Coherence Shifts in Attribute Evaluations

Douglas Lee[1] Keith J. Holyoak[2]

[1] California Institute of Technology, Division of the Humanities and Social Sciences

[2] University of California, Los Angeles, Department of Psychology

E-mail for correspondence:

DouglasGLee@gmail.com

Word count: 6,539

Abstract

In two experiments, people repeatedly judged individual options with respect to both overall value and attribute values. When required to choose between two snacks, each differing in two attributes (pleasure and nutrition), people's assessments of value shifted from pre- to post-choice in the direction that spread the alternatives further apart so as to favor the winner, thereby increasing confidence in the choice. This shift was observed not only for ratings of overall value, but also for each of the two individual attributes. The magnitude of the coherence shift increased with choice difficulty as measured by the difference in initial ratings of overall value for the two options, as well as with a measure of attribute disparity (the degree to which individual attributes "disagree" with one another as to which option is superior). In Experiment 2, a task requiring comparison of value similarity generated the same qualitative pattern of value and confidence changes, even though the similarity judgments entirely preceded the actual choice task. These findings support the hypothesis that active comparison of values is the core component of the decision process during which perceived values are shifted so as to more sharply distinguish the items being compared.

Keywords: choice, value, coherence shifts, spreading of alternatives, confidence

Introduction

The traditional view of multi-attribute decision making assumes that when people choose between options, they compare them with respect to a set of decision-relevant attributes. According to one theoretical perspective, the attributes of each option are assessed individually, weighted according to their importance, and summed together to map onto an estimate of overall value for each option (e.g., Tversky & Simonson, 1993; Rieskamp, Busemeyer, & Mellers, 2006). Alternatively, attributes may be assessed and compared separately across options, with differences at the level of individual attributes being used to choose the preferred option (e.g., Tversky, 1969; Hunt, Dolan, & Behrens, 2014). In either case, attribute values constitute pre-existing and stable elements that serve as the basic inputs to the decision process.

An alternative perspective emphasizes the constructive nature of preference (e.g., Ariely, Loewenstein, & Prelec, 2006; DeKay, Stone, & Miller, 2011; Holyoak & Simon, 1999; Russo, Medvec, & Meloy, 1996; Payne, Bettman, & Schkade, 1999; Svenson, 1992; also see Lichtenstein & Slovic, 2006; Warren, McGraw, & Van Boven, 2011). Under this view, subjective assessments of attribute values and attribute importance can shift during the decision process so as to cohere with the emerging decision. These shifts will spread the perceived values of the options, increasing the advantage of the eventual winner relative to its competitor(s) and thereby increasing confidence in the decision. Such coherence shifts have been shown to impact decisions about such everyday matters as choosing a restaurant or an apartment to rent (Russo et al., 1996; Simon, Krawzcyk, & Holyoak, 2004; Simon, Krawczyk, Bleicher, & Holyoak, 2008), adjudicating legal disputes (Holyoak & Simon, 1999; Carlson & Russo, 2001; Simon, 2012), and evaluating complex issues with moral implications (Spellman, Ullman, & Holyoak, 1993; Holyoak & Powell,

2015; Simon, Stenstrom, & Read, 2015). Cognitive evaluations and emotional reactions can jointly undergo coherence shifts during decision making (Simon, Stenstrom, & Read, 2015). In contrast to classical cognitive dissonance theory (Festinger, 1957; see Harmon-Jones & Harmon-Jones, 2007, for a review), which claims that value shifts are post-decisional, the construction of preference emphasizes that coherence shifts play a critical role in the pre-decisional dynamics that drive the selection of an option (Simon & Holyoak, 2002). A number of studies have shown that coherence shifts are observed prior to making a firm decision or public commitment (Holyoak & Simon, 1999; Russo et al., 1996; Russo, Carlson, Meloy, & Yong, 2008; Russo & Chaxel, 2010; Simon et al., 2004; Simon, Pham, Le, & Holyoak, 2001).

Although the impact of coherence shifts on decision making has been firmly established, less is known about the intra-decisional dynamics that drive attribute reevaluation. The way in which people process information during decision making may be subject to metacognitive control (e.g., Chaxel, 2015), which may lead to *rational inattention* to different attributes in different choice contexts (Sims, 2003; Caplin & Dean, 2015). The magnitude of shifts in overall evaluations of options tends to increase with the difficulty of the decision (Svenson, 1992), thereby increasing confidence in the choice (Simon, Snow & Read, 2004; Simon & Spiller, 2016). However, previous studies relating choice difficulty to the magnitude of coherence shifts have not assessed potential changes in the perceived values of those individual attributes that are hypothesized to determine overall value. In the present paper we report two experiments that used a multi-attribute choice paradigm to examine coherence shifts in both overall evaluations and in perceived values of individual attributes. We expected that more difficult choices—those

in which two options are initially relatively close in overall value—will trigger larger coherence shifts both for overall value and for individual attributes.

In addition, we examined the impact of a different factor potentially related to choice difficulty. We hypothesized that decision dynamics may also be influenced by differences in attribute composition. Consider the two hypothetical choices situations depicted in Figure 1. Both choices involve two options based on two relevant and equally-weighted attributes, A1 and A2. The first choice is between option 1 with attribute values of 90 for A1 and 30 for A2, versus option 2 with values 20 and 80. The second choice is between option 3 with values 65 and 55, versus option 4 with values 45 and 55. Under the assumption of equal importance weights for A1 and A2, the two choices are identical at the level of overall value (60 for option 1 and option 3, 50 for option 2 and option 4). Thus option 1 is favored for the first choice, and option 3 for the second choice, by equal amounts. But notice that the disparity between the attributes is much greater for the first choice (option 1 versus option 2) than for the second (option 3 versus option 4).

We propose that the choice with higher disparity will result in larger coherence shifts and thus greater confidence. When choice options are disparate, one option dominates along one attribute dimension, whereas the other dominates along the other dimension. It has been argued that the more dominant is an attribute of a particular option (relative to another option or the average across all other options), the more attention it will receive during the deliberation leading to a decision (Bordalo, Gennaioli, & Shleifer, 2020, 2013). Such increased attention would be expected to enhance the processes that produce spreading of alternatives. An increased

coherence shift for high disparity choices would in turn increase choice confidence and decrease response time (RT), as the choice effectively becomes easier (Lee & Daunizeau, 2020a).

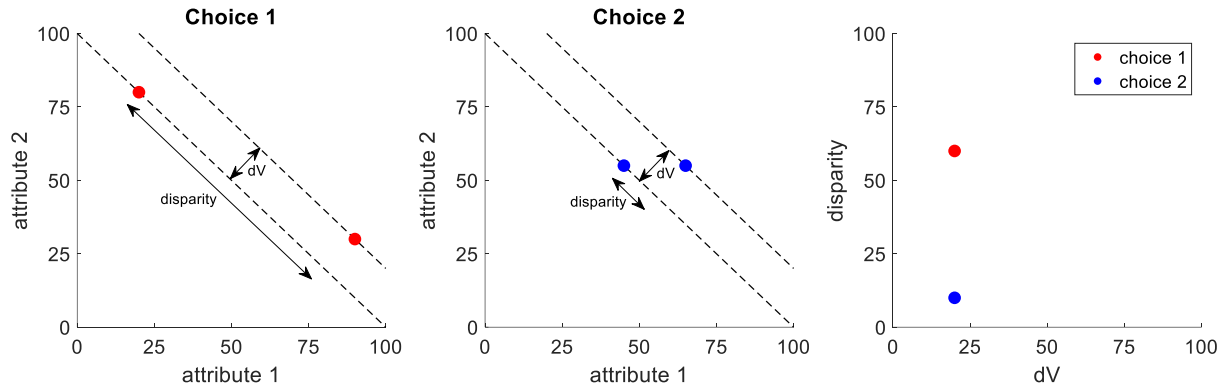


Figure 1. A schematic illustration of orthogonal components of choice difficulty: dV and disparity. The left plot illustrates a "high disparity" choice and the middle plot illustrates a "low disparity" choice. The red and blue dots represent the alternative options for each choice, each plotted according to its measurements on two attribute dimensions. The example assumes equal importance weights for each attribute, so the iso-value curves are represented by parallel lines with slope -1. The difference in overall value of the options, dV, is the distance between the iso-value curves on which the options lie. Disparity is the distance between the options in the dimension orthogonal to overall value (see Equation 1 below for a mathematical formulation). The right plot shows the location of each choice pair in the transformed dV-disparity space.

