructions given to the participants in the physical condition. Page 2/6.

Your earnings in this experiment
At the beginning of the experiment, you receive €14.50. You can use this money to buy one item of snack food from us. Whatever you do not spend remains yours, just like in everyday transactions. Although you will place bids on 44 items, you will only be allowed to buy one of them.

How do we determine whether you bought or not?
At the end of the experiment, **one** of the rounds will be randomly chosen. You will be asked to draw a card from a stack of cards numbered from 1 to 44. The round with the number that you draw will be the one that counts. Note that every round has the chance to be selected, and only one round will be selected. Therefore, you don't need to worry about spreading your €14.50 budget. In fact, you can treat every round as if it were the only round. Each time you have to bid on an item, it is in your best interest to report exactly your maximum willingness to pay for being allowed to eat this item at the end of the experiment.

After you picked a card to randomly select the round that counts, you can see which item was presented in this round and what was your bid on this item. Then, the actual price for the item will be determined randomly. Therefore, you will be asked to throw 3 dice that will determine the price. The price shown is the actual price for which you can buy the item from us. Please note that your bid does not influence the actual price of the item.

Whether you indeed buy the item from us depends on your bid and the actual, randomly determined price. If your bid is higher than or equal to the actual price (so you would be willing to pay the actual price) you will buy the item at the actual price and keep the rest of the €14.50.
On the other hand, if your bid is lower than the actual price(so you would not be willing to pay this actual price), then you do not buy the item and keep the €14.50.

Note that if you buy you never pay more than the actual, randomly determined price! If your bid is higher than the actual price you do not have to pay your bid, but just the actual price! Therefore, the best you can do in each round is simply to estimate what the item is worth to you (the maximum you would be willing to pay for it) and bid exactly this amount.

Figure S2 (continued). Instructions given to the participants in the physical condition. Page 3/6.

You should not bid more money on an item than you actually are willing to pay. Stating higher bids increases the chance that you will buy the item. However, the downside of this is that this involves the risk of buying the item at a price that is higher than what you are willing to pay for it.

For example: Suppose that the most you would like to pay for a bag of biscuits is €3, but in order to increase the chances of getting the biscuits you decide to bid €4. The actual price is randomly determined at €3.60. Then, you have to purchase the biscuits for €3.60, a price that is higher than what the biscuits are actually worth to you (€3).

You might think that your best strategy is to bid lower than your actual valuation for the item. This is incorrect. The price that you pay is determined by the numbers you throw with the dies and not by your bid. Bidding lower than your true value you would not affect the price that you pay, but you run the risk of not buying although the price is acceptable to you.

For example: Suppose that the maximum you would like to pay for a chocolate bar is €3.50, but in order to keep more money you decide to bid only €1. The actual price turns out to be €2. You will not buy the chocolate bar because you bid only €1. Had your bid been your true value of €3.50, you would have purchased the chocolate bar for €2 and kept €12.50 in cash.

To sum up, the best you can do in your own interest is to bid exactly the amount which you are maximally willing to pay for the item at stake.

Here are a few examples, to make this mechanism clear:

Example 1: In the round selected for payment, Manuel was bidding on chips. Manuel's bid in this round was €2.65. The randomly selected actual price turns out to be €2.00.
Manuel buys the chips because the actual price (€2.00) is lower than his maximum willingness to pay (€2.65). Manuel gets the chips and pays €2.00. He keeps €12.50 from his initial €14.50.

Example 2: In the selected round, Manuel was bidding on popcorn. Manuel's bid was €1.35. The randomly selected actual price turns out to be €4.00.

4

Figure S2 (continued). Instructions given to the participants in the physical condition. Page 4/6.

Manuel doesn't buy the popcorn because the store's price (€4) is higher than Manuel's maximum willingness to pay (€1.35). He keeps the €14.50.

Example 3: In the selected round, Manuel bid €2.80 on gummibears. The actual price for the gummibears turns out to be €2.50. Manuel buys the gummibears for €2.50 because his bid is greater than the actual price. He keeps €12.00 and gets the gummibears.

Please ask questions NOW if anything remains unclear.

You will have a chance to practice this task for 5 rounds that will not count towards determining your final earnings.

Please let the experimenter know that you finished reading.

Figure S2 (continued). Instructions given to the participants in the physical condition. Page 5/6.

