



OPEN

Healthy decisions in the cued-attribute food choice paradigm have high test-retest reliability

Zahra Barakchian^{1✉}, Anjali Raja Beharelle^{2,3} & Todd A. Hare^{2,3}

Food choice paradigms are commonly used to study decision mechanisms, individual differences, and intervention efficacy. Here, we measured behavior from twenty-three healthy young adults who completed five repetitions of a cued-attribute food choice paradigm over two weeks. This task includes cues prompting participants to explicitly consider the healthiness of the food items before making a selection, or to choose naturally based on whatever freely comes to mind. We found that the average patterns of food choices following both cue types and ratings about the palatability (i.e. taste) and healthiness of the food items were similar across all five repetitions. At the individual level, the test-retest reliability for choices in both conditions and healthiness ratings was excellent. However, test-retest reliability for taste ratings was only fair, suggesting that estimates about palatability may vary more from day to day for the same individual.

Food decisions are an integral part of our daily lives. The quality and quantity of the foods we eat have a substantial impact on our health and well-being. There has been a great deal of basic science and clinical research into the determinants and consequences of food choices, as well as numerous efforts to design and validate interventions that lead to healthier eating behaviors^{1–5}. Food choice paradigms play a central role in much of this research.

There are several ways food choice paradigms are used in basic and clinical research. Food choice paradigms are a critical tool for investigating the nature of the decision process itself at the behavioral, computational, and neural levels^{6–18}. In addition, they can be used to compare groups^{19–23}, evaluate the effectiveness of a separate intervention (e.g. behavioral, pharmacological, surgical)^{2–4,16,24–26}, or even be incorporated into the intervention itself^{27–30}.

Food choice paradigms come in many different flavors. The exact details of the food choice task used in a given study will depend on the hypothesis being investigated and the practical constraints of the experimental setting. For example, in some cases, participants make choices over foods types and quantities in real buffet-like settings, fixed-option ad libitum meals, or snacking opportunities, while in other experiments plastic food replicas, pictures, or text are used to indicate the available options¹. Often in studies using plastic replicas, pictures, or text, participants are asked to make many different food choices with the understanding that one of those choices will be selected to count, where participants will consume the chosen food “for real”. In general, individuals’ choices over food representations such as pictures are associated with aspects of their actual eating behaviors when measured on the same day or in the future and anthropomorphic measures^{20,21,31}, indicating that these paradigms can have a useful degree of ecological and external validity.

Here, we report on decision patterns across five repetitions of a food-picture-based choice paradigm that explicitly cued participants to consider certain food attributes before making their choices on a subset of trials. We refer to this paradigm as the cued-attribute food choice task. Several previous studies have shown that, when participants complete the cued-attribute food choice task a single time, the attribute cues have significant effects on both choice outcomes and brain activity measured by functional magnetic resonance imaging (fMRI)^{7,13,32,33}. Specifically, on trials in which they were cued to consider the healthiness of the food options before making a consumption decision, participants selected healthier items more often compared to baseline and other cue conditions (e.g., cues to consider tastiness or cues to choose naturally based on whatever comes to mind). Food

¹Department of Cognitive Neuroscience, Institute for Research in Fundamental Sciences, Tehran, Iran. ²Zurich Center for Neuroeconomics, Dept. of Economics, University of Zurich, Zurich, Switzerland. ³Zurich Center for Neuroscience, University of Zurich, ETH Zurich, Zurich, Switzerland. ✉email: zbarakchian@ipm.ir

choices in the health-cued condition have been associated with the structure and functional activity within and connectivity between the dorsolateral and ventromedial prefrontal cortex^{7,32,33}.

However, it is unknown whether repeated experience or practice with the cued-attribute food choice task will lead to changes in choices during health-cued or baseline trials. Does the influence of the health cues increase, decrease, or remain stable with experience and/or time? Does repeated experience with explicitly considering health-related attributes in the cued trials spill over to affect choices in the baseline trials? Previous work has shown that choosing one food over another in an experiment setting can lead to both short and long-term changes in the relative evaluations of the chosen and unchosen items^{29,30,34}. Thus one might expect changes in behavior when repeating the cued-attribute or any other food choice task. Furthermore, explicitly reappraising positive or negative aspects of food items during regulation of craving training has been shown to change subsequent dietary choices as well^{27,28}. The cued-attribute food choice task does not necessarily involve active re-evaluation or reappraisal, but it does direct attention toward healthiness attributes that are often overlooked by the average decision maker when choosing between food items. It is plausible that once a food item appears in a health-cued trial, some of its healthiness aspects remain salient and influence its evaluation if it subsequently appears in a natural, baseline trial. This carry over from health-cued to natural decisions may increase with repeated exposures in health-cued trials leading to smaller differences in choice behavior between health-cued and natural trials after extended experience with the cued-attribute food choice task. Alternatively, more experience shifting into a healthiness-oriented mindset in response to the task cues might lead to greater differences between the health-cued and natural trials. However, past studies using this task found no evidence for changes in choice behavior within or across conditions over the course of a single session⁷. Moreover, food preferences in daily life are quite stable and behavior in non-cued food choice paradigms shows high test-retest reliability within a laboratory setting³⁵. Thus, there are reasons to think that repeated administration of the cued-attribute food choice task will or will not lead to changes in dietary decision making.

We show that the average pattern of choices in the baseline condition as well as the influence of the health cues remain fairly stable over five repetitions of the task across 14 days. Furthermore, the test-retest reliability of individuals' choices was high across both conditions, consistent with previous reports of high test-retest reliability in other food choice paradigms³⁵. Interestingly, the reliability of subjective healthiness ratings was also high, but taste or palatability rating reliability was only fair.

Results

Figure 1 shows a schematic representation of the food choice task used in this study. Participants were instructed to fast three hours prior the experiment in order to increase the value of foods. In the beginning of the experiment, participants completed a rating phase for 180 images in which they judged in two different phases, how tasty or how healthy they thought each food item to be. In each trial of the subsequent choice task participants have to choose between two food items. Within the choice task, there were two types of decision conditions that differed in the cues that are provided for the participants. In the health-cued condition, subjects are cued to consider the healthiness of the foods while making decisions. In the natural-cued condition, subjects are cued to make decisions naturally based on whatever freely comes to their mind. At the end of each task session one of their choices was randomly selected and participants received and had to eat the chosen food while in the lab. Thus, in both the health-cued and natural-cued conditions, participants had to keep in mind the fact that they may have to eat the food they select at the end of that day's session.

Choice outcomes. We examined the effects of attribute (taste, health), context (natural-cued and health-cued condition), and session (1:5) on healthy choice behavior using a Bayesian hierarchical logistic regression. This regression estimated the probability of selecting the healthier of the two food options as a function of the differences in tastiness and healthiness ratings between the foods, the cue condition (natural-cued, health-cued), and the experimental session (1–5). Table S1 show that, during the baseline (i.e. natural-cued) condition on the first session, the difference in the taste attribute was significantly associated with the choice outcome (1.22, 95% highest density interval (HDI) = [0.94, 1.51]). In contrast, the difference in the healthiness attribute was not significantly related to choice outcomes during natural-cued trials (– 0.02, 95% HDI = [– 0.20, 0.15]). In session 1, healthier choices were made more often in the health-cued relative to natural-cued condition. There was a significant main effect of health-cued trials (1.52, 95% HDI = [1.04, 1.98]), such that the healthier options were chosen significantly more in the health-cued condition compared to the baseline (natural-cued condition). There were significant interactions between the health cues and the differences in both the taste (– 0.73, 95% HDI = [– 1.02, – 0.45]) and healthiness attributes (1.22, 95% HDI = [0.85, 1.60]). This means that the taste attribute had a significantly stronger effect in the natural-cued condition compared to the health-cued condition, and the health attribute had a significantly stronger effect in the health-cued condition compared to the natural-cued condition.

