The impact of computation time and experience on decision values

SUPPLEMENTARY MATERIALS

K. Carrie Armel

Antonio Rangel

I. Experiment 1

Method. 77 Stanford University undergraduates and local residents between the ages of 18 and 32 participated in the experiment. All had normal or corrected to normal vision. Individuals were excluded if they had a history of eating disorders, used drugs regularly, had dieted in the past year, were vegetarian, disliked junk food, or were pregnant. These selection criteria were designed to recruit individuals who liked junk food and were not trying to control their diet. In addition to the payoffs described below, individuals received $20 for their participation. All subjects provided informed consent prior to their participation.

Participants were required to sleep a minimum of six hours the night before the experiment, and to fast for four hours prior to the experiment, but to have eaten just prior to that time. The latter requirement increased subject's willingness to pay for food. All testing took place in mid-afternoon.

Five subjects were tested simultaneously in a lab with multiple cubicles. The experiment was self-paced with the instructions, practice trials, and tasks presented through the Eprime software (Psychology Software Tools Inc, Pittsburgh, PA). For each part of the experiment subjects silently read through the instructions, played several practice trials, and then performed the task. No deception was used.

The timeline for the experiment is summarized in Figure S1(a). Subjects performed two tasks. First they received an endowment of $10 that they could use to purchase food from us by placing bids on 70 different food items. Next they provided a liking rating for each food. At the end of the experiment one of the bidding trials was selected by drawing a ball from an urn. That was the only trial that counted: following the rules described below, the subject's bid on this trial determined whether he got the item and the price that he had to pay for it. At the end of the experiment subjects had to stay in the lab for an additional 30 mins. During this time they were allowed to eat as much as they wanted of the food that they purchased from us, but no other foods or drinks were allowed. Importantly, subjects were taught the experimental timeline during the instruction period and had complete knowledge of how their actions affected their payoffs.

Bidding trials were structured as follows (see Fig S1(b)). A 1000ms fixation point was followed by a picture of a food item that was presented for a predetermined amount of time. To decrease predictability and prevent boredom the foods appeared in one of six random locations equidistant from the center of the screen. Subjects then had an unlimited time to enter their bid. Note that since only one of the trials counted, subjects treated each decision as if it were the only one, and thus they could bid up to $10 every time. They bid by clicking with a mouse on a horizontal bar labeled $0 to $10 (the rest of the bar was not marked). A random inter-trial interval of either 2000 or 4000ms followed. Subjects bid twice on each food item for a total of 140 bidding trials. The bidding task was divided into two parts. In the first half of the trials each food item was presented in random order for 500 ms. After subjects had placed bids on all foods at that duration, the items were presented in random order a second time for either 500, 2000, 3500, 5000, or 6500 ms. Thus, each item belonged to one of six conditions defined by its presentation times.

Liking rating trials were structured as follows (see Fig S1(c)). Each item was presented once in random order for 3000ms. Subjects then had unlimited time to type their liking rating according to a scale of -50 ("worst food ever tasted") to 50 ("best food every tasted"), where 0 denotes a neutral item. Liking ratings were anchored to the question "how much would you like to eat this item at the end of the experiment". The inter-trial interval lasted 500ms.

We sold the items to the subjects by applying the rules of a Becker-DeGroot Auction to the trial that was selected to count. A random number between $0 and $10 dollars (in $0.25 increments) was selected from an urn. Let n denote the random number that was selected and let b denote the subject's bid. If $b \geq n$, the subject got the food item and paid $\$n$. If $b < n$, the subject did not get the item but kept the full $10 of bidding money. Note that since the subjects kept whatever funds they did not use, they were *de facto* spending their own money to purchase the food.

The key feature of the auction procedure is that it satisfies incentive compatibility: if a subject's true value for an item is v, her best response is to bid exactly v. Any deviation from this strategy results in a lower expected payoff. Since the rules of the auction are somewhat complicated, we spent significant time training the subjects. In particular, we emphasized that their best strategy was to "go with their gut feeling" about how much each item was worth to them, and then to bid that amount. Debriefing during a pilot experiment confirmed that subjects complied with these instructions.