Comprehension questions: (Choose all that apply)

1. *Suppose that in the round that is selected for payment you entered a **bid of €4** for a bag of chips. The randomly selected actual price turns out to be **€3**. What happens? Choose all that apply.*

   a) I buy the chips for €4
   b) I buy the chips for €3
   c) My earnings in cash are €10.50
   d) My earnings in cash are €14.50
   e) My earnings in cash are €11.50
   f) I don't buy the chips

2. *Suppose that in the selected round you had **bid €1** for a pack of crackers**.** The randomly selected actual **price** turns out to be **€2.75**. What happens? Choose all that apply.*

   a) I buy the crackers for €2.75
   b) I buy the crackers for €1
   c) My earnings in cash are €14.50
   d) My earnings in cash are €11.25
   e) My earnings in cash are €13.50
   f) I don't buy the crackers

3. *Suppose that in the round that is selected for payment, **you bid €3.20** for a bag of peanuts. **The actual price turns out to be €3.20**. What happens? Choose all that apply.*

   a. I buy the peanuts for €3.20
   b. I buy the peanuts for €2.50
   c. My earnings in cash are €11.30
   d. My earnings in cash are €12.00
   e. My earnings in cash are €14.50
   f. I don't buy the peanuts

6

Figure S2 (continued). Instructions given to the participants in the physical condition. Page 6/6.

Instructions

Task
In this experiment, you will have an opportunity to buy a snack food from our store using €14.50 that you receive from us. You will receive the €14.50 once you have read the instructions and answered the comprehension questions correctly.

At the end of the experiment, you will be asked to stay in an adjacent room for 30 minutes. During this time, the only food that you will be allowed to eat is whatever snack you bought from us during the experiment.

Please note: All of the snack food items that are available in this experiment are regular size, as available in the supermarket. They have been recently purchased for the purpose of this experiment.

Your task in this experiment is to decide and tell us the maximum amount that you would currently be willing to pay for each of these items.

Round structure
The experiment consists of many rounds, all of which have a similar structure. In each round:

1. On the screen you will see the name and a description of the food item that is on offer in this round.
2. Then you will enter the maximum amount that you are willing to pay for this item (in €). You will need to enter a number between €0 and €4. You can also enter decimal numbers like €3.47 using the dot as a decimal point.
3. You will answer a few questions about the item.

There is no strict time limit for giving an answer. Nevertheless, try to answer spontaneously, without thinking too much. After 22 of these rounds there will be a break, then there will be another 22 rounds. You will see each item only once.

During this task a wristband will be attached to each of your arms.

2

Figure S3. Page 2 of the instruction given to the participants in the computer condition. The other pages of the instructions were identical to the physical condition.

2 Analysis

Additional Descriptive Statistics

As part of our item randomization strategy, we formed pairs of similar items. Within one subject, items from the same pair were never presented in the same condition. Thus, it is possible to compute for each subject the difference in WTP across the two items of each pair. Figure S4 and Figure S5 show the difference in willingness to pay for each item pair and each participant separately for the physical and computer condition, respectively. It should be noted that, although items of the same pair were of the same type of snack food, they were still different, and any strong individual preferences (e.g. for or against cheese flavor on crisps, for or against orange flavor on chocolate, for or against raisins) still play a role.

In the manuscript we show subject-level differences across the two conditions for WTP. Here, we show this data also for liking and wanting. Figure S6 and Figure S7 show the average difference in liking and wanting across weight conditions for each participant, respectively.