The overall pattern of behavior observed in session 1, remained similar in sessions 2–5 (Fig. 2). There was a slight increase in the average number of healthy choices during the natural condition in session 2 relative to session 1, but no significant difference in natural-cued choices between session 1 and any other session (Table S1). There was a significant increase in the number of healthy choices during the health-cued compared to natural-cued trials within all five sessions, although the difference between health and natural-cued trials decreased slightly after the first session (Table S1, interaction term condition:session). Figure 2 shows that there were slightly more healthy choices during the natural-cued and slightly few healthy choices in health-cued trials in sessions 2–5 versus 1. In addition, the influences of the taste and healthiness attributes differed less between the natural and health-cued trials in sessions 2–5 versus 1 (Table S1 interaction terms td:condition:session and hd:condition:session). Notably, the effects of health cues on the way the two attributes influenced choice outcomes

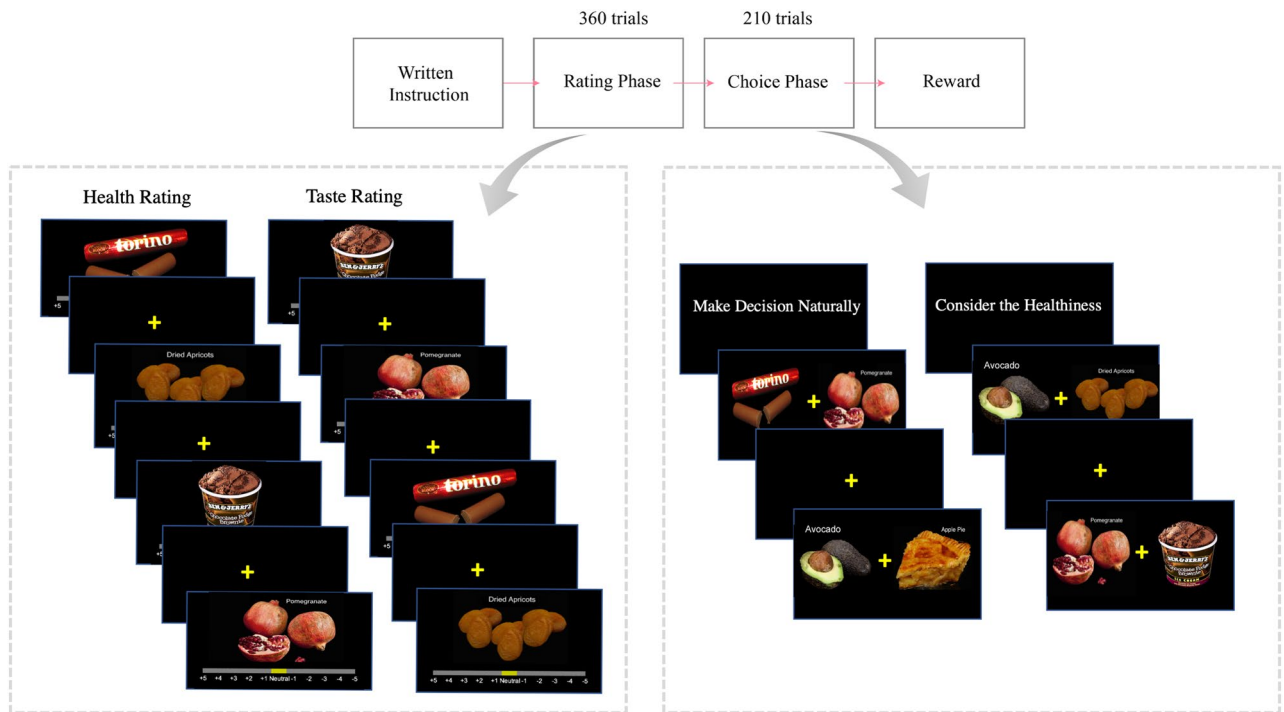


Figure 1. Experiment structure. The experiment had two phases, ratings and choices. In the ratings phase, participants rated the taste and health aspects of the foods using a visual analog scale ranging from -5 to $+5$. The order of the two ratings was counterbalanced. In the choice task, participants had to choose one of the two food items to eat at the end of the experiment. Within the choice phase, there were two conditions that differed in the attention cues given to the participants. In the health-cued condition, subjects were cued to consider the healthiness of the foods while making decisions. In the natural-cued condition, subjects were cued to make decisions naturally using whatever features freely came to mind. At the end of the session on each day, one of the participant's choices was randomly selected and the participant was given the chosen food to eat in the behavioral laboratory.

and the overall proportion of healthy choices did not significantly differ in sessions 2–5. In other words, from session 2 onward, choice behavior was quite stable.

In addition to choice outcomes, we also tested the repeated subjective ratings for taste and healthiness attributes. The group average ratings did not significantly change across sessions (Fig. S1). The ratings for the taste and healthiness attributes give us information about the participants' subjective opinions of those attributes. This in turn allows us to test how the subjective opinions about the taste and healthiness attributes influence the food choices participants make. This influence is not necessarily fixed, indeed it changes substantially across the two conditions as illustrated in Fig. 2b. Together, the repeated choices and ratings indicate that, aside from a small change between the first and second sessions, the subjective ratings for taste and healthiness as well as the impact of those attributes on choice outcomes within each condition remain similar at the group-average level.

To address the question of how well ratings from the first session relate to choices in subsequent sessions, we have done two sets of analyses. We used the fitted regression coefficients from the regression model model described in Table S1 to calculate the probability of choosing the healthier option with respect to the health and taste ratings (always from session 1) of the two options. If we treat a probability ≤ 0.5 as choosing the less healthy option, and probabilities ≥ 0.5 as choosing the healthier option, then this model would coincide with the participants' actual choice behavior 72% of the time in session 1 compared to $73 \pm 1\%$ in sessions 2–5.

We also conducted a different set of analyses to directly predict choices from the taste and healthiness ratings separately. In this case the predictions were simply that the participants would choose the item that they rated as tastier (healthier) in session 1 over the alternative item. Taste ratings from session 1 predicted session 1 choices correctly $72 \pm 3\%$ of the time in natural-cued trials, and $57 \pm 2\%$ of the time in health-cued trials. Healthiness ratings from session 1 predicted session 1 choices correctly $53 \pm 2\%$ of the time in natural-cued trials, and $59 \pm 2\%$ of the time in health-cued trials. The regressions summarized in Tables S4 and S5 show that these prediction rates did not significantly change when using session 1 ratings to predict choices in the subsequent sessions, 2–5.