All of the items sold to the subjects were junk foods (e.g., Snickers Bars) so that self-control processes did not differentially affect the evaluations of specific items. Pre-testing showed that the foods were familiar to our participants. The food items were presented using 4x4 inch photographs with resolution a resolution of 72 dpi (see Figure S1(d) for some examples). In the photographs, both the food and packages were visible. We edited the pictures so that the largest dimension of the food completely filled either the horizontal or vertical dimension of the photograph.

About the design. Several aspects of the design are worth emphasizing. First, every item was presented twice: the first time at 500ms, and the second time at either 500, 2000, 3500, 5000, or 6500 ms. Presenting items twice allows to test for the presence of previous exposure effects. Presenting the items at different lengths on the second presentation allows us to test for the presence of attention duration effects. Second, we controled computation time by manipulating the amount of time that items were displayed on the computer screen. We asked the subjects to look at the item the entire time that it displayed, but we did not collect a measure of compliance. Third, we cannot rule out the possibility that subjects kept computing a value during the time it took to enter the bid, after the disappearance of the picture. However, a simple regression analysis shows that every second of additional exposure to the stimuli decreased reaction times by only 32ms ($p<0.006$), which is only 2% of the change in exposure times. Therefore, *changes* in exposure time map almost one-to-one to *changes* in computation time, even if subjects kept computing the DVs after the picture had been removed from the screen. Fourth, we use the liking-rating measure collected at the end of the experiment as our measure of the consumption value of the items. This is justified by the model: when computation time is kept equal for all items, as it is the case for the liking ratings, the resulting DVs provide on average an ordinally correct ranking of the items.

Data cleaning. Data from 6 subjects is excluded from the analysis because they did not adequately follow instructions. Three subjects rated their hunger state as "completely full", even though they were asked not to eat for 4 hours prior to the experiment, one subject rated his sleepiness as "ready to fall asleep", and two subjects expressed bids and liking ratings that were inconsistent with each other (they rated a large fraction of the items as desirable but bid $0 in almost every trial).

Negatively rated items (rating < 50) are dropped from the analysis because the bids are only a good measure of DVs for positive and neutral items (since subjects were not allowed to place negative bids, their best strategy was to bid zero for all negative items). In addition, we drop trials in which reaction times are greater than three standard deviations from the mean of all bid trials, as well as trials in which bids are greater than three standard deviations from each subject's own mean. The former criterion helps to eliminate variation in responses due to subjects' distractions. The later criterion cleans up egregious mistakes in entering bids through the mouse clicks. For the remaining data, the average liking rating is 16.9 (SD =12.6), the average reaction time is 2189ms (SD = 853ms), the maximum reaction time is 4852ms, and the average bid is $1.8 ($SD$ = $1.8). The magnitude of the reaction times is reasonable given the complexity of the response involved in bidding: subjects need to recognize the end of the trial, transform their "gut feeling" about the DV of the item into a motor response, and execute the motor response by moving the mouse to the appropriate location in the screen.

II. Experiment 2

Method. 59 Stanford undergraduates and local residents participated in the experiment. The screening criteria were identical to those of experiment 1 with two exceptions: we allowed individuals who did not like junk food, and we excluded foreigners and those raised by more than one guardian who grew up in a foreign country. The latter exclusion criteria was instated because during pilot testing it became clear that many subjects in this group find our food items appetizing rather than aversive. In addition to the payoffs described below, individuals received $30 for their participation. All subjects provided informed consent prior to their participation.

The key difference with Experiment 1 is that items are now aversive instead of appetitive. The stimuli included items such as spam and spinach baby food. Because of the difficulty in finding foods that subjects consistently rated as negative, we limited the stimulus set to 35 food items.

The procedures are similar to those for Experiment 1, including the sleep and fasting requirements. Subjects were told at the beginning of the experiment that they would have to eat five spoonfuls of the item shown in a randomly selected trial. They were also endowed with $10 (in addition to their $30 participation fee) that they could use to bid for avoiding having to eat the item. In the first part of the experiment they bid twice on each food item, for a total of 70 bidding trials. In the second part they provided liking ratings for each item by clicking with a mouse on a continuous scale from -100 to +100.

As before, the bidding task was divided into two parts. In the first half of the trials each food item was presented in random order for 1000ms. Afterwards the items were presented again in random order for 1000, 2500, 4000, 5500, and 7000ms. We increased the presentation times because pilot work showed that the negative food items were less familiar than the positives and subjects needed longer times to recognize them.

At the end of the experiment one of the bidding trials was randomly selected by drawing a ball from an urn and the outcome was implemented using a Becker-DeGroot Auction. In this case the rules of the auction were as follows. Let b denote the bid by the subject and n denote a random number drawn from an urn with 40 balls labeled \$0, \$0.25, … , \$10. If $b \geq n$, the subject did not have to eat the item but had to pay an amount equal to n. If $b < n$ the subject had to eat at least 5 spoonfuls of the item but got to keep the entire $10 of bidding money. As before, the auction is incentive compatible and the best strategy for a subject is to bid its true value for avoiding having to eat a food.

During the instruction period we emphasized that no deception was used in the experiment and that we expected the subjects to eat the foods if they lost the auction. The size of the bids shows that the foods were aversive and that subjects were trying to avoid them. Subject's well-being was not jeopardized by

the experiment since they could always avoid having to eat an item by bidding $10 (a fact that was emphasized during the instruction period).

Data cleaning. Data from 8 subjects is excluded from the analysis because they did not adequately follow instructions. Four subjects rated their hunger as "completely full" and one rated his sleepiness as " ready to fall asleep". An additional three subjects bid $0 as a general strategy to maximize their earnings (less than $0.03 on 70% or more of the trials) even though they rated some of the items as highly aversive. Two subjects decided not to participate in the experiment after reading the instructions.

Positively rated items (liking rating > 0) are dropped from the analysis because the bids are only a good measure of DVs for negative and neutral items (since subjects were not allowed to place negative bids, their best strategy was to bid zero for all positive items). For the same reasons as before, we drop trials in which reaction times are greater than three standard deviations from the mean of all bid trials, as well as trials in which bids are greater than three standard deviations from each subject's own mean. For the remaining data, the average liking rating is -57.1 (*SD* =34.5), the average reaction time is 2723ms (*SD* = 1541ms), the maximum reaction time is 8421ms, and the average bid is $5.24 (*SD* = $3.74).

III. Experiment 3

Methods. 56 Stanford University undergraduates and local residents participated in the experiment. All had normal or corrected to normal vision. Individuals with neurological or psychiatric disorders, or those who use drugs regularly, were not allowed to take part in the experiment. Individuals received $20 for their participation. Informed consent was obtained from the subjects at the beginning of the experiment.

The stimuli evaluated by the subjects consisted of 40 art and 40 music posters displayed at 3x6 inches and 72 dpi. The experiment was divided into three parts. In each part the subjects submited a liking rating for each poster, which were shown in random order. In part 1 posters were shown for 500ms and subjects entered a liking rating by clicking with a mouse on a continuous scale labeled -100 to +100. In part 2 posters were shown for 500, 2000, 3500, 5000, or 6500 ms and the liking ratings were entered using a similar method. In part 3 posters were shown for 3000ms and the liking-rating scale was replaced by a continuous scale labeled from -2 to +2. In all cases liking-ratings were anchored to the question "how much would you like to own this poster at the end of the experiment?", and thus provide an alternative measure of DVs. Subjects did not get a poster and thus there were no material outcomes contingent on the subject's decisions.

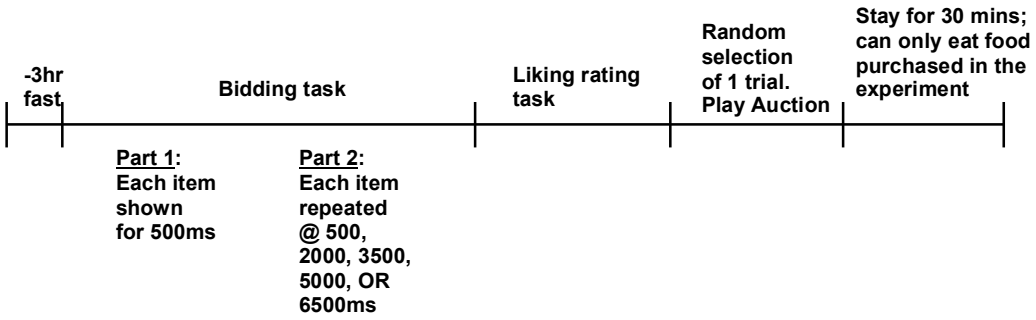
Data cleaning. Data from one subject is excluded from the analysis because he rated his tiredness at the time of the experiment as "ready to fall asleep". For the same reasons as before we drop trials in which reaction times are greater than three standard deviations from the mean of all bid trials, as well as items for which a subject's first or second liking rating was greater than three standard deviations from his or her own mean. After these cutoff criteria, the average and maximum reaction times are 1743ms (SD = 493ms) and 3054ms respectively. The average liking rating for parts 1 and 2 is -2.17 (SD = 19.72), and for part 3 is -0.10 (SD = 0.41).

We use the third liking rating to categorize items into positives and negatives. Since liking ratings can be used to measure both positive and negative DVs, all items are retained in the dataset. Partitioned by valence, the average liking rating for positives in parts 1 and 2 is 27.10 (SD = 35.02), and for negatives is -32.28 (SD = 38.65).

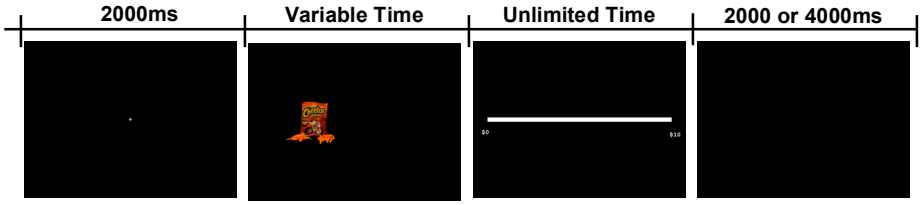
We use a linear mixed model to test the reliability of the liking ratings. A linear mixed effects regression of liking ratings in part 2 on ratings in part 1 generates a coefficient of 0.88 ($z=88.53$, $p < 0.0000$) with an intercept of -2.19 (z = -2.44, $p < 0.05$). A similar regression of liking ratings in part 3 on ratings in part 2 generates a coefficient of 0.93 ($z=85.78$, $p < 0.0000$) with an intercept of -1.78 (z = -1.21, p = 0.226). This suggests that the liking ratings are noisy, but highly reliable.

Figure S1. Methods for experiment 1. (A) Timeline. (B) Bidding trials. (C) Liking rating trials. (D) Examples of positive food items.

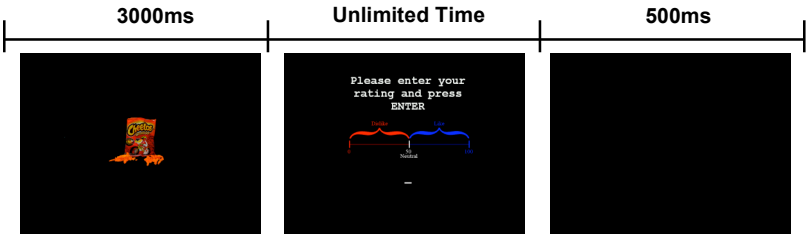
A



B



C



D

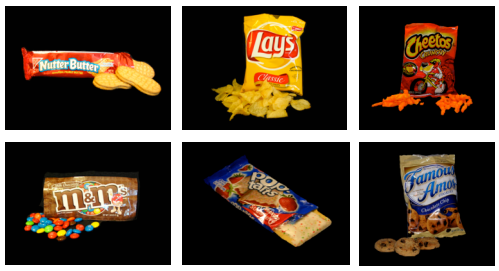


Figure S2. Results for experiment 1. (A) Previous exposure effect: Estimated change in bid between the first exposure and the second exposure at 500ms. (B) Attention duration effect: Estimated change in bid between 500 and 3500 ms in the second exposure. (C) Estimated total change in bid between the first exposure and the second exposure at 3500 ms. (D) Estimated marginal effect of 1500ms of additional exposure for different presentation times.

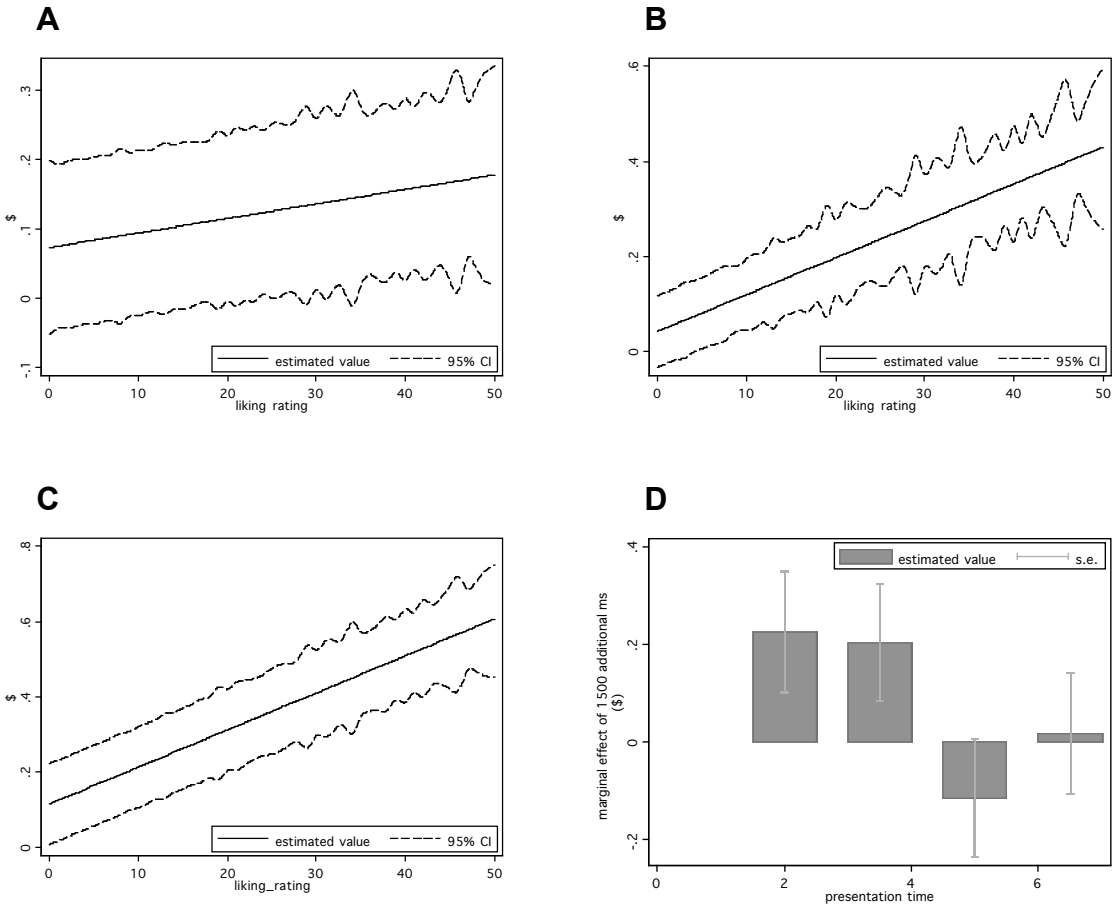


Figure S3. Results for experiment 2. (A) Previous exposure effect: Estimated change in bid between the first exposure and the second exposure at 1000ms. (B) Attention duration effect: Estimated change in bid between 1000 and 5500 ms in the second exposure. (C) Estimated total change in bid between first exposure and second exposure at 5500 ms. (D) Estimated marginal effect of 1500ms of additional exposure for different presentation times.

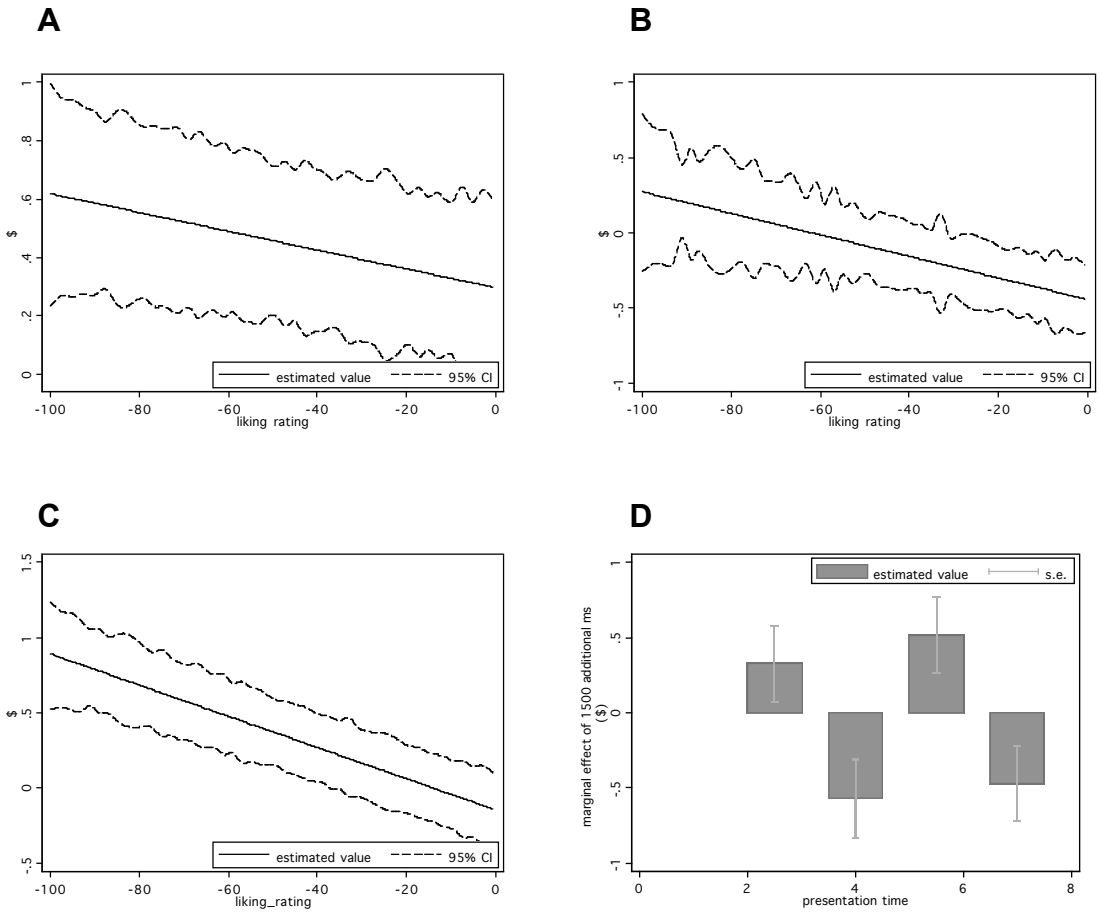


Figure S4. Results for experiment 3. (A) Previous exposure effect: Estimated change in bid between the first exposure and second exposure at 500ms. (B) Attention duration effect: Estimated change in bid between 500 and 5000 ms in the second exposure. (C) Estimated total change in bid between first exposure and second exposure at 5000 ms. (D) Estimated marginal effect of 1500ms of additional exposure for different presentation times.