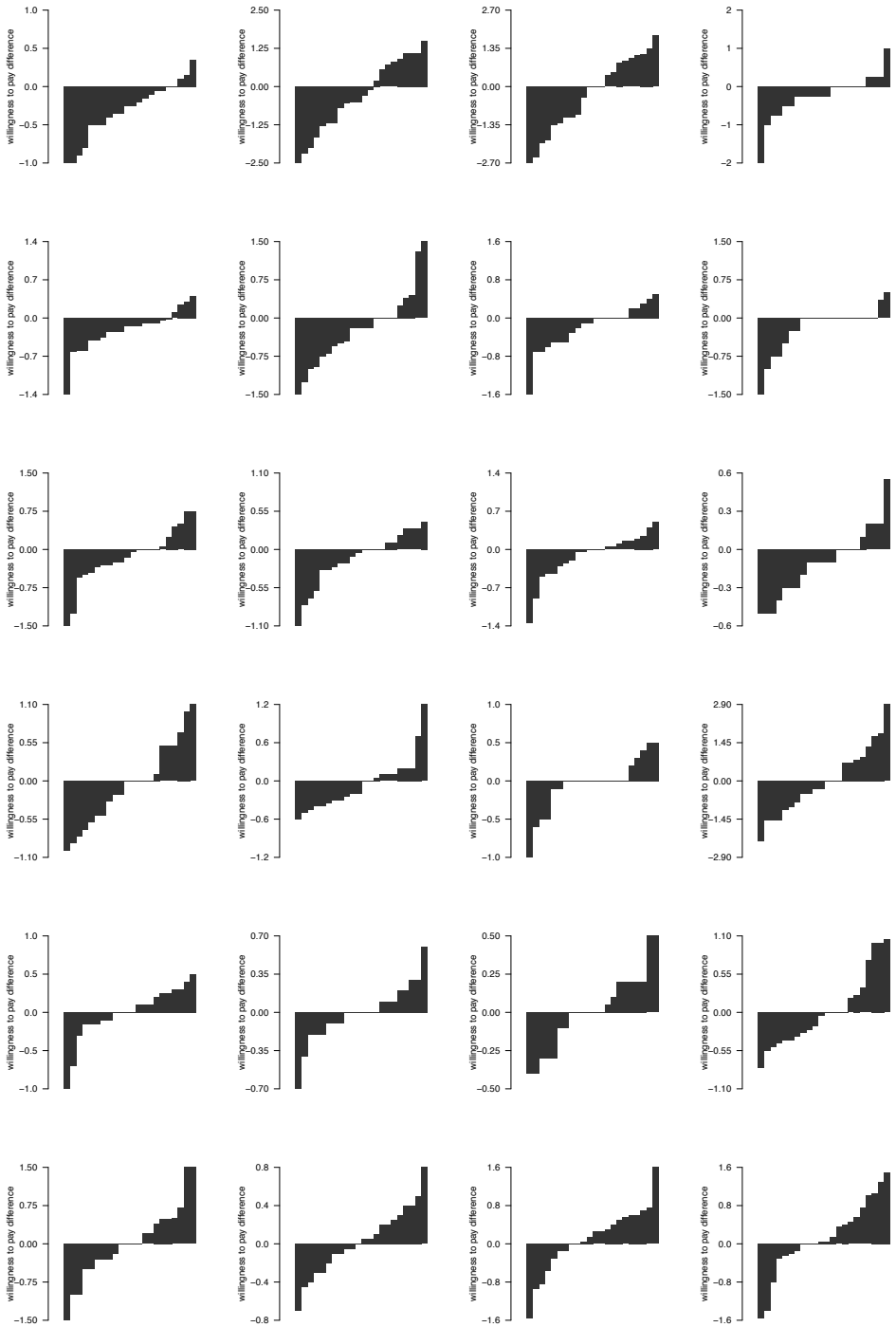

Figure S4. Willingness to pay differences across weight and no weight condition for each item pair and individual. Each barplot shows the willingness to pay differences of one subject in the physical condition. Each bar represents the difference in willingness to pay for one item pair.