Response times. Participants responded faster in both choice conditions as their experience with the cued-attribute food choice task increased. However, the effects of the health cue and differences between the healthiness and taste attributes remained consistent over all five sessions. Health cues increased the influence of healthiness attributes on choice response times. The differences between the healthiness attributes of the two food options had no significant effect on the response times in the natural condition (0.00, 95% HDI = $[-0.01, 0.01]$). However, differences in healthiness did influence response times for health-cued trials, such that par-

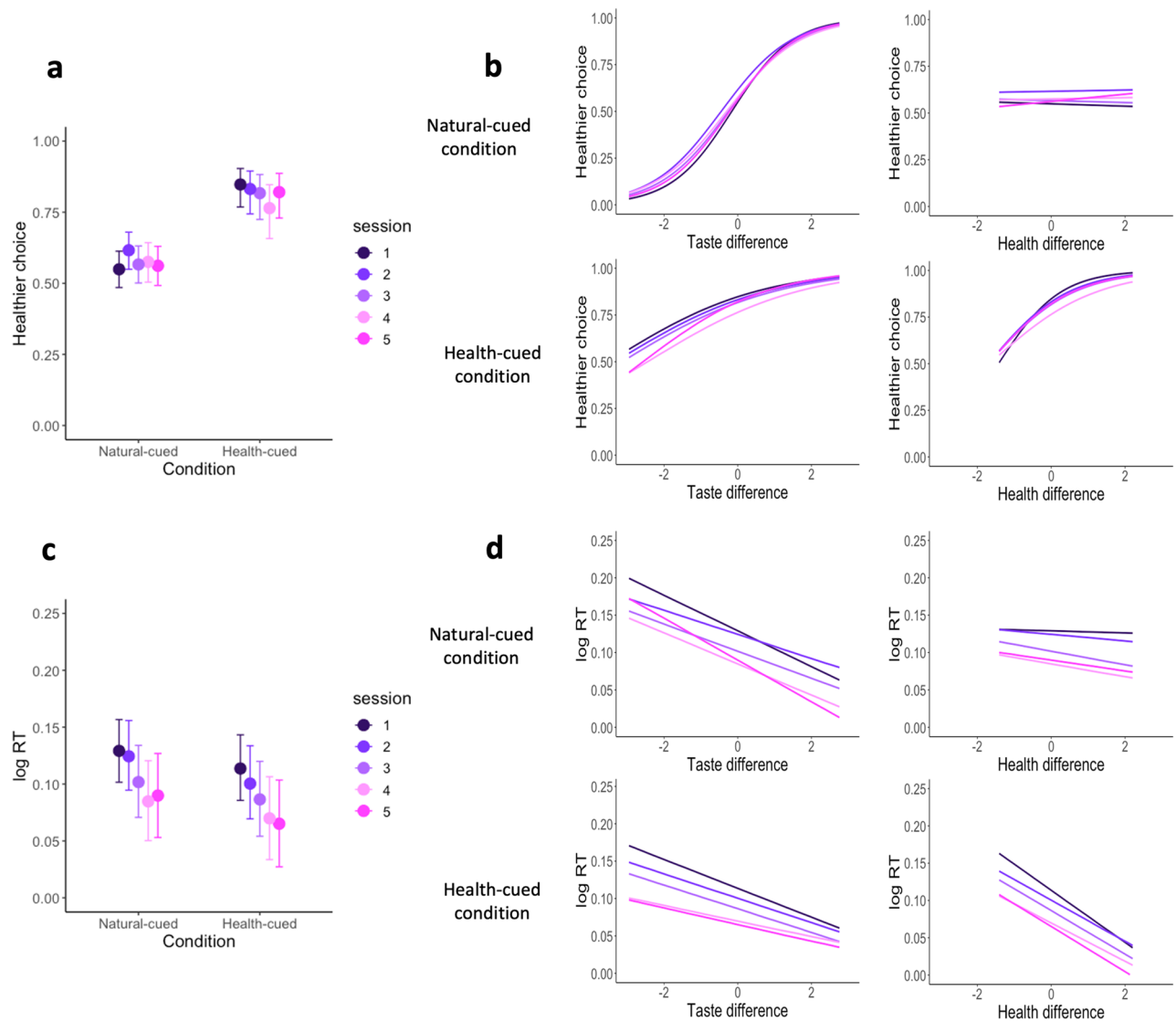


Figure 2. Effects of health-cues and attribute differences on choice outcomes and RTs over the five sessions. (a,b) Shows the proportion of healthier choices on the y-axis as a function of condition (health-cued, natural-cued choice) or attribute differences (computed as healthier item minus less-healthy item) on the x-axis. The results in each session are indicated by the separate colors shown in the legend. (a) The panel shows that healthier choices were higher in the health-cued condition compared to the natural-cued condition in all five sessions. (b) The panel shows how the two attributes (taste and healthiness) relate to choice outcomes in each of the two conditions (natural-cued and health-cued). (c,d) are analogous to (a) and (b) except that they show the logarithm of response times on the y-axis as a function of condition or attribute differences on the x-axis. The error bars in both (a) and (c) represent the 95% HDIs.

Participants responded faster when differences in health were larger (health difference interaction effect = -0.03 , 95% HDI = $[-0.05, -0.02]$). In contrast, the differences between taste attributes were significantly associated with response times in both natural and health-cued trials, with larger differences in taste leading to faster response time in both conditions (taste difference main effect = -0.02 , 95% HDI = $[-0.03, -0.01]$). In addition to the interaction between the health-cued condition and health ratings, there was also a main effect condition. Response times were significantly faster overall in the health-cued relative to the natural-cued condition (main effect of health-cue = -0.02 , 95% HDI = $[-0.03, 0.00]$). The response time patterns in both conditions remained fairly stable across all five sessions, although participants became faster with more experience in the task and there were small changes in the sensitivity to taste and healthiness differences (Tables S2 and S3).

Time-varying diffusion decision model analysis. We used a time-varying diffusion decision model (tDDM) to examine if repeated experience with the cued-attribute food choice task changed the evidence accumulation process assumed to occur during food decisions. The specific tDDM that we used allows for each attribute (here taste and health) to begin to influence the decision process at a different time^{36–38}. We refer to this model as the relative-starting-time DDM (rstDDM). The RST parameter represents the relative advantage

| Parameters | Session 1 | Session 2 | Session 3 | Session 4 | Session 5 |
|-------------------------------|--------------|--------------|--------------|--------------|--------------|
| Natural-cued condition | | | | | |
| Drift bias | 0.02 ± 0.08 | −0.03 ± 0.1 | 0.02 ± 0.11 | 0.08 ± 0.17 | −0.03 ± 0.13 |
| Health weight | −0.13 ± 0.31 | 0.05 ± 0.48 | 0.03 ± 0.47 | −0.01 ± 0.62 | 0 ± 0.49 |
| Taste weight | 0.89 ± 0.48 | 0.69 ± 0.28 | 0.73 ± 0.31 | 0.72 ± 0.39 | 0.87 ± 0.49 |
| RST | 0.18 ± 0.08 | 0.14 ± 0.13 | 0.02 ± 0.02 | 0.04 ± 0.03 | 0.1 ± 0.06 |
| NDT | 0.6 ± 0.16 | 0.59 ± 0.09 | 0.58 ± 0.15 | 0.54 ± 0.11 | 0.55 ± 0.09 |
| SP bias | 0.49 ± 0.03 | 0.5 ± 0.04 | 0.5 ± 0.03 | 0.48 ± 0.04 | 0.5 ± 0.04 |
| Noise | 1.02 ± 0.13 | 1.02 ± 0.14 | 1.08 ± 0.13 | 1.08 ± 0.16 | 1.09 ± 0.12 |
| Health-cued condition | | | | | |
| Drift bias | 0.05 ± 0.07 | 0.07 ± 0.13 | 0.01 ± 0.12 | 0.08 ± 0.07 | −0.05 ± 0.09 |
| Health weight | 0.9 ± 0.46 | 0.82 ± 0.55 | 0.8 ± 0.5 | 0.68 ± 0.65 | 0.8 ± 0.63 |
| Taste weight | 0.32 ± 0.32 | 0.29 ± 0.23 | 0.3 ± 0.29 | 0.33 ± 0.31 | 0.32 ± 0.44 |
| RST | −0.03 ± 0.02 | −0.04 ± 0.02 | −0.03 ± 0.03 | −0.02 ± 0.02 | −0.05 ± 0.04 |
| NDT | 0.59 ± 0.13 | 0.57 ± 0.1 | 0.53 ± 0.15 | 0.54 ± 0.1 | 0.52 ± 0.1 |
| SP bias | 0.49 ± 0.03 | 0.48 ± 0.04 | 0.49 ± 0.04 | 0.48 ± 0.04 | 0.51 ± 0.04 |
| Noise | 0.98 ± 0.13 | 1.01 ± 0.12 | 1 ± 0.09 | 1.08 ± 0.12 | 1.03 ± 0.13 |

Table 1. rstDDM Parameters. This table reports the group-level estimates (mean ± standard error (SE)) for each of the rstDDM parameters over 5 sessions in both the natural-cued and health-cued conditions. The parameters are: stochastic component of evidence accumulation (*noise*), starting point bias (*SP bias*; a value of 0.5 indicate no starting point bias), non-decision time (*NDT*), relative start time for health (*RST*; computed as starting time for health minus starting time for taste), weight of taste attribute on drift rate (*taste*), weight of health attribute on drift rate (*health*), and intercept for drift rate (*Drift bias*). The decision thresholds were held fixed at $[-1, 1]$ when fitting the model.