In addition to addressing the above issues, the present study attempted to identify the core component of the decision-making mechanisms that generates coherence shifts. Coherence shifts have typically been assessed in a choice context (hence the common term, *choice-induced preference change*; Izuma et al., 2010; Voigt, Murawski, Speer, & Bode, 2019). There is evidence, however, that coherence shifts can be produced by meaningful processing of options without

introducing an overt choice task (DeKay 2015; DeKay, Stone, & Sorenson, 2012; Janis & Mann, 1977; Montgomery & Willen, 1999; Russo et al., 1996, 2008; Simon et al., 2001; Svenson, 1992; for reviews see Brownstein, 2003; Russo, 2015). We propose that the core component of decision making that triggers coherence shifts is a *comparison* of options to assess their similarity in value. People can be asked to compare the value of two options (i.e., to judge how similar is their overall value) without requiring an overt choice. In Experiment 2, we tested the prediction that the task of comparing option values for the sole purpose of judging their similarity, without having to choose one over the other, will lead to the same sorts of coherence shifts in judgments of overall value and attribute values as does a choice task.

We conducted two behavioral experiments in which people judged individual options with respect to both overall value and attribute values, both prior to and after making a choice (Experiment 1) or making a similarity comparison (Experiment 2). In both experiments we assessed how the relationship between the options (in terms of both difference in overall value and in attribute disparity) impacted eventual choice, choice confidence, RT, and shifts in overall value and in attribute values.

Experiment 1

Experiment 1 examined shifts in overall option values and in attribute values triggered by the choice between two snacks with varying initial values on two attributes, pleasure and nutrition.

Method

Participants

A total of 58 people participated in Experiment 1 (32 female; age: mean = 39 years, SD = 8, range 25-50). (The data for one participant were corrupt, so we excluded this participant from analyses.) This sample size was chosen in line with previous studies that used a similar paradigm. All participants were recruited using Amazon Mechanical Turk (MTurk) (https://mturk.com). All were classified as “masters” by Mturk. They were residents of the United States or Canada, and all were native English speakers. Each participant received a payment of $7.50 as compensation for approximately one hour of time. Our experiments involved de-identified participant data, and protocols were approved by the Institutional Review Board of the University of California, Los Angeles. All participants gave informed consent prior to commencing the experiments.

Materials

The experiments were constructed using the online experimental platform Gorilla (gorilla.sc). The experimental stimuli were drawn from a set of 200 digital images used in a previous study (Lee & Coricelli, 2020), each representing a distinct snack food item. For Experiment 1, we used a subset of 100 stimuli, identical for all participants. We determined this subset by selecting the 100 items for which ratings from a previous study (Lee & Coricelli, 2020) varied the least (within each item) across participants. Figure 2 shows examples of images of snacks, indicating their initial values on the two attributes of pleasure and nutrition (averaged across participants).

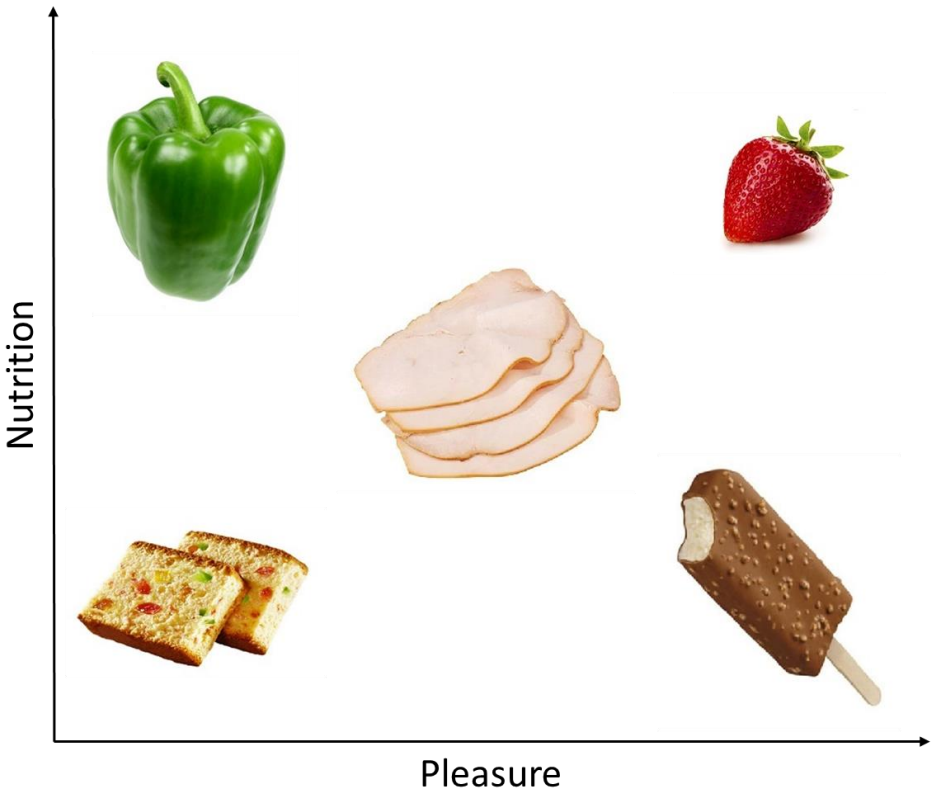


Figure 2. An example of snack foods included in the stimulus set, plotted according to their respective ratings along the dimensions of pleasure and nutrition (averaged across participants). From lower left to upper right: fruitcake rated low on both pleasure and nutrition; bell pepper rated low on pleasure but high on nutrition; sliced turkey rated medium on both pleasure and nutrition; ice cream bar rated high on pleasure but low on nutrition; strawberry rated high on both pleasure and nutrition.

Design and Procedure

The experiment consisted of a pre-exposure phase, followed by initial value ratings, choice, and final ratings. No time limits were imposed for any of the constituent tasks or for the overall experiment.

In the pre-exposure phase, participants simply observed as all individual items were displayed in a random sequence for 750 ms each. The purpose of the pre-exposure phase was to familiarize participants with the full set of items that they would later evaluate, allowing them to form an impression of the range of subjective values across the item set.