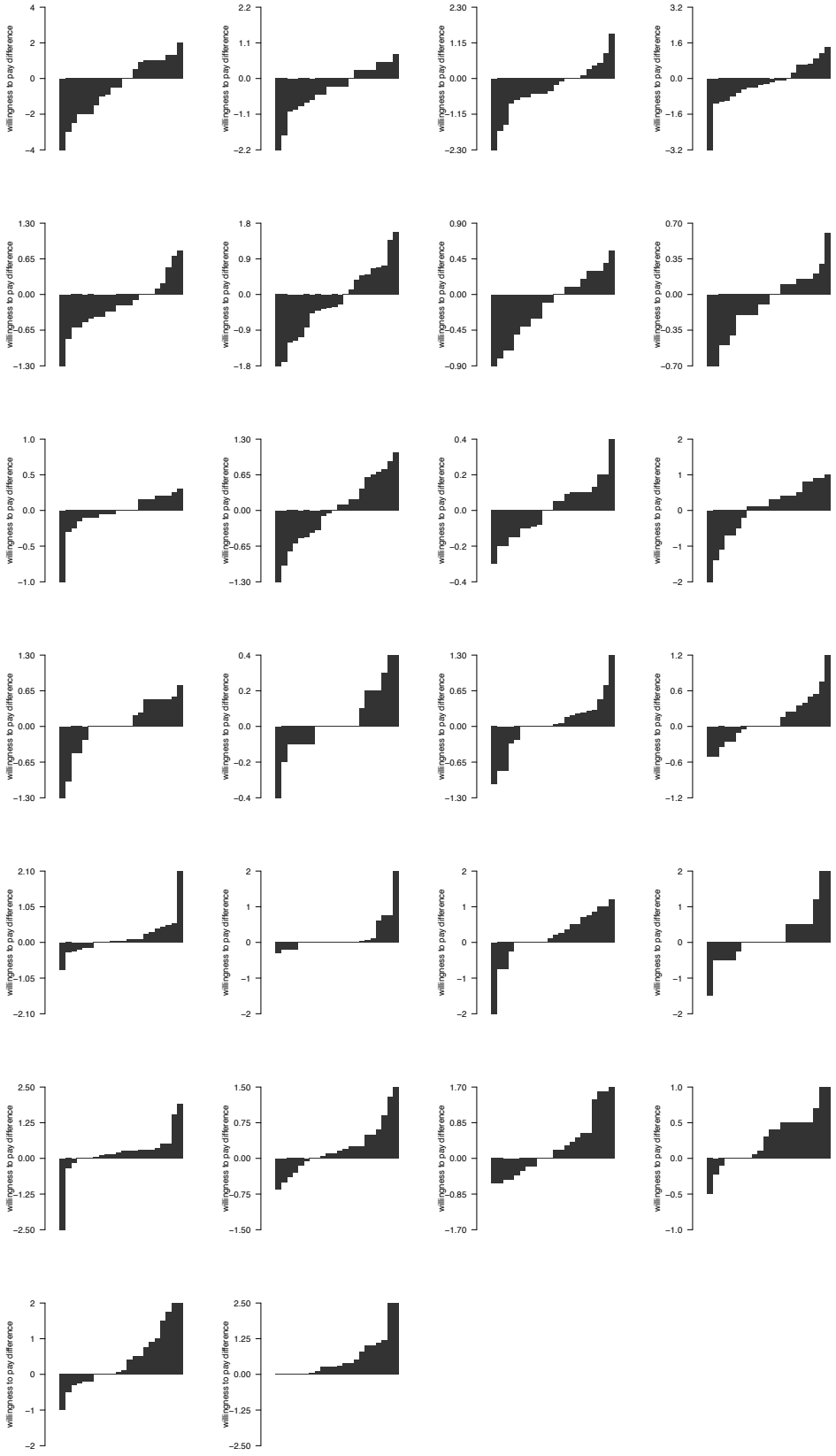

Figure S5. Willingness to pay differences across weight and no weight condition for each item pair and individual. Each barplot shows the willingness to pay differences of one subject in the computer condition. Each bar represents the difference in willingness to pay for one item pair.

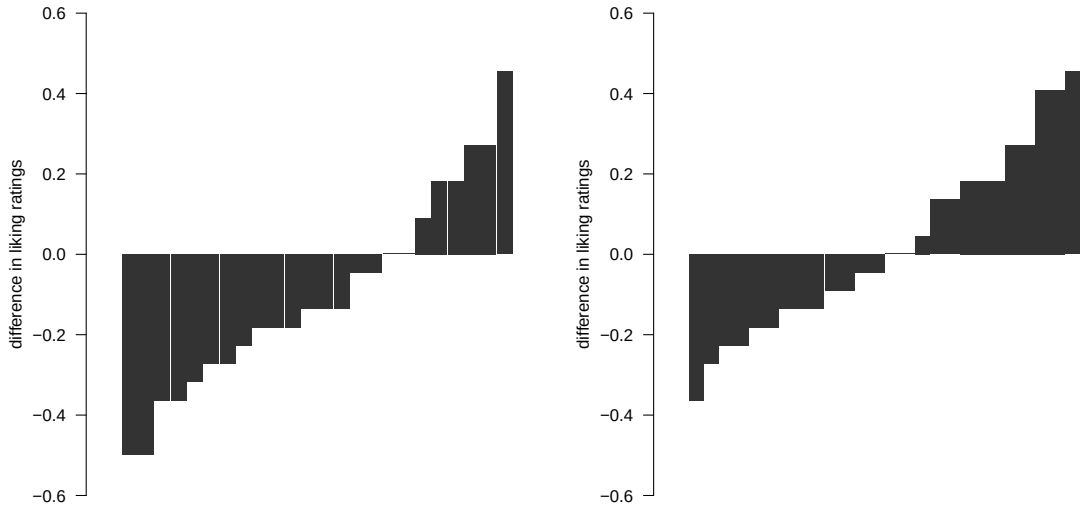

Figure S6. Average liking difference per individual across weight and no weight condition and across physical condition (a) and computer condition (b). Each bar shows the average difference in liking of one participant across weight conditions. Negative values indicate that participants gave a lower average liking rating for items in the weight condition.