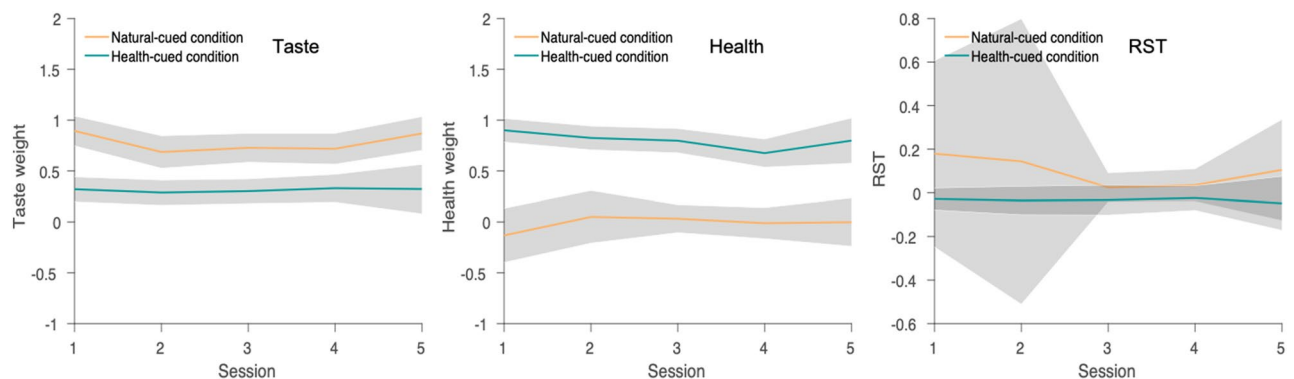


Figure 3. Group-level rstDDM parameters by condition across sessions. These three plots show the overall trends for the taste weight, health weight, and RST parameters in the natural-cued and health-cued conditions across sessions. The values for the health and taste weights are given in arbitrary units, while the RST parameter is specified in seconds. The taste weight was higher in the natural-cued condition compared to the health-cued condition in all sessions (left panel). The health weight was higher in the health-cued condition compared to the natural-cued condition in all sessions (middle panel). The RST parameter was qualitatively lower (i.e. healthiness was considered earlier) in the health-cued condition compared to the natural-cued condition in all sessions (right panel), although there was substantial individual variability in the RST parameter during the natural-cued trials. All group-level parameters other than the RST parameter for natural-cued trials were quite stable and did not differ with repeated experience across sessions. The shaded bars indicate the 95% HDIs for each parameter. The natural-cued condition is indicated by orange color and health-cued condition is indicated by green color.

in initial processing time for the taste attributes. If RST is positive, then taste-related attributes are considered before healthiness attributes, whereas if RST is negative then healthiness attributes are considered first. The drift weighting parameters in the rstDDM indicate the relative contribution of taste and health to the evidence accumulation process and decision outcome. Figure 3 shows the drift weighting and RST parameters by session and condition (see Table 1 for the full set of parameters).

Health cues had a significant and stable influence on rstDDM parameters across all five sessions. On average, participants weighted taste attributes more and considered them sooner than health attributes in the natural-cued condition during the first-session (taste weight: 0.8940, 95% HDI = [0.8200, 0.9660], health weight: −0.132, 95%

HDI = $[-0.275, -0.009]$, rst : 0.1790, 95% HDI = $[0.021, 0.448]$, and this pattern held across all sessions (Fig. 3). However, during the health-cued trials health weights were higher than in the natural trials (0.8120, 95% HDI = $[0.7520, 0.8780]$), while in contrast taste weights showed the opposite pattern and were significantly higher in the natural-cued condition relative to the health-cued condition (0.4670, 95% HDI = $[0.4180, 0.5150]$). Comparing the health and taste weights within each condition revealed that the health weight was significantly higher than taste weight in the health-cued condition (0.4870, HDI = $[0.4290, 0.5430]$), and taste weight were significantly higher than health weight in the natural-cued condition (0.7910, HDI = $[0.7240, 0.8620]$). The total contribution of taste plus health attributes to the drift rate was significantly higher in the health-cued than natural-cued condition (0.3450, 95% HDI = $[0.2760, 0.4130]$), which is consistent with the pattern of faster mean RTs in the health-cued trials shown in Fig. 2. In addition to weights, health cues also significantly changed the relative-starting times for each attribute within the decision process. In the first session, the RST was less than zero (i.e., healthiness was considered earlier than tastiness) in the health-cued condition, while it was greater than zero in the natural-cued condition (i.e., tastiness was considered earlier healthiness). Moreover, the RST parameter was significantly smaller in the health-cued compared to natural-cued trials (-0.1310 , 95% HDI = $[-0.2680, -0.0530]$). Health-cues continued to promote the consideration of health attributes before taste across all five sessions. Across sessions, the consideration start times for healthiness and tastiness in natural cued trials became more similar (i.e. RST was closer to 0), and less variable across individuals (Fig. 3).

In general, individual participants' rstDDM parameters were fairly stable across sessions. Figure 4 shows that the ranges ($\text{max} - \text{min}$) for the parameter estimates across the five sessions were fairly small, with the notable exception of the RST for natural-cued trials. Comparisons across conditions showed that the ranges were significantly smaller in the health-cued condition than in the natural-cued condition (Taste: $tstat = -3.1051$, $pvalue = 0.0052$, Health: $tstat = -2.4076$, $pvalue = 0.0249$, RST: $tstat = -14.5627$, $pvalue = 8.8958e - 13$).

Test-retest reliability at the individual level. Having found that group average decisions within the cued-attribute food choice task are fairly stable over 14 days and 5 repetitions of the task, we next tested the reliability at the individual level. Specifically, we computed the intra-class correlation coefficients (ICC), for three different task measures: 1) frequency of healthier choices, 2) subjective ratings of taste and healthiness, and 3) rstDDM parameters (health/taste weight, and RST). In the text below we interpret the ICC values as: Poor (< 0.40), Fair ($0.4 - 0.6$), Good ($0.6 - 0.75$), and Excellent ($0.75 - 1.0$)³⁹.

Individuals' choices showed excellent test-retest reliability across the five sessions. The frequency of healthier choices showed excellent reliability across the 5 sessions both within the natural- and health-cued conditions separately and as a difference score between conditions. (natural-cued condition, $ICC(A, 1) = 0.808$, $F(22, 86.8) = 23.3$, $p = 1.74e - 27$, 95%-Confidence Interval: $0.688 < ICC < 0.901$, health-cued condition, $ICC(A, 1) = 0.812$, $F(22, 91.9) = 22.6$, $p = 6.06e - 28$, 95%-Confidence Interval: $0.695 < ICC < 0.903$, difference between conditions, $ICC(A, 1) = 0.812$, $F(22, 89) = 23.5$, $p = 4.79e - 28$, 95%-Confidence Interval: $0.694 < ICC < 0.902$).