The initial value ratings comprised three sets of ratings: for overall value of the snack, and for the value of each of two individual attributes, nutrition and pleasure. In each rating task, all stimuli were displayed on the screen, one at a time, in a sequence randomized across participants. At the onset of each trial, a fixation cross appeared at the center of the screen for 750 ms. Next, an image of a single food item appeared at the center of the screen. For the rating of overall value, participants responded to the question, “How much would you like this as a daily snack?” using a horizontal slider scale. This question was intended to motivate participants to think carefully while assessing the overall subjective quality of each option, as the choice was to determine long-term daily consumption, rather than a “one off” snack. The leftmost end of the scale was labeled “Not at all.” and the rightmost end was labeled “Very much!” The scale appeared to participants to be continuous, the data was captured in increments of 1 (ranging from 1 to 100). Participants could revise their rating as many times as they liked before finalizing it. Participants clicked the “Enter” button to finalize their value rating response and proceed to the next screen. At that time, the next trial began.

The overall value ratings were followed by attribute ratings, which were performed separately for nutrition and pleasure, with the order counterbalanced across participants. One attribute was rated for every alternative (in one task section), then the other attribute was rated for every alternative (in the next task section). In each attribute rating task, all stimuli were

displayed on the screen, one at a time, in a random sequence (randomized across participants and across sections for each participant). The format of this task was identical to that of the overall value rating task, except now participants responded to the question, “How nutritious do you consider this item to be?” or “How pleasurable do you consider this item to be?” For both attribute ratings, the leftmost end of the slider scale was labeled “Very low!” and the rightmost end was labeled “Very high!”.

The choice task was then administered. For this task, 50 pairs of stimuli were displayed on the screen, one pair at a time, in a sequence randomized across participants. The pairings of items for each choice trial were created so as make the choices relatively difficult, as assessed by small differences in value ratings between the two items in a choice pair in a previous study (Lee & Coricelli, 2020). Each individual item occurred in a single choice pair. At the onset of each trial, a fixation cross appeared at the center of the screen for 750 ms. Next, a pair of images of food items appeared on the screen, one left and one right of center. Participants responded to the question, “Which would you prefer as a daily snack?” by clicking on the image of their preferred item. Participants then responded to the question, “How sure are you about your choice?” using a horizontal slider scale. The leftmost end of the scale was labeled “Not at all!” and the rightmost end was labeled “Absolutely!” Participants could revise their confidence report as many times as they liked before finalizing it. Participants clicked the “Enter” button to finalize their confidence report and proceed to the next screen.

Finally, participants made final ratings of overall value and attribute values, exactly as in the initial ratings, except with stimuli presented in new random orders. Note that this procedure (randomizing order of individual items) serves to separate the ratings from the prior choice

context. Prior to completing these final ratings, participants were instructed not to try to remember their earlier ratings, but rather to simply rate the stimuli as they now evaluated them.

Results and Discussion

Unless stated otherwise, all statistical effect sizes reported below were first calculated at the individual level and then tested for significance at the population level. All reported p-values represent the probability of non-zero effect sizes, based on standard two-sided t-tests. For every generalized linear model (GLM) regression analysis reported below, all variables were first converted to *z*-scores (within participant) before being entered into the models.

Coherence Shifts in Overall Value

Our primary analysis focus is on coherence shifts, which result in spreading of alternatives (*SoA*) from initial to final ratings. The choice defines the winning option, and SoA is defined in terms of changes that relatively favor the winner. SoA can be defined for both overall value ratings and for ratings of individual attribute values. For overall value, SoA is defined as the change in overall value for the chosen option from initial to final rating minus the change in overall value for the unchosen option. In accord with previous work showing coherence shifts over the course of decision making, we observed a reliable SoA in overall value across all choice trials (cross-participant mean of within-participant mean SoA 3.70, $p < .001$; cross-participant mean of within-participant median SoA = 2.86, $p < .001$), which is comparable to the magnitude observed in previous studies (Lee & Daunizeau, 2020a; Voigt et al., 2019; Izuma et al., 2010). (See the Supplementary Materials for more details on SoA.)

We then assessed the relationship between choice difficulty measured by the absolute difference in initial overall value ratings between the options, $|dV_1|$, where the subscript

distinguishes the initial ratings (1) from the final ratings (2). As in previous studies (Lee & Coricelli, 2020; Lee & Daunizeau, 2020a), $|dV_1|$ had a reliable negative relationship with overall SoA, a reliable positive relationship with choice confidence, and a reliable negative relationship with RT. We used GLM to regress $|dV_1|$ on overall SoA, on confidence, and on RT, separately. The beta weights for $|dV_1|$ on each dependent variable were significant and in the predicted direction: mean beta for SoA = -0.247 ($p < .001$); mean beta for confidence = 0.348 ($p < .001$); mean beta for RT = -0.191 ($p < .001$).

Lee and Daunizeau (2020a) proposed that decision makers refine their value estimates and certainty for the choice options during deliberation, prior to committing to the choice. In support of this hypothesis, these investigators showed that the impact of $|dV|$ on confidence and RT is larger when $|dV|$ is calculated using post-choice ratings (i.e., $|dV_2|$) rather than pre-choice ratings. This basic finding (also observed by Simon et al., 2004; Simon & Spiller, 2016) was replicated in Experiment 1. The mean GLM beta weight for $|dV_2|$, calculated from post-choice ratings of overall value, as a predictor of confidence was 0.407 ($p < .001$). The magnitude of this beta weight was reliably greater than that obtained using $|dV_1|$ ($p = .003$). The mean GLM beta weight for $|dV_2|$ as a predictor of RT was -0.231 ($p < .001$). The magnitude of this beta weight was greater than that obtained using $|dV_1|$ ($p = .056$).

Lee and Daunizeau (2020a) also found that SoA positively influences choice confidence, as would be predicted given that the impact of a coherence shift is to effectively make the choice easier prior to entering a response. In brief, a high value of $|dV_1|$ will make the choice an easy one, and confidence will therefore be high. A low $|dV_1|$ will make the choice difficult, thereby encouraging deliberation before responding. Deliberation will on average result in a large SoA

(which is essentially an increment in $|dV_1|$), and therefore will lead to higher confidence. When we included both $|dV_1|$ and SoA in a GLM regression model to predict confidence, the cross-participant mean beta weight for $|dV_1|$ was 0.389 ($p < .001$) and for SoA was 0.161 ($p < .001$). We also checked for a similar effect of SoA on RT. When we included both $|dV_1|$ and SoA in a GLM regression model to predict RT, the cross-participant mean beta weight for $|dV_1|$ was -0.212 ($p < .001$) and for SoA was -0.062 ($p = .007$).

Finally, we tested our prediction that attribute disparity would correlate negatively with both SoA and choice confidence. To explicitly define our disparity variable, we transformed the space representing choice options from attribute space (i.e., a two-dimensional space composed of pleasure and nutrition axes) to a new space in which one dimension is $|dV_1|$ and the other is a value of disparity. Each point in the original attribute space (representing an individual item) necessarily resides on a specific iso-value line (i.e., an imaginary line that connects all points of equal overall value), the slope of which is determined by the (participant-specific) marginal rate of substitution (MRS) of the attributes. We calculated the MRS for each participant as $-b_P/b_N$, where b_P and b_N are the beta weights from the regression of pleasure and nutrition on overall value. The difference in the overall value of the two options being compared is thus the distance between the two iso-value lines on which the options reside (measured in the direction orthogonal to the lines). The disparity measure that we seek is the distance between the two options in the direction parallel to the iso-value lines (see Figure 3 for an illustrative example). This is simply the distance between the scalar projections of the two points (i.e., the location of the options in attribute space) onto the iso-value vector ($[-b_P\ b_N]'$):

$$disparity_{i,j} \triangleq \left| \frac{[P_i\ N_i]*\begin{bmatrix}-b_P\\ b_N\end{bmatrix} - [P_j\ N_j]*\begin{bmatrix}-b_P\\ b_N\end{bmatrix}}{\left\|\begin{matrix}-b_P\\ b_N\end{matrix}\right\|} \right|, \text{for options } i, j. \quad (1)$$