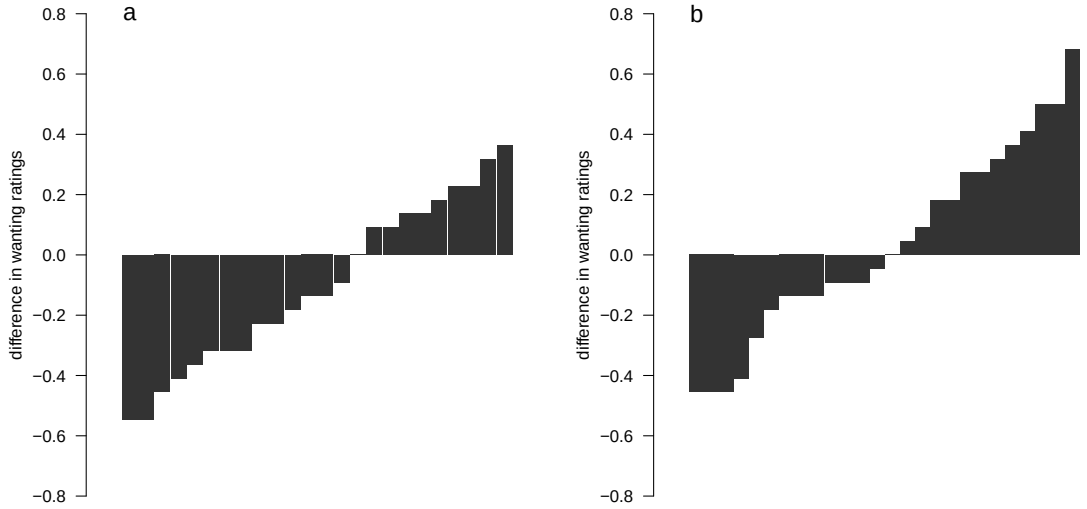

Figure S7. Average wanting difference per individual across weight and no weight condition and across physical condition (a) and computer condition (b). Each bar shows the average difference in wanting of one participant across weight conditions. Negative values indicate that participants gave a lower average wanting rating for items in the weight condition.

Hypothesis Tests / Regression Models

In this section we elaborate on our choice of analysis strategy, and on the underlying assumptions. In the experimental psychology literature, it is common to aggregate data on the subject level. For comparison, we therefore present the results of t-tests as well as of an ANOVA analysis, both of which rely on subject-level averages. Below we detail the assumptions underlying the regression model that we report in the paper.

t-Tests

The simplest way to test our hypothesis that wristbands with weight decrease the valuation for snack food items in the physical but not in the computer condition is by computing the average willingness to pay, liking and wanting for each subject in each of the weight conditions, and compare these with a paired t-test separately for the physical and the computer condition. This does not, however, test directly whether the effect of the weights is different between the computer condition and physical condition.

In the physical condition participants showed a significant decrease in willingness to pay (paired samples t-test, $t(23) = 3.74$, $p < .01$, two-sided), while in the computer condition, average willingness to pay under heavy weights was not significantly different from average willingness to pay under no weight (paired samples t-test, $t(25) = -0.81$, $p = .43$, two-sided).

The same analysis can be applied to liking and wanting ratings. Liking ratings were marginally lower in the physical condition when wearing heavy wristbands (paired samples t-test, $t(23) = 1.98$, $p = .06$, two-sided). This was not the case in the computer condition (paired samples t-test, $t(25) = -0.47$, $p = .64$, two-sided).

A similar result was obtained for wanting ratings (physical condition: paired samples t-test, $t(23) = 1.85$, $p = .08$, two-sided; computer condition: paired samples t-test, $t(25) = -0.53$, $p = .60$, two-sided).

Analysis of Variance

By comparing the data in the aforementioned way we are not able to test the interaction of reachability $\times$ physical effort directly and we cannot control for the order of the weight conditions. This is possible when using Analysis of Variance (ANOVA). As for the t-tests, data was aggregated on the subject level. Thus for each subject we calculated the average willingness to pay, liking and wanting for each weight condition and each participant.

The ANOVA model contained one within subject factor (anticipated effort: no weight vs. weight condition), two between subject factors (reachability condition: physical vs. computer condition and order: starting the experiment with no weight or starting the experiment with weight) as well as the covariate familiarity (aggregated on subject-level) and all possible interactions of the main factors. As can be seen in Table S2, the analysis for willingness to pay revealed a significant interaction of reachability (physical vs. computer) $\times$ physical effort (no weight vs. weight). For the wanting and liking ratings as dependent variable, as already indicated by the t-tests, the interaction of reachability $\times$ physical effort did not reach the significance threshold of $p < .05$ in the ANOVA models (Table S3 and Table S4).