The test-retest reliability for participants' subjective healthiness and tastiness ratings was excellent and fair, respectively. Note that participants did not face the exact same choices in each session, but rather similar types of choices drawn from the same set of 180 food items. In addition to making choices, the participants rated each of the 180 food items for taste and healthiness in sessions 1 and 5. They rated a subset of 60 food items on taste and health in all five sessions. We computed the ICC both for full set of 180 and subset of 60 items. Rating reliability at the individual level was excellent for health and fair for taste ratings (60 items in all 5 sessions, health: 0.808 ± 0.171 , taste: 0.579 ± 0.184 , Table S6, 180 items in sessions 1 and 5, health: 0.778 ± 0.216 , taste: 0.532 ± 0.190 , Table S6). We also analyzed rating reliability as a function of food image rather than individual. At the food image level, we found that reliability was fair for both health and taste ratings (across five sessions: health: 0.552 ± 0.146 , taste: 0.574 ± 0.120 , Table S7, between session 1 and 5 alone: health: 0.468 ± 0.251 , taste: 0.516 ± 0.197 , Table S8).

Lastly, we computed the test-retest reliability of the rstDDM parameters fit to the participants' choice outcomes and response times. Table 2 reports the ICC values for each parameter by condition. The reliability of the parameters ranged from excellent to good in the health-cued condition. However, the reliability of the same parameters fit to the same individuals was only fair to poor in the natural-cued condition. As a consequence of the lower reliability of the parameters in the natural-cued condition, the reliability of the difference scores for taste and health weights is also lower.

Discussion

Previous research has shown that food choice tasks using photographic images as stimuli have good test-retest reliability when participants make choices naturally (i.e. without explicit instructions about how to choose or what attributes to consider)³⁵. However, unlike the previously tested paradigms, there are clear differences in the ways the same participant will make choices across the different conditions within the cued-attribute food choice task^{7,13,32,33}. Participants decision outcomes and response times are more strongly influenced by healthiness attributes and they choose the healthier item more often in health-cued compared to natural-cued trials. We tested whether these intra-individual differences changed over the course of 14 days and five repetitions of the cued-attribute food choice task. Overall, choice outcome patterns were stable and had excellent test-retest reliability within and across the decision conditions over the five task repetitions.

Interestingly, healthiness ratings had excellent reliability, but the reliability for taste ratings was only fair. This suggests that an individuals' estimates of the tastiness of food items may vary more from day to day, or even moment to moment than their opinions about healthiness. Previous work has found that liking or attractiveness ratings for food items, which are presumably closely tied to taste, vary from the first to second round of ratings

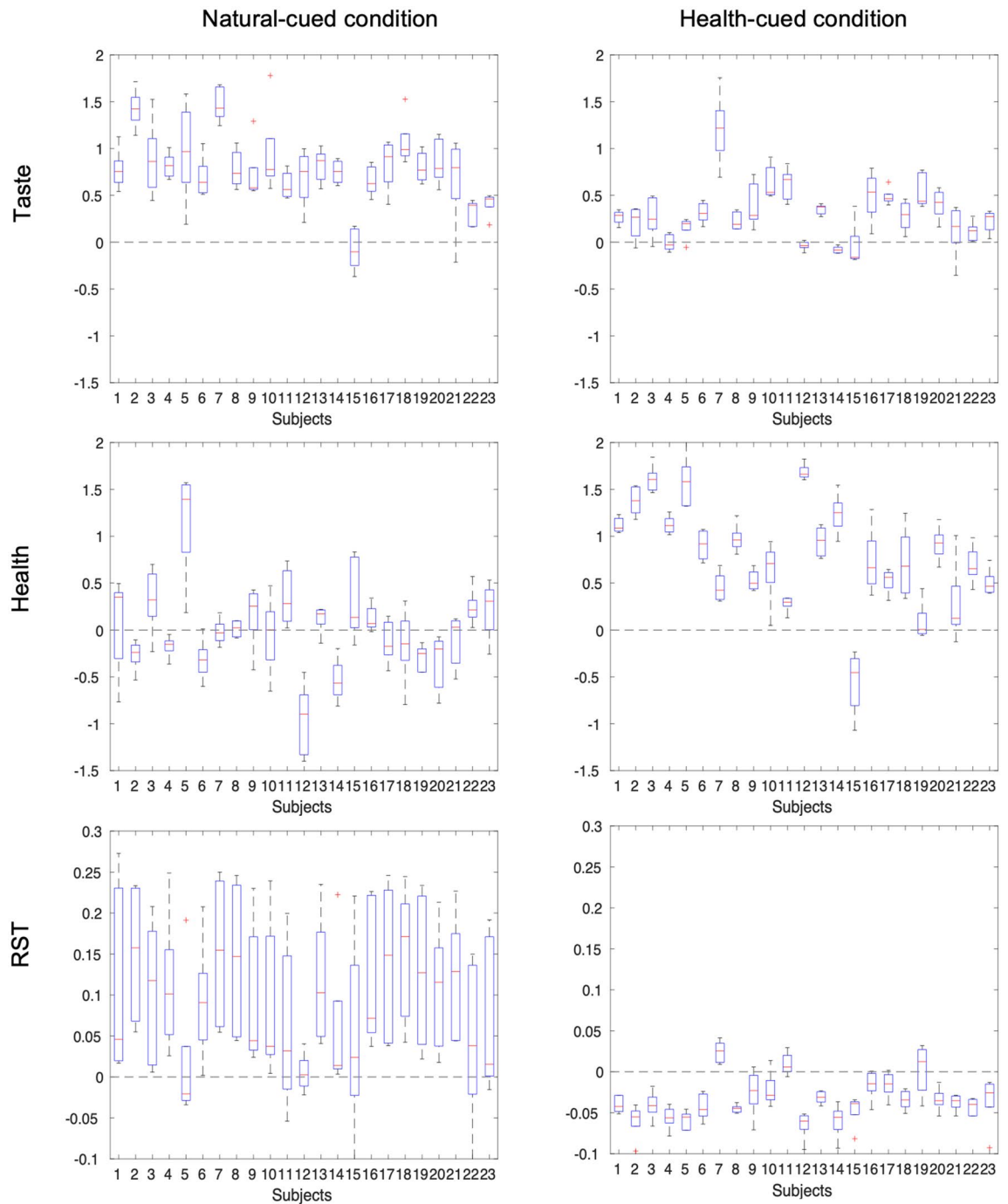


Figure 4. rstDDM parameters stability. The boxplots show that the values of most subject-level parameters didn't change much across the 5 sessions. The values for the health and taste weights are given in arbitrary units, while the RST parameter is specified in seconds. However, the changes in attribute weights on the drift rate over time/experience were significantly greater in the natural-cued condition compared to the health-cued condition. We address the high variability of the RST parameters in the natural-cued condition further in the discussion section.

within a given day^{40,41}. Palatability is more subjective than healthiness and may be influenced more by previously rated items, context effects, or noise during the process of retrieving the past consumption experiences that are thought to be used as a basis for present valuations⁴². Healthiness is not something that is directly experienced and thus may not be computed using samples from episodic memory the same way tastiness is presumed to be. While the current study was not designed to directly address the sources of consistency and variability within and across different food attributes, this is an important avenue for future research.

| Parameter | Natural-cued condition | Health-cued condition | Health—natural |
|---------------|------------------------|-----------------------|----------------|
| Drift bias | − 0.0149 | − 0.0679 | − 0.129 |
| Health weight | 0.592 | 0.841 | 0.784 |
| Taste weight | 0.53 | 0.696 | 0.541 |
| RST | 0.095 | 0.58 | 0.0655 |
| NDT | 0.447 | 0.625 | 0.0277 |
| SP bias | 0.0054 | − 0.0368 | − 0.0264 |
| Noise | 0.522 | 0.477 | − 0.00835 |

Table 2. Intra-class correlation coefficients (ICC). ICC of rstDDM parameters over 5 sessions for both conditions of natural-cued and health-cued and also for the difference between these two conditions. *NDT* non-decision time.