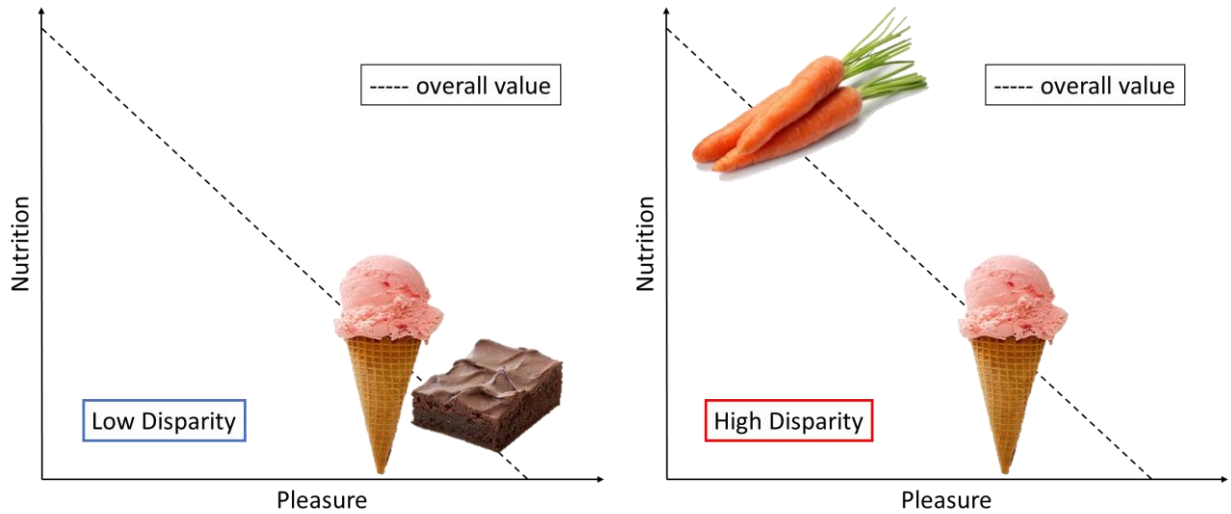
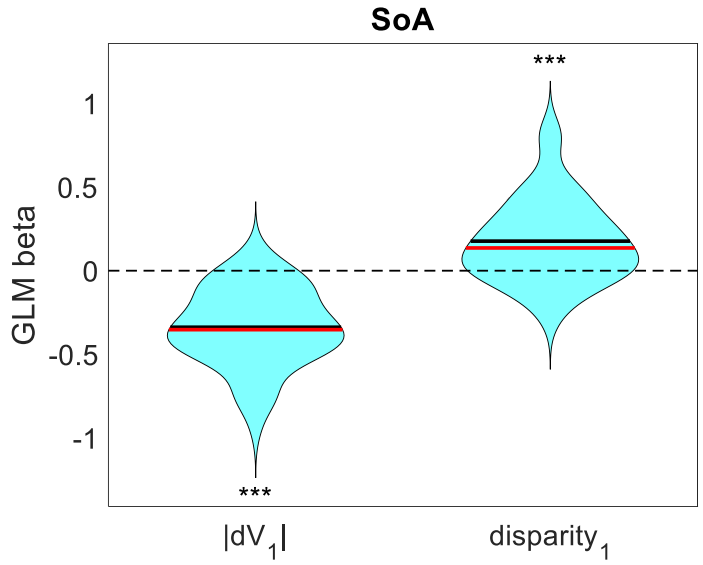
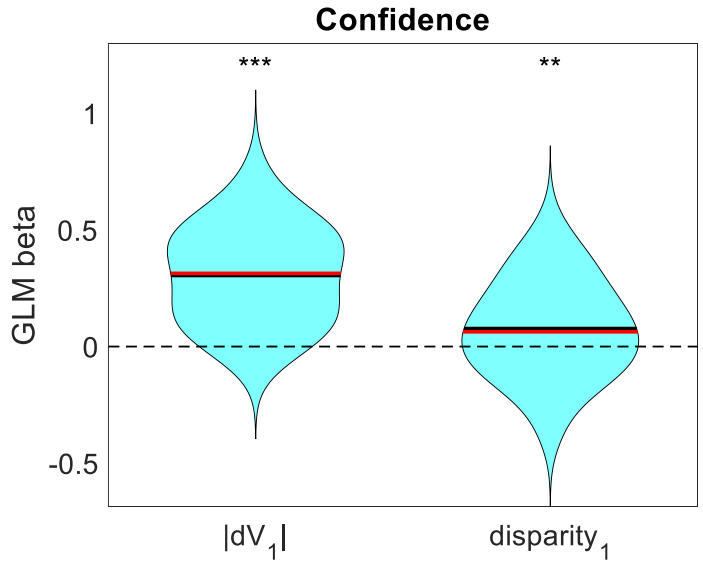


Figure 3. An illustration of two choice sets for snack foods, one low disparity (left plot), one high disparity (right plot). As shown by the dashed iso-value lines, all of the available snacks are of comparable overall value (and thus each choice pair is of comparable low dV). However, the two choice pairs are of very different disparity. In the low disparity pair (left), both options score high on pleasure and low on nutrition. In the high disparity pair (right), one option scores high on pleasure but low on nutrition, while the other option scores low on pleasure but high on nutrition.

Using this formal definition of disparity, we performed GLM regressions at the participant level of $|dV_1|$ and $disparity_1$ (where the subscript indicates disparity based on initial ratings; for simplicity we henceforth drop the subscripts for the two options) on SoA, choice confidence, and RT. Across participants, all beta weights were significant and in the predicted direction (mean $|dV_1|$ beta for SoA = -0.286, $p < .001$; mean $disparity_1$ beta for SoA = 0.125, $p < .001$; mean $|dV_1|$

beta for confidence = 0.333, $p < .001$; mean disparity$_1$ beta for confidence = 0.055, $p = .013$; mean $|dV_1|$ beta for RT = -0.164, $p < .001$; mean disparity$_1$ beta for RT = -0.045, $p = .040$; see Figure 4).





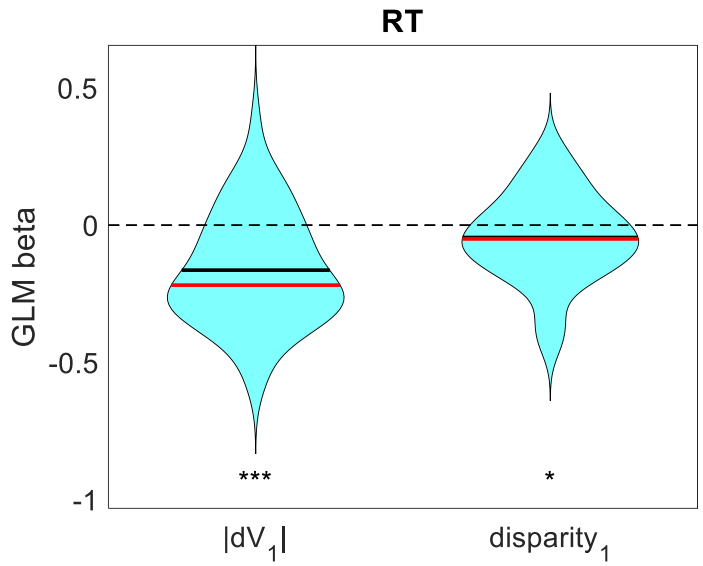


Figure 4. Impact of $|dV_1|$ and $disparity_1$ on SoA (top), confidence (middle), and RT (bottom). Violin plots represent cross-participant distributions of GLM beta weights; black lines represent cross-participant mean values, red lines represent cross-participant median values; *** $p < .001$, ** $p < .01$, * $p < .05$.