Table S2
Repeated measures analysis of variance.
Dependent variable: willingness to pay.

	df	SS	MS	F	p
within subjects					
anticipated effort	1	0.01	0.01	0.01	0.93
familiarity	1	1.59	1.59	2.81	0.10
order	1	0.98	0.98	1.74	0.19
anticipated effort × order	1	0.43	0.43	0.76	0.39
error	45	25.40	0.56		
between subjects					
reachability	1	0.02	0.01	1.10	0.30
familiarity	1	0.09	0.09	5.56	0.02
anticipated effort × reachability	1	0.08	0.08	4.75	0.03
reachability × order	1	0.02	0.02	0.88	0.35
anticipated effort × reachability × order	1	0.01	0.01	0.39	0.54
error	45	0.76	0.02		

Note. 50 subjects. Data aggregated over items, two data points for each subject. Within-subject factor: Anticipated effort (no weights vs. weights). Between-subject factor: reachability (physical vs. computer screen). All p values are two-sided.

Table S3
Repeated measures analysis of variance.
Dependent variable: wanting ratings.

	df	SS	MS	F	p
within subjects					
anticipated effort	1	0.04	0.04	0.16	0.69
familiarity	1	1.46	1.46	6.54	0.01
order	1	1.87	1.87	8.37	< 0.01
anticipated effort × order	1	0.34	0.34	1.53	0.22
error	45	10.05	0.22		
between subjects					
reachability	1	0.03	0.03	0.69	0.41
familiarity	1	0.22	0.22	5.72	0.02
anticipated effort × reachability	1	0.07	0.07	1.81	0.19
reachability × order	1	0.15	0.15	3.78	0.06
anticipated effort × reachability × order	1	0.08	0.08	1.96	0.17
error	45	1.75	0.04		

Note. 50 subjects. Data aggregated over items, two data points for each subject. Within-subject factor: Anticipated effort (no weights vs. weights). Between-subject factor: reachability (physical vs. computer screen). All p values are two-sided.

Table S4
Repeated measures analysis of variance.
Dependent variable: liking ratings.

	df	SS	MS	F	p
within subjects					
anticipated effort	1	0.05	0.05	0.22	0.64
familiarity	1	2.73	2.73	11.13	< 0.01
order	1	0.43	0.43	1.74	0.19
anticipated effort $\times$ order	1	0.72	0.72	2.92	0.09
error	45	11.03	0.25		
between subjects					
reachability	1	0.03	0.03	1.44	0.24
familiarity	1	0.16	0.16	6.78	0.01
anticipated effort $\times$ reachability	1	0.06	0.06	2.32	0.13
reachability $\times$ order	1	0.12	0.12	4.90	0.03
anticipated effort $\times$ reachability $\times$ order	1	0.01	0.01	0.59	0.45
error	45	1.08	0.02		

Note. 50 subjects. Data aggregated over items, two data points for each subject. Within-subject factor: Anticipated effort (no weights vs. weights). Between-subject factor: reachability (physical vs. computer screen). All p values are two-sided.

Regression Analysis

By aggregating the data on the subject-level, we unnecessarily reduce the amount of information (e.g. on within-subject variability), and we cannot control for familiarity on an item by item basis.

For this reason we fitted random intercept regression models to the data as described in the manuscript. In using this model, we make the following assumptions: We assume that observations belonging to different subjects are independent, that is we assume that the residual error term is uncorrelated across individuals. In contrast to a regression model with normal standard errors, we do not assume that observations have equal variance. For example, some subjects might vary their bid substantially from item to item, and others very little. We do also not assume that the residual error term is independent within individuals. That is, even after accounting for the fact that some subjects may bid higher than others with a random intercept, we may still expect some form of dependency across the observations belonging to the same subject. This could for example arise if subjects made a mistake on one trial, and reacted to that mistake by changing bidding behavior, or if subjects made their bids for items dependent on the previously encountered items. We account for these possible dependencies within subjects, as well as for possible heteroscedasticity, by using cluster-robust standard errors (Rogers, 1993). An additional requirement when using cluster-robust standard errors is that the number of clusters (in our case subjects) needs to be sufficiently large. With 50 clusters of equal size our dataset is large enough for accurate inference with this method (Kezdi, 2004; Miller and Cameron, 2013). The model is estimated using generalized least squares.

Table S5
Random intercept regression model with control variables.
Dependent variable: wanting ratings.