Although choice outcomes had excellent reliability in both the natural and health-cued task conditions, the reliability of rstDDM parameters differed across conditions. A much stronger reliance on tastiness than healthiness attributes may have led to the lower test-retest reliability for rstDDM parameters fit to the natural-cued choices. Both the rstDDM and the logistic regression analyses showed that, on average, healthiness had little influence on food choice outcomes during natural-cued choices in our sample of participants. Figures 2b and 3 show that, holding tastiness fixed, the influence of healthiness on food decision is approximately zero at the group level. This is not too surprising given that the individuals were recruited because they consumed sweet or savory snacks on a regular basis, but attempting to eat a healthy diet in daily life was not an inclusion criterion. Thus, concerns about the palatability of the food items dominated the decision process in natural trials. The near-zero influence of healthiness on food choices in natural trials means that the rstDDM has little power to accurately identify the relative attribute weighting and consideration start timing parameters (*RST*).

The rstDDM model we fit to the data can accurately identify and distinguish between attribute weights and *RST* parameters when both weights are non-zero^{36,38}, but (near) zero weights are a problem for it. Either a sufficiently low drift weighting coefficient or long delay in consideration starting time can effectively eliminate the influence of an attribute on choice outcomes within the rstDDM framework. The excellent reliability of the drift weighting parameters and good reliability of the *RST* parameter in the health-cued trials proves that the rstDDM can yield highly reliable parameter estimates if fit to a suitable set of data. Thus, the lower reliability of the rstDDM in the natural-cued trials serves to demonstrate the importance of checking that the data conform to the expectations and requirements of a given modeling approach. If one of the drift weights from the rstDDM is close to zero for a given data set, then one is probably better off fitting the more common version of the DDM⁴³, which fixes the *RST* parameter to zero and only estimates the drift weights (see Table S9 for the full set of parameters, and Table S10 for the reliability values).

Given its high level of test-retest reliability, the cued-attribute food choice task appears to be well suited to serve as an outcome measure for cognitive, physical, or pharmacological interventions that target food consumption decision processes. This task may be useful as an outcome measure because examines decisions with and without an explicit cue/instruction to consider healthiness aspects during the choice. The health cues cause participants to make more choices in favor of the healthier option and exhibit brain activity pattern similar to those observed in participants that exert self-control over dietary choices endogenously^{7,13,19,32,33}. Although, there is a substantial increase in the proportion of healthier choices made in the health-cued condition, on average, participants do not always choose the healthier item in these trials. Thus, the task could be used to determine if an intervention increased cued dietary self-regulation. In addition, the natural, uncued trials allow experimenters to assess the baseline, unprompted choice behavior. In this case, it provides insight in to endogenous self-regulation. It may be that a specific intervention eliminates the difference between health-cued and natural trials by bringing the natural attention and decision mechanisms in line with those prompted by the cues. Ideally, an intervention designed to promote healthier food choices would cause its participants to make healthy choices at near-ceiling rates in both task conditions. If it turns out that an intervention boosts healthy choices in one condition, but not the other, then we would know that the intervention is more effective for externally vs internally cued regulation or vice versa. We think these features of the cued-attribute food choice task and its high level of test-retest reliability recommend it as one of a suite of intervention outcome measures. However, it is important to note that here we have only tested healthy young adults. Further examinations of test-retest reliability in individuals with obesity, eating disorders, and other conditions are warranted.

The cued-attribute food choice task explicitly prompts participants to incorporate health-related attributes into their decision process. In combination with functional magnetic resonance imaging, the health-cued condition within this task has been shown to engage prefrontal brain activity patterns similar to those observed during endogenous dietary self-control in individuals who successfully lost or maintained their desire weight^{7,19,32,33}. Thus, the cued attribute food choice task may be a means of testing how well an intervention shifts externally cued and endogenous health-promoting decision processes.

Methods

Participants. Twenty-three participants completed the five sessions of the experiment across 14 days. All the participants (10 female, ages between 18 and 30 years) gave written informed consent for their participation in accordance with the regulations of the Zurich Cantonal Ethics commission. They received a flat fee in addition

to their food reward for the time they spent for the task. We screened participants by email and phone to ensure that they did not follow any specific food diet, and they consumed sweet and/or savory snacks regularly. We also rechecked the exclusion criteria on each day of testing. All participants were healthy and without any current or recent psychiatric, metabolic or neurological illness. On the day of the study, participants sent us a photograph of the meal they consumed 3 hours before their appointment as an indication that they followed our instructions to eat a small meal and then fast for 3 hours before coming into the lab. All five laboratory visits took place between 17:15 and 19:15. The first visit occurred on a Wednesday, visits 2:3 took place on the Monday, Wednesday, and Friday of the following week. The fifth and final visit took place on a Wednesday two weeks after the initial visit. All procedures were approved by and performed in accordance with the guidelines and regulations of the Zurich Cantonal Ethics commission.

Behavioral task. Participants were asked to eat a small meal three hours before their appointment and consume nothing but water in the meantime, in order to be hungry and thereby increase the value of the foods during the experiment. The task involved two sequential phases of food rating and food choice, which participants performed in all 5 sessions. Before each phase, subjects performed a short training session to become familiar with the task. In the food rating phase, participants judged how tasty or healthy each food item is in two different blocks. The ratings were done on continuous scale with anchors of -5 and $+5$ at each end. The order of taste and health rating blocks were counterbalanced, and the order of images within each block was randomized. In the first and last sessions (1 and 5), the ratings were done over all 180 food images, but in the middle sessions (2–4), the ratings were done for a subset of 60 food images whose average ratings in previous studies spanned the total ranges for taste and healthiness. We reduced the size of the rating set in order to save time in sessions 2:4 and maximize participant retention. We used the healthiness and taste ratings provided in first session when analyzing the choice data in all 5 sessions.

In the food choice phase, subjects had to choose which of two food items they would eat at the end of the experiment. The food pairs were constructed so that choosing the healthier item often required forgoing the subjectively tastier one, and we refer to these cases as challenge trials. The food pairings of all 5 sessions for each participant were constructed based on the ratings of his or her first session. One of the participant's decisions were randomly selected and implemented after the task. There was a 30-minute waiting period following the end of the session during which participants ate their selected food item. This 30-minute waiting period made the decisions more meaningful because participants did not have access to other food sources at that time, and they had not eaten for approximately 4.5 hours at that point.

Within the choice phase, there were two types of conditions that differed in the attention cues that are provided for the participants. In the health-cue condition, subjects were cued to consider the healthiness of the foods while making decisions. In this condition they were supposed to choose the healthier of the foods as often as they could, while keeping in mind that they would potentially have to eat the food they choose. In the natural-cue condition, subjects were cued to make decisions naturally based on whatever freely came to their minds. The food choice task consisted of 3 runs, with each run consisting of 5 or 6 blocks, 210 trials in total. The order of condition blocks was pseudo-randomized across subjects. Each subject faced 9 health-cue condition blocks (110.6 ± 5.2 trials) and 8 natural-cue condition blocks (99.3 ± 5.2 trials).

Choices were presented on the screen for 3 s and there was a jittered inter-trial interval was of 2–6 s. After the task was completed, one trial randomly was selected to be realized for each participant according to his or her choices. The participants stayed in the laboratory for 30 min to eat their food reward.