Coherence Shifts in Attribute Values

The next set of analyses focused on coherence shifts at the level of the individual attributes, pleasure and nutrition. SoA values at the attribute level (SoA_P and SoA_N for pleasure and nutrition, respectively) were calculated in the analogous manner as overall SoA: in terms of the magnitude of the shift between the initial and final ratings of attribute values in the direction favoring the winner of the choice. Similarly, we defined the absolute value difference for each attribute separately. For the initial ratings, these differences are termed $|dP_1|$and $|dN_1|$, respectively.

We first checked to see if there was a reliable positive average SoA_P and SoA_N across trials. We observed a reliable SoA_P across all items (cross-participant mean of within-participant mean SoA_P = 1.27, p = .002; cross-participant mean of within-participant median SoA_P = 1.18, p < .001). We observed a smaller and slightly less reliable SoA_N across all items (cross-participant mean of within-participant mean SoA_N = 0.40, p = .179; cross-participant mean of within-participant median SoA_N = 0.48, p = .004). (See the Supplementary Materials for more details on SoA_P and SOA_N.)

We tested our predictions that $|dP_1|$ would predict SoA_P and that $|dN_1|$ would predict SoA_N using GLM regression. Across participants, both beta weights were negative and significant (for dP_1 as a predictor of SoA_P, mean beta = -0.245, p < .001; for dN_1 as a predictor of SoA_N, mean beta = -0.050, p = .023) (see Figure 3). The larger beta value for pleasure than nutrition reflects the greater subjective importance of the former attribute.

We then ran GLM regressions of both $|dP_1|$ and $|dN_1|$ on choice confidence and on RT. Across participants, beta weights for confidence were positive and significant for both independent variables (mean $|dP_1|$ beta = 0.241, p < .001; mean $|dN_1|$ beta = 0.081, p < .001), and beta weights for RT were negative and significant for both independent variables (mean $|dP_1|$ beta = -0.139, p < .001; mean $|dN_1|$ beta = -0.049, p = .042; see Figure 5). When post-choice instead of pre-choice attribute ratings were used as predictors of confidence, the beta weight for $|dP_2|$ was reliably greater than that for $|dP_1|$ (mean beta difference = 0.039, p = .010), but that for $|dN_2|$ did not differ from that for $|dN_1|$ (mean beta difference = -0.012, p = .432). When post-choice ratings were used as predictors of RT, the beta magnitude for $|dP_2|$ was

reliably greater than that for $|dP_1|$ (mean beta difference = -0.035, p = .019), but that for $|dN_2|$ did not differ from that for $|dN_1|$ (mean beta difference = 0.001, p = .974).

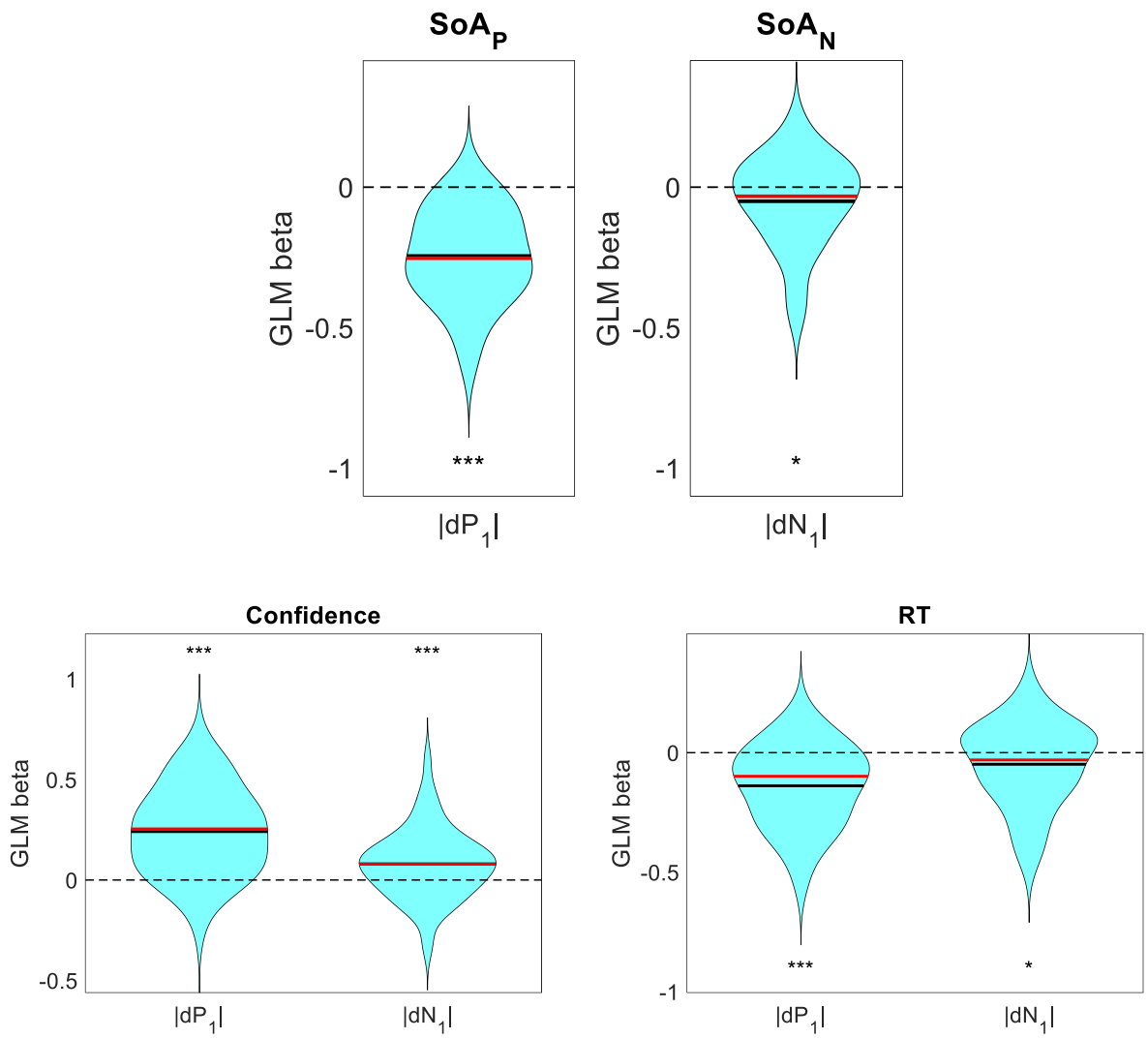


Figure 5. Impact of $|dP_1|$ on SoA_P and $|dN_1|$ on SoA_N (top); and of both $|dP_1|$ and $|dN_1|$ on confidence (bottom left) and on RT (bottom right). Violin plots represent cross-participant distributions of GLM beta weights; black lines represent cross-participant mean values, red lines represent cross-participant median values; *** $p < .001$, ** $p < .01$, * $p < .05$.