	Coef.	**95% CI**	**p**
constant (computer condition, no weight)	0.45	[0.30, 0.61]	< 0.01
physical condition	0.10	[-0.11, 0.32]	0.35
weight	0.14	[-0.00, 0.28]	0.06
physical $\times$ weight	-0.21	[-0.41, -0.02]	0.03
familiarity	0.35	[0.28 ,0.41]	< 0.01
order of weight condition	0.29	[0.02, 0.57]	0.04
order $\times$ weight	-0.26	[-0.47, -0.05]	0.01
order $\times$ physical	0.11	[-0.30, 0.53]	0.59
order $\times$ physical $\times$ weight	0.22	[-0.08, 0.52]	0.15
σ_u (SD between subjects)	0.31		
σ_e (SD within subjects)	0.88		

Note. 2200 trials, nested within 50 subjects. Standard errors are corrected for potential heteroscedasticity and autocorrelations at the subject level. All p values are two-sided. Order indicates whether participant started with the weight (Order = 1) or no weight condition (Order = 0).

The regression results for WTP are shown in the manuscript. Table S5 and Table S6 additionally show the results for wanting and liking ratings as the dependent variable, respectively.

Model Diagnostics

Figure S8 to Figure S10 show diagnostic statistics of the fitted models. Figure Figure S8a reveals that there is indeed some heteroscedasticity in the willingness to pay data, with larger residuals for higher predicted values. As outlined above, this is accounted for by using cluster-robust standard errors. Figure Figure S8b suggests that residuals are identically distributed across the experimental conditions.

For liking and wanting, some effect of limiting the scale from 1-4 is evident in the residuals in Figures S9a and S10a, but no striking differences appear across the experimental conditions (see Figures S9b and S10b).

Table S6
Random intercept regression model with control variables.
Dependent variable: liking ratings.

	Coef.	**95% CI**	**p**
constant (computer condition, no weight)	0.81	[0.65, 0.98]	< 0.01
physical condition	-0.07	[-0.32, 0.19]	0.60
weight	0.09	[-0.02, 0.20]	0.11
physical $\times$ weight	-0.13	[-0.29, 0.02]	0.08
familiarity	0.35	[0.30, 0.41]	< 0.01
order of weight condition	0.06	[-0.23, 0.34]	0.70
order $\times$ weight	-0.19	[-0.33, -0.04]	0.01
order $\times$ physical	0.27	[-0.13, 0.67]	0.19
order $\times$ physical $\times$ weight	0.09	[-0.14, 0.33]	0.44
σ_u (SD between subjects)	0.33		
σ_e (SD within subjects)	0.85		

Note. 2200 trials, nested within 50 subjects. Standard errors are corrected for potential heteroscedasticity and autocorrelations at the subject level. All p values are two-sided. Order indicates whether participant started with the weight (Order = 1) or no weight condition (Order = 0).

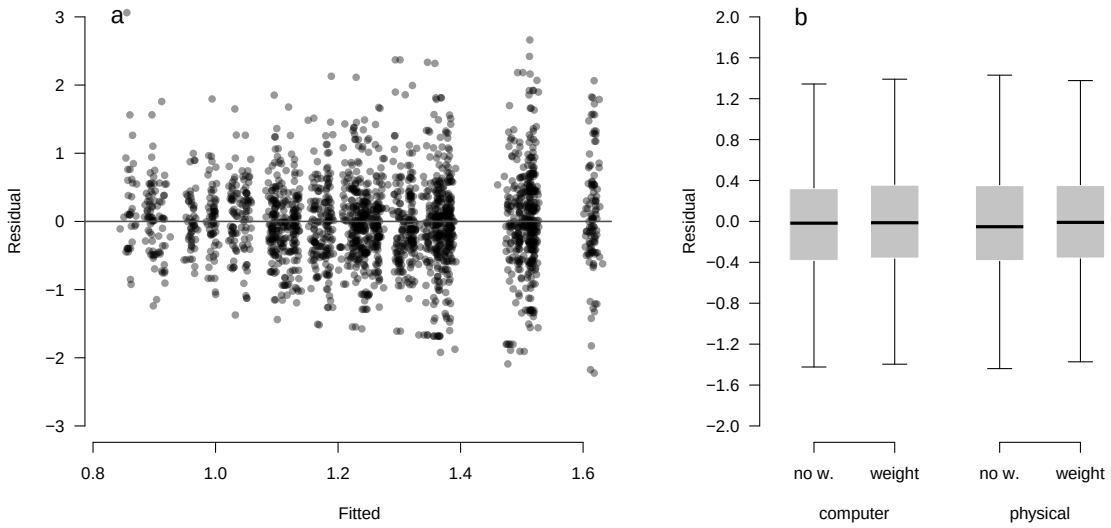

Figure S8. Residual plots of the willigness to pay random intercept regression. (a) Residuals vs. fitted values (slightly jittered) and (b) residual distribution across conditions.