Computational tools. *Logistic regression for choice outcome.* We ran a Bayesian hierarchical logistic regression analysis with the *brms* package^{44,45} in *R*⁴⁶, to estimate the participants choice. We used the uninformative priors default in the *brms* package. We estimated the probability of choosing the healthier food option, as a function of the difference in taste ratings (td) and health ratings (hd) between the healthier and less-healthy options, condition (natural-, and health-cued conditions), and session (5 sessions). The taste difference (td) is calculated by the taste of the more healthy option minus the taste of the less healthy option. The health difference (hd) is also computed in the same way. The model formula in *brms* syntax was:

$$\begin{aligned} \text{choice}(\text{healthier}) \sim & 1 + (\text{hd} + \text{td}) * \text{condition} * \text{session} \\ & + (1 + (\text{hd} + \text{td}) * \text{condition} * \text{session}) | \text{subject} \end{aligned} \quad (1)$$

Note that asterisks indicate interaction terms and thus the model included all main effects and interactions between regressors except for interactions between taste and health differences. The first line lists the regressors and interactions that were estimated at the group or population level, and the final portion on the second lined inside parentheses shows those that were estimated at the subject level. In this case, all terms were estimated at both levels.

Linear regression for response time. We ran another Bayesian hierarchical regression analysis with the *brms* package in *R*, to estimate the response times (logarithm) as a function of difference in taste ratings and health ratings between two competing options, condition (natural-, and health-cued conditions), session (5 sessions), challenge type (when there is a conflict between taste and health and they have similar ratings). The health and taste ratings were z-scored, and all other variables were categorical. We used the uninformative priors default in the *brms* package. The model in *brms* syntax was:

$$\log(RT) \sim 1 + (hd + td) * condition * session * challenge + (1 + (hd + td) * condition * session * challenge|subject) \quad (2)$$

Note that asterisks indicate interaction terms and thus the model included all main effects and interactions between regressors except for interactions between taste and health differences. The first line lists the regressors and interactions that were estimated at the group or population level, and the final portion on the second lined inside parentheses shows those that were estimated at the subject level. In this case, all terms were estimated at both levels.

Sequential sampling model. We used the relative-starting-time DDM (rstDDM) model to estimate the response times and choice outcomes of the participants. This model was previously tested and validated for binary food choices³⁶. This analyses have been implemented in R⁴⁶. In this work, we used the hierarchical Bayesian fitting procedure described here: https://github.com/galombaridi/method_HtSSM_aDDM. We fit each session and condition separately. The following seven free parameters were estimated in each fitting: stochastic component of evidence accumulation (*noise*), starting point bias (*SP bias*), non-decision time (*NDT*), relative start time for health (*RST*), weight of taste attribute on drift rate (ω_{taste}), weight of health attribute on drift rate (ω_{health}), and intercept for drift rate (*Drift bias*). The decision thresholds were fixed to $[-1, 1]$. Note that the target variable in this analysis is the choosing left versus right option.

Test-retest reliability(ICC). To determine the test-retest reliability of the task across the five sessions, we computed intra-class correlation coefficients (ICC). We computed ICCs for three different task dimensions including frequency of healthier choices, and DDM parameters (focusing on drift bias, health/taste weight, and RST), and participants' ratings (health and taste). All of the variables were continuous, therefore we used the *icc* function in the *irr* library in R⁴⁷ with the following configuration: *icc(data, model = "twoway", type = "agreement", unit = "single", r0 = 0, conf.level = 0.95)*. We follow Shrout and Fleiss³⁹ and interpret the ICC values according to the following convention: Poor (< 0.40), Fair (0.4 – 0.6), Good (0.6 – 0.75), and Excellent (0.75 – 1.0).

Data availability

The datasets analyzed for this study can be found in the OSF repository, and the corresponding analysis code is available at this Github repository.

Received: 3 March 2021; Accepted: 2 June 2021

Published online: 18 June 2021

References

1. Symmank, C. *et al.* Predictors of food decision making: A systematic interdisciplinary mapping (SIM) review. *Appetite* **110**, 25–35. <https://doi.org/10.1016/j.appet.2016.11.023> (2017).
2. Leng, G. *et al.* The determinants of food choice. *Proc. Nutr. Soc.* **76**, 316–327. <https://doi.org/10.1017/S002966511600286X> (2017).
3. Apolzan, J. W. *et al.* Long-term weight loss with metformin or lifestyle intervention in the diabetes prevention program outcomes study. *Ann. Intern. Med.* **170**, 682–690. <https://doi.org/10.7326/M18-1605> (2019).
4. Walker, L. A., Chambers, C. D., Veling, H. & Lawrence, N. S. Cognitive and environmental interventions to encourage healthy eating: Evidence-based recommendations for public health policy. *Royal Society Open Science* **6**, 190624. <https://doi.org/10.1098/rsos.190624> (2019).
5. Val-Laillet, D. *et al.* Neuroimaging and neuromodulation approaches to study eating behavior and prevent and treat eating disorders and obesity. *NeuroImage* **8**, 1–31 (2015).
6. Plassmann, H., O'Doherty, J. & Rangel, A. Orbitofrontal cortex encodes willingness to pay in everyday economic transactions. *J. Neurosci.* **27**, 9984–9988. <https://doi.org/10.1523/JNEUROSCI.2131-07.2007> (2007).
7. Hare, T. A., Malmaud, J. & Rangel, A. Focusing attention on the health aspects of foods changes value signals in vmPFC and improves dietary choice. *J. Neurosci.* **31**, 11077–11087 (2011).
8. Smeets, P. A. M. *et al.* Good practice in food-related neuroimaging. *Am. J. Clin. Nutr.* **109**, 491–503. <https://doi.org/10.1093/ajcn/nqy344> (2019).
9. de Araujo, I. E., Schatzker, M. & Small, D. M. Rethinking food reward. *Annu. Rev. Psychol.* **71**, 139–164. <https://doi.org/10.1146/annurev-psych-122216-011643> (2020) [arXiv:31561741](https://arxiv.org/abs/31561741).
10. Sullivan, N. J., Fitzsimons, G. J., Platt, M. L. & Huettel, S. A. Indulgent foods can paradoxically promote disciplined dietary choices. *Psychol. Sci.* **30**, 273–287. <https://doi.org/10.1177/0956797618817509> (2019).
11. Krajbich, I., Armel, C. & Rangel, A. Visual fixations and the computation and comparison of value in simple choice. *Nat. Neurosci.* **13**, 1292–1298 (2010).
12. Brunstrom, J. M. & Rogers, P. J. How many calories are on our plate? Expected fullness, not liking, determines meal-size selection. *Obesity* **17**, 1884–1890. <https://doi.org/10.1038/oby.2009.201> (2009).
13. Hege, M. A. *et al.* Eating less or more: Mindset induced changes in neural correlates of pre-meal planning. *Appetite* **125**, 492–501. <https://doi.org/10.1016/j.appet.2018.03.006> (2018).
14. Rangel, A. Regulation of dietary choice by the decision-making circuitry. *Nat. Neurosci.* **16**, 1717–1724. <https://doi.org/10.1038/nn.3561> (2013).
15. van der Laan, L. N., de Ridder, D. T. D., Viergever, M. A. & Smeets, P. A. M. The first taste is always with the eyes: A meta-analysis on the neural correlates of processing visual food cues. *NeuroImage* **55**, 296–303. <https://doi.org/10.1016/j.neuroimage.2010.11.055> (2011).
16. Cornil, Y. *et al.* Obesity and Responsiveness to Food Marketing Before and After Bariatric Surgery. *J. Consum. Psychol.* <https://doi.org/10.1002/jcpy.1221> (2021).
17. Kramani, Q., Krajbich, I., Enax, L., Brustkern, L. & Weber, B. Salient nutrition labels shift peoples' attention to healthy foods and exert more influence on their choices. *Nutrition Research* (2020).
18. Ha, O.-R. *et al.* Healthy eating decisions require efficient dietary self-control in children: A mouse-tracking food decision study. *Appetite* **105**, 575–581 (2016).