Experiment 2

Experiment 2 was designed to test the hypothesis that comparing the similarity of options, without any requirement to make a choice between them, is sufficient to trigger coherence shifts. The general materials and design were basically the same as those of Experiment 1, except that ratings were obtained in three phases: initial ratings, ratings after a comparison of similarity, and finally ratings after a choice. As in Experiment 1, the final choice defined the winning option with respect to which all SoA measures were defined.

Method

Participants

A total of 63 people participated in this experiment (31 female; age: mean = 41 years, SD = 8, range 26-50). This sample size was chosen in line with previous studies that used a similar paradigm. All participants were MTurk “masters”, residents of the United States or Canada, and native English speakers. Each participant received a payment of $7.50 as compensation for approximately one hour of time.

Materials

For Experiment 2 we selected 60 stimuli, identical for all participants, from the full set of snacks (each presented as a digital image). The experimental set consisted of the 30 choice pairs that generated the highest overall SoA across participants in Experiment 1.

Design and Procedure

The basic design was very similar to that of Experiment 1, but with three phases of ratings (rather than two). After pre-exposure to all individual snacks, participants gave their initial ratings (phase 1) of overall value and of attribute values. This was followed by a comparison task for the

30 pairs, after which another set of ratings was obtained (phase 2). The choice task was then administered in the same way as in Experiment 1, and a final set of ratings was obtained (phase 3). Participants were not informed that any task required a choice between options prior to the choice task.

The pre-exposure, rating, and choice tasks were identical in format to those in Experiment 1. For the comparison task, the 30 pairs of stimuli were displayed on the screen, one pair at a time, in a sequence randomized across participants. These item pairs were the same as those that would be used in the choice task, except for the presentation sequence. Each individual item occurred in a single comparison pair. At the onset of each trial, a fixation cross appeared at the center of the screen for 750 ms. Next, a pair of images of food items appeared on the screen, one left and one right of center. Participants responded to the question, “How similarly would you like these as daily snacks?” using a dropdown menu. The response choices were, “I like them {totally equal, very similar, fairly similar, slightly similar, totally different} amounts.” Participants could revise their response as many times as they liked before finalizing it. Participants clicked the “Enter” button to finalize their comparison rating and proceed to the next screen. No time limits were imposed for any of the constituent tasks or for the overall experiment.

Results and Discussion

Experiment 2 included three phases of ratings. We will refer to measures of value difference (overall and for individual attributes) calculated using the initial round of ratings (prior to the comparison task) with the subscript 1 (e.g., $|dV_1|$), those based on the intermediate round of ratings (following the comparison task but prior to the choice task) with the subscript 2, and those calculated using the final round of ratings (following the choice task) with the subscript 3.

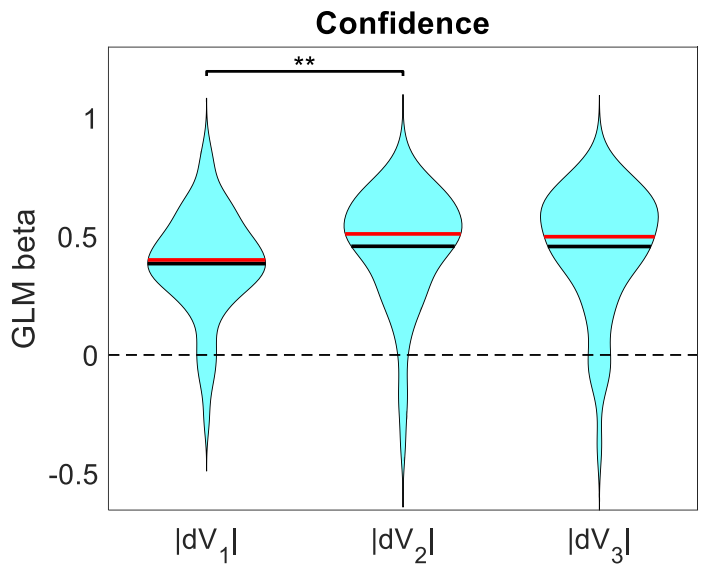
For SoA, subscript 1 is used for those measures calculated from pre- to post-comparison (i.e., from phase 1 to phase 2). The subscript 2 is used for those measures calculated from post-comparison to post-choice (i.e., from phase 2 to phase 3). For all measures of SoA, the winning choice is defined by the option eventually selected during the choice task. Critically, SoA_1 values reflect coherence shifts that occurred *before* the choice task was even administered. Positive values of SoA_1, $SoAP_1$, and $SoAN_1$ will thus indicate changes in the direction of the eventual winner triggered by a comparison task that did not itself require making a choice. Similarly, any impact of comparison on eventual choice confidence would precede the posing of a choice task.

Coherence Shifts in Overall Value During Comparison

Our main goal in Experiment 2 was to demonstrate that each of the effects related to choice that were observed in Experiment 1 would also be triggered by our novel comparison task. Indeed, we found a reliable level of SoA_1 when we examined the change in overall value ratings before and after the comparison task (mean of mean SoA_1 = 2.46, $p < .001$; mean of median SoA_1 = 1.80, $p < .001$).

We then ran GLM regressions of $|dV_1|$ on SoA_1 and on eventual choice confidence and RT. All beta weights were significant and in the predicted direction: mean beta for SoA_1 = -0.206 ($p < .001$); mean beta for confidence = 0.385 ($p < .001$); mean beta for RT = -0.260 ($p < .001$). We then performed GLM regressions of $|dV_1|$ and SoA_1 on confidence and RT. For confidence, the mean beta weight for $|dV_1|$ was 0.438 ($p < .001$) and for SoA_1 was 0.215 ($p < .001$); for RT, the mean beta weight for $|dV_1|$ was -0.291 ($p < .001$) and for SoA_1 was -0.153 ($p < .001$). We also conducted GLM regressions of $|dV_2|$ on confidence and RT. Both beta weights for $|dV_2|$ were significant and in the predicted direction (mean beta for confidence = 0.458, $p < .001$; mean beta

for RT = -0.292, $p < .001$). Replicating the comparable finding in Experiment 1, the beta magnitudes for $|dV_2|$ on both confidence and RT were greater than for $|dV_1|$ (mean difference for confidence = 0.073, $p = .001$; mean difference for RT = -0.032, $p = .069$; see Figure 6). Thus, value differences after coherence shifts were better predictors of choice confidence than were the initial value differences.



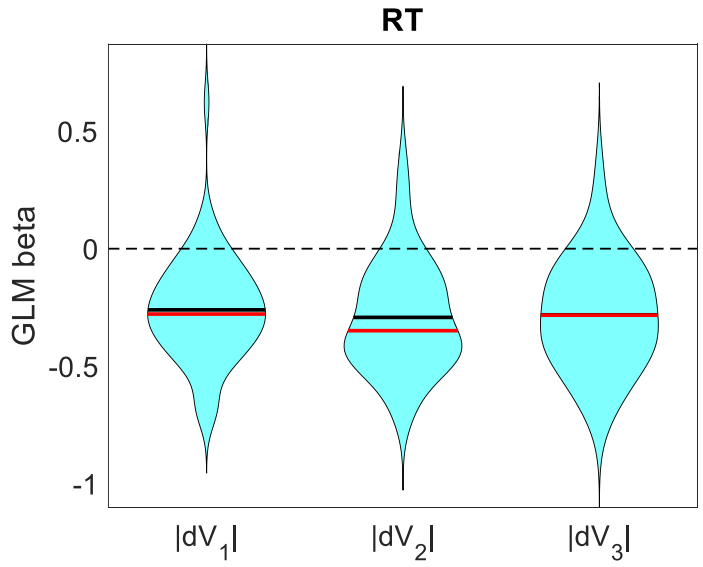
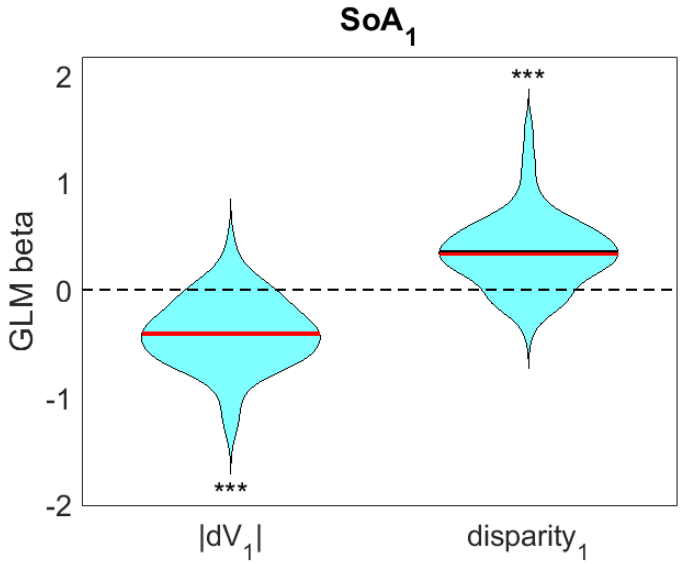


Figure 6. Impact of |dV| on choice confidence (top) and on RT (bottom), separately for each round of ratings. Violin plots represent cross-participant distributions of GLM beta weights; black lines represent cross-participant mean values, red lines represent cross-participant median values; ** $p < .01$.

Finally, we ran GLM regressions of $|dV_1|$ and $disparity_1$ on SoA_1, on choice confidence, and on RT. Across participants, all beta weights were significant and in the predicted direction (mean $|dV_1|$ beta for SoA_1 = -0.405, $p < .001$; mean $disparity_1$ beta for SoA_1 = 0.350, $p < .001$; mean $|dV_1|$ beta for confidence = 0.299, $p < .001$; mean $disparity_1$ beta for confidence = 0.166, $p < .001$; mean $|dV_1|$ beta for RT = -0.211, $p < .001$; mean $disparity_1$ beta for RT = -0.098, $p = .002$; see Figure 7).



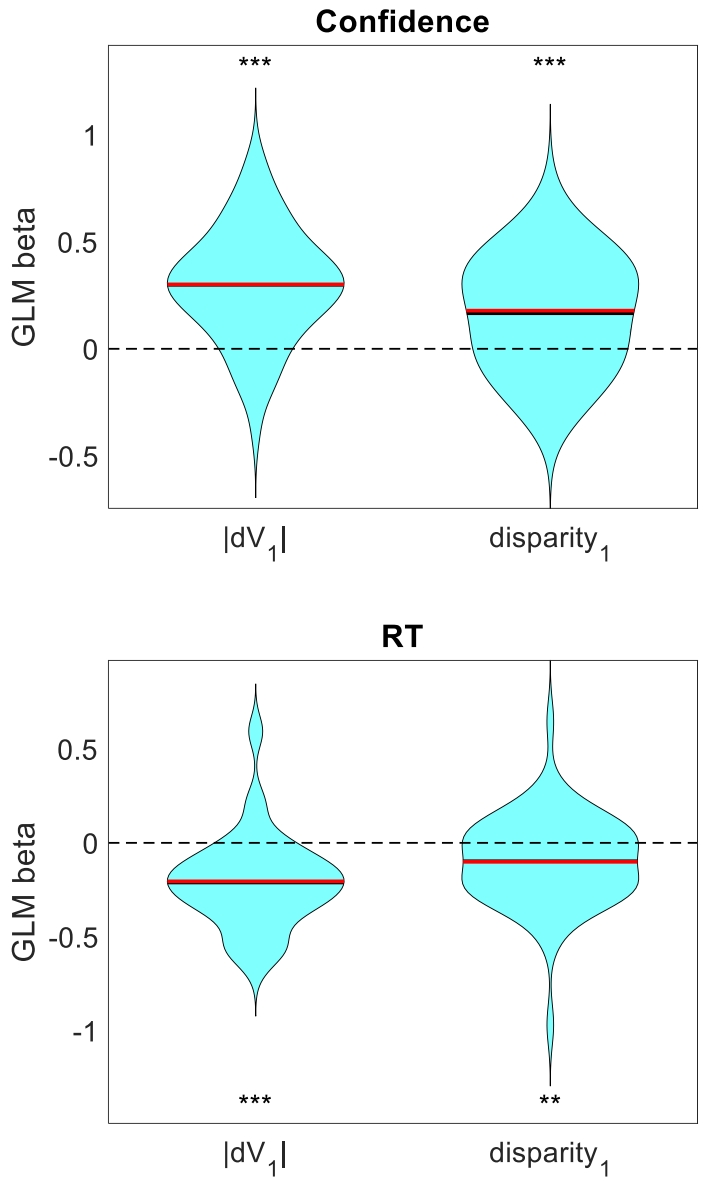


Figure 7. Impact of $|dV_1|$ and $disparity_1$ on SoA_1 (top), confidence (middle), and RT (bottom). Violin plots represent cross-participant distributions of GLM beta weights; black lines represent cross-participant mean values, red lines represent cross-participant median values; *** $p < .001$, ** $p < .01$.

Additional Shifts Triggered by Choice Task

The findings reported above show that a task requiring comparison of value similarities, but not a choice between options, was sufficient to produce the same qualitative pattern of coherence shifts and of changes in eventual choice confidence as we had observed in Experiment 1 following an actual choice task. For Experiment 2, we repeated all the analyses reported above using data from the third phase of ratings (post-choice) to determine whether the actual choice task yielded additional shifts beyond those triggered by the comparison task. The basic pattern was that additional shifts were relatively small and statistically weak. For example, whereas the cross-participant mean SoA_1 was 2.46 (mean median SoA_1 was 1.80), the cross-participant mean SoA_2 (i.e., the additional shift from phase 2 to final phase 3) was only 1.07 (mean median SoA_2 was 0.40), an increment that fell short of statistical significance (mean of mean $p = .070$; mean of median $p = .123$). Similarly, while $|dV_2|$ was a more accurate predictor of confidence and RT than $|dV_1|$ (see above), $|dV_3|$ was no more accurate as a predictor than was $|dV_2|$ (mean difference for confidence < -0.001, $p = .950$; mean difference for RT = 0.011, $p = .441$) (see Figure 4 above). Thus, the comparison task was indeed sufficient to trigger coherence and confidence shifts comparable to those produced by an actual choice task.

General Discussion

Summary and Implications

The present study provides a number of important findings concerning the role of coherence shifts in the construction of preferences. When required to choose between two options (snacks), each differing in two attributes that determine value (pleasure and nutrition), people's assessments of value shifted from pre- to post-choice in the direction that spread the

alternatives further apart so as to favor the winner, thereby increasing confidence in the choice. This shift was observed not only for ratings of overall value, but also for each of the two individual attributes that determined value. Moreover, the magnitude of the coherence shift increased with the difficulty of the choice as measured by the difference in initial ratings of overall value for the two options. Coherence shifts were in turn predictive of increased choice confidence and decreased RT, with confidence and RT being more accurately predicted by value difference measured by ratings obtained after the choice.

We also found that coherence shifts are predicted not only by the overall value difference between options, but also by the pattern of attribute composition across options (which we believe is a novel contribution to the literature). We introduced a formal measure of attribute *disparity* (Equation 1), reflecting the degree to which individual attributes "disagree" with one another as to which option is superior. After accounting for the impact of overall value difference, the magnitude of the coherence shift was also positively correlated with disparity. Moreover, coherence shifts associated with disparity increased confidence and decreased RT in the eventual choice.

Experiment 2 provided evidence that active comparison of option values is the core component of the decision process that drives coherence shifts. To isolate the impact of the comparison process, Experiment 2 used a three-phase design in which value ratings were obtained at the outset, after a comparison task, and finally after an actual choice task. Critically, the comparison task was administered prior to informing the participants that any choice would be required. Instead, they were simply asked to rate the similarity of values for two snacks. This comparison task generated the same qualitative pattern of changes in value ratings and in

confidence and RT in the eventual choice, even though the similarity judgments entirely preceded the actual choice task. Moreover, the comparison task sufficed to generate most of the entire coherence shift, with the later choice task itself adding little additional change. These findings support the hypothesis that active comparison of values is the core component of the decision process during which perceived values are shifted so as to more sharply distinguish the items being compared.

Lee and Daunizeau (2020b) have recently proposed a computational account of why and how coherence shifts can be generated during intra-decisional processing. According to their metacognitive control of decision-making (MCD) model, decision makers should exert mental effort only to the degree necessary to reach some subjective threshold of confidence that the chosen option is indeed the best of the candidate set. People process additional information about choice options up until the point at which the options are sufficiently distinguishable to provide a satisfactory level of confidence in the emerging choice. The MCD model can account for empirical evidence that people change their subjective value estimates when required to choose between options that they initially estimated to have similar values, and that this change in value correlates with both choice confidence and response time.

The behavioral evidence concerning intra-decisional processing provided by the present study fits well with the picture emerging from studies of the emergence of decisions at the neural level. It has been shown that the brain computes attribute values separately and then integrates them before making a choice (Lim, O'Doherty, & Rangel, 2013). Other neural evidence indicates that the assessment of individual attributes, as well as their relative weights on the final choice, evolve during the decision process (Hunt et al, 2014). Löffler, Haggard, and Bode (2020; also Voigt

et al, 2019) have provided compelling neuroimaging evidence that shifts in valuation that support the eventual choice occur prior to the choice response. Future studies using multiple methodologies will hopefully provide a deeper understanding of the timing of coherence shifts at the attribute level.

Directions for Future Research

In Experiment 2 we found that making judgments of similarity in item value sufficed to trigger coherence shifts in overall and attribute values. It would be interesting to examine the impact of other types of similarity judgments (for example, judgments based on all features of items, rather than specifically on their values). Perhaps any type of meaningful processing of paired items will yield coherence shifts. Alternatively, a task set that encourages a focus on comparison of values may be necessary.

In the present study, we found that smaller differences in value between choice alternatives increased both deliberation time and the spreading of the alternatives, on average. This result may seem intuitive; however, it is also possible to imagine that participants would not have bothered to deliberate for choices between options of similar value, because the value would be almost the same regardless of which they chose. Future studies might search for individual differences, as some people may show a non-monotonic relationship between choice difficulty and deliberation time. If some participants do indeed show such non-monotonicity, it would be interesting to examine whether they also show less pronounced coherence shifts.

Another useful direction for future research would be to collect self-reported estimates of attribute-specific importance weights, rather than relying on statistical regression to infer these weights. Recall that in the present experiments, final ratings were obtained using all

individual items after completing all choices between paired options. Each individual item occurred in only one choice trial; hence any value shift for an item can be attributed to that one specific choice. However, the two attributes were likely assessed on all 50 choice trials, and hence their relative importance would have been pushed in different directions on different trials (presumably canceling out any coherence shifts due to trial-by-trial changes in attribute importance weights). A different paradigm would be required to meaningfully assess coherence shifts in attribute importance. Such a paradigm could also be used to assess whether the magnitude of coherence shifts along a particular attribute dimension is systematically correlated with the self-reported importance of that attribute.

Acknowledgements

Preparation of this paper was supported by NSF Grant BCS-1827374. We thank Dan Simon for comments on an earlier draft of this paper.

References

Ariely, D., Loewenstein, G., & Prelec, D. (2006). Tom Sawyer and the construction of value. *Journal of Economic Behavior & Organization*, *60*, 1-10.

Bordalo, P., Gennaioli, N., & Shleifer, A. (2020). Memory, attention, and choice. *Quarterly Journal of Economics*, *135*(3), 1399-1442.

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

Brownstein, A. (2003). Biased pre-decision processing. *Psychological Bulletin*, *129*, 545–568.

Caplin, A., & Dean, M. (2015). Revealed preference, rational inattention, and costly information acquisition. *American Economic Review*, 105(7), 2183-2203.

Carlson, K. A., & Russo, J. E. (2001). Biased interpretation of evidence by mock jurors. *Journal of Experimental Psychology: Applied*, *7*, 91-103.

Chaxel, A.-S. (2015). The impact of a relational mindset on information distortion. *Journal of Experimental Social Psychology*, *60*, 1-7.

DeKay, M. L. (2015). Predecisional information distortion and the self-fulfilling prophecy of early preferences in choice. *Current Directions in Psychological Science*, *24*(5), 405-411.

DeKay, M. L., Stone, E. R., & Miller, S. A. (2011). Leader-driven distortion of probability and payoff information affects choices between risky prospects. *Journal of Behavioral Decision Making*, *24*(4), 394–411.

DeKay, M. L., Stone, E. R., & Sorenson, C. M. (2012). Sizing up information distortion: Quantifying its effect on the subjective values of choice options. *Psychonomic Bulletin and Review*, *19*(2), 349–356.

Festinger, L. (1957). *A theory of cognitive dissonance*. Evanston, IL: Row, Peterson.

Harmon-Jones, E., & Harmon-Jones, C. (2007). Cognitive dissonance theory after 50 years of development. *Zeitschrift für Sozialpsychologie*, *38*(1), 7-16.

Holyoak, K. J., & Powell, D. (2016). Deontological coherence: A framework for commonsense moral reasoning. *Psychological Bulletin, 142*, 1179–1203

Holyoak, K. J., & Simon, D. (1999). Bidirectional reasoning in decision making by constraint satisfaction. *Journal of Experimental Psychology: General*, *128*, 3-31.

Hunt, L. T., Dolan, R. J., & Behrens, T. E. (2014). Hierarchical competitions subserving multi-attribute choice. *Nature Neuroscience*, *17*(11), 1613-1622.

Izuma, K., Matsumoto, M., Murayama, K., Samejima, K., Sadato, N., & Matsumoto, K. (2010). Neural correlates of cognitive dissonance and choice-induced preference change. *Proceedings of the National Academy of Sciences*, *USA*, *107*(51), 22014-22019.

Janis, I. L., & Mann, L. (1977). *Decision making: A psychological analysis of conflict, choice, and commitment*. New York, NY: Free Press.

Lee, D., & Coricelli, G. (2020). An empirical test of the role of value certainty in decision making. bioRxiv, 155234.

Lee, D., & Daunizeau, J. (2020a). Choosing what we like vs liking what we choose: How choice-induced preference change might actually be instrumental to decision-making. *PLoS ONE*, *15*(5), e0231081.

Lee, D., & Daunizeau, J. (2020b). Trading mental effort for confidence: The metacognitive