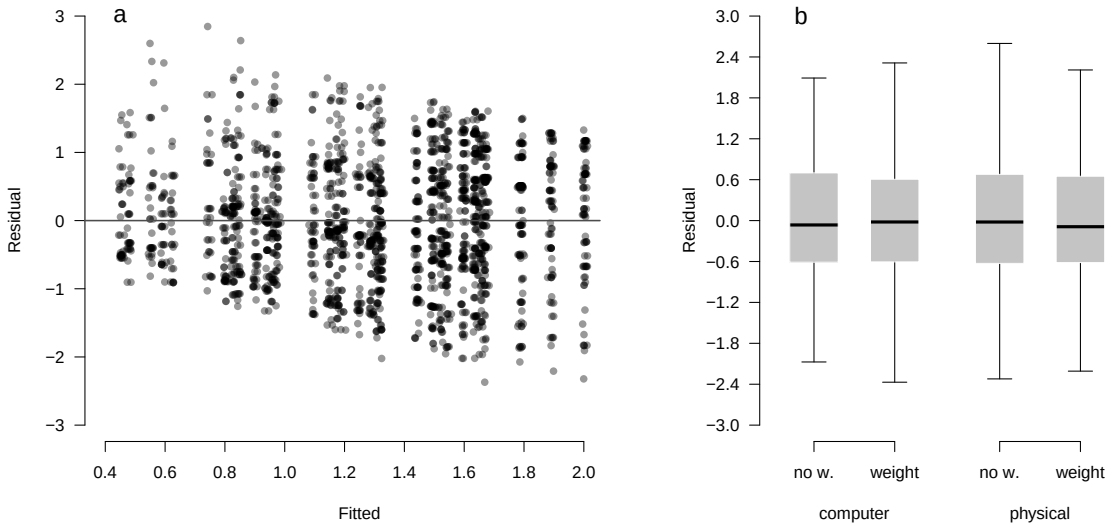

Figure S9. Residual plots of the wanting ratings random intercept regression. (a) Residuals vs. fitted values (slightly jittered) and (b) residual distribution across conditions.

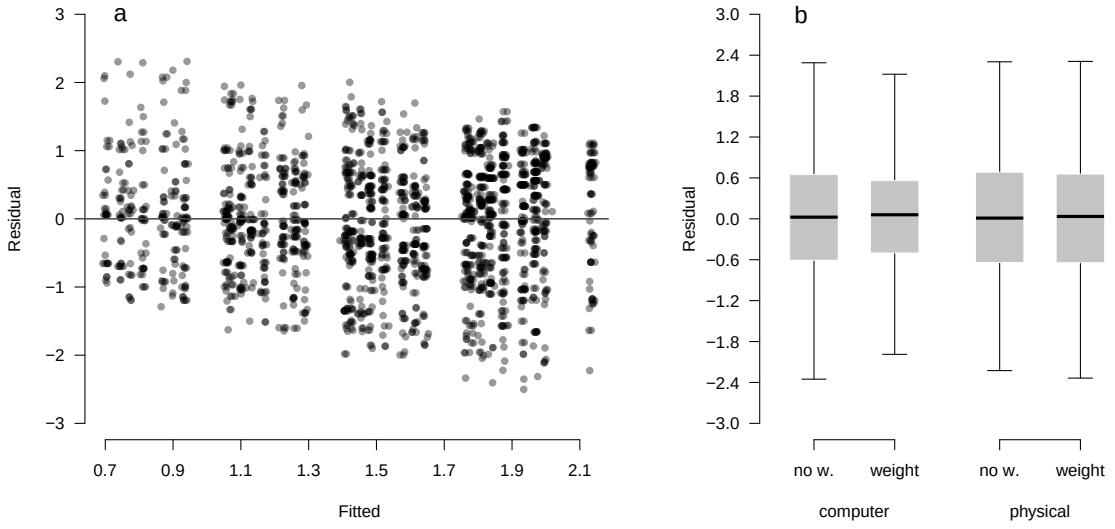

Figure S10. Residual plots of the liking ratings random intercept regression. (a) Residuals vs. fitted values (slightly jittered) and (b) residual distribution across conditions.

References

Kezdi, G. (2004). Robust standard error estimation in fixed-effects panel models. *Hungarian Statistical Review*, 9:96–116.

Miller, D. L. and Cameron, C. A. (2013). A practitioner's guide to cluster-robust inference. Working paper, University of California, Davis.

Rogers, W. H. (1993). Regression standard errors in clustered samples. *Stata Technical Bulletin*, 13:19–23.