19. Hare, T. A., Camerer, C. F. & Rangel, A. Self-control in decision-making involves modulation of the vmPFC valuation system. *Science* **324**, 646–648 (2009).
20. Foerde, K., Steinglass, J. E., Shohamy, D. & Walsh, B. T. Neural mechanisms supporting maladaptive food choices in anorexia nervosa. *Nat. Neurosci.* **18**, 1571–1573. <https://doi.org/10.1038/nn.4136> (2015).
21. van Meer, F. *et al.* Development and body mass inversely affect children's brain activation in dorsolateral prefrontal cortex during food choice. *NeuroImage* **201**, 116016. <https://doi.org/10.1016/j.neuroimage.2019.116016> (2019).
22. Uniacke, B. *et al.* A comparison of food-based decision-making between restricting and binge-eating/purging subtypes of anorexia nervosa. *Int. J. Eat. Disord.* **53**, 1751–1756 (2020).
23. Lim, S.-L., Penrod, M. T., Ha, O.-R., Bruce, J. M. & Bruce, A. S. Calorie labeling promotes dietary self-control by shifting the temporal dynamics of health-and taste-attribute integration in overweight individuals. *Psychol. Sci.* **29**, 447–462 (2018).
24. Demos, K. E. *et al.* Identifying the mechanisms through which behavioral weight-loss treatment improves food decision-making in obesity. *Appetite* **114**, 93–100. <https://doi.org/10.1016/j.appet.2017.03.013> (2017).
25. Kohl, S. H. *et al.* Real-time fMRI neurofeedback training to improve eating behavior by self-regulation of the dorsolateral prefrontal cortex: A randomized controlled trial in overweight and obese subjects. *NeuroImage* **191**, 596–609. <https://doi.org/10.1016/j.neuroimage.2019.02.033> (2019).
26. McDermott, K. E. D., Lillis, J., McCaffery, J. M. & Wing, R. R. Effects of cognitive strategies on neural food cue reactivity in adults with overweight/obesity. *Obesity* **27**, 1577–1583. <https://doi.org/10.1002/oby.22572> (2019).
27. Boswell, R. G., Sun, W., Suzuki, S. & Kober, H. Training in cognitive strategies reduces eating and improves food choice. *Proc. Nat. Acad. Sci. USA* **115**, E11238–E11247. <https://doi.org/10.1073/pnas.1717092115> (2018).
28. Sun, W. & Kober, H. Regulating food craving: From mechanisms to interventions. *Physiol. Behav.* **112878**, <https://doi.org/10.1016/j.physbeh.2020.112878> (2020).
29. Schonberg, T. *et al.* Changing value through cued approach: An automatic mechanism of behavior change. *Nat. Neurosci.* **17**, 625–630. <https://doi.org/10.1038/nn.3673> (2014).
30. Salomon, T. *et al.* The cue-approach task as a general mechanism for long-term non-reinforced behavioral change. *Sci. Rep.* **8**, 3614. <https://doi.org/10.1038/s41598-018-21774-3> (2018).
31. Gobbi, S. *et al.* Reduced neural satiety responses in women affected by obesity. *Neuroscience* **447**, 94–112. <https://doi.org/10.1016/j.neuroscience.2020.07.022> (2020).
32. van Meer, F., van der Laan, L. N., Viergever, M. A., Adan, R. A. H. & Smeets, P. A. M. Considering healthiness promotes healthier choices but modulates medial prefrontal cortex differently in children compared with adults. *NeuroImage* **159**, 325–333. <https://doi.org/10.1016/j.neuroimage.2017.08.007> (2017).
33. Schmidt, L. *et al.* Neuroanatomy of the vmPFC and dlPFC predicts individual differences in cognitive regulation during dietary self-control across regulation strategies. *J. Neurosci.* **38**, 5799–5806 (2018).
34. Lee, D. & Holyoak, K. Coherence shifts in attribute evaluations. *bioRxiv* (2020).
35. Foerde, K. *et al.* Assessment of test-retest reliability of a food choice task among healthy individuals. *Appetite* **123**, 352–356. <https://doi.org/10.1016/j.appet.2018.01.010> (2018).
36. Maier, S. U., Raja Beharelle, A., Polania, R., Ruff, C. C. & Hare, T. A. Dissociable mechanisms govern when and how strongly reward attributes affect decisions. *Nat. Hum. Behav.* **4**, 949–963. <https://doi.org/10.1038/s41562-020-0893-y> (2020).
37. Sullivan, N., Hutcherson, C., Harris, A. & Rangel, A. Dietary self-control is related to the speed with which attributes of healthfulness and tastiness are processed. *Psychol. Sci.* **26**, 122–134. <https://doi.org/10.1177/0956797614559543> (2015).
38. Amasino, D. R., Sullivan, N. J., Kranton, R. E. & Huettel, S. A. Amount and time exert independent influences on intertemporal choice. *Nat. Hum. Behav.* **3**, 383–392. <https://doi.org/10.1038/s41562-019-0537-2> (2019).
39. Shrout, P. E. & Fleiss, J. L. Intraclass correlations: Uses in assessing rater reliability. *Psychol. Bull.* **86**, 420 (1979).
40. Chen, Z., Veling, H., Dijksterhuis, A. & Holland, R. W. How does not responding to appetitive stimuli cause devaluation: Evaluative conditioning or response inhibition? *J. Exp. Psychol. Gen.* **145**, 1687–1701. <https://doi.org/10.1037/xge0000236> (2016).
41. Polania, R., Woodford, M. & Ruff, C. C. Efficient coding of subjective value. *Nat. Neurosci.* **22**, 134–142 (2019).
42. Shadlen, M. N. & Shohamy, D. Decision Making and Sequential Sampling from Memory. *Neuron* **90**, 927–939. <https://doi.org/10.1016/j.neuron.2016.04.036> (2016).
43. Forstmann, B., Ratcliff, R. & Wagenmakers, E.-J. Sequential sampling models in cognitive neuroscience: Advantages, applications, and extensions. *Annu. Rev. Psychol.* **67**, 641–666. <https://doi.org/10.1146/annurev-psych-122414-033645> (2016).
44. Bürkner, P.-C. brms: An R package for Bayesian multilevel models using Stan. *J. Stat. Softw.* **80**, 1–28. <https://doi.org/10.18637/jss.v080.i01> (2017).
45. Bürkner, P.-C. Advanced Bayesian multilevel modeling with the R package brms. *The R Journal* **10**, 395–411. <https://doi.org/10.32614/RJ-2018-017> (2018).
46. R Core Team. *R: A Language and Environment for Statistical Computing*. (R Foundation for Statistical Computing, 2019).
47. Matthias Gamer, I. F. P. S., Jim Lemon. *irr: Various Coefficients of Interrater Reliability and Agreement*. R package version 0.84.1.

Acknowledgements

This work was supported by a grant from the Swiss National Science Foundation (SNSF; Grant number 32003B_166566).

Author contributions

A.R.B. and T.A.H. designed the study. A.R.B. and Z.B. acquired the data. Z.B. analyzed the data with input from A.R.B. and T.A.H. T.A.H. and Z.B. wrote the paper with input from A.R.B.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1038/s41598-021-91933-6>.

Correspondence and requests for materials should be addressed to Z.B.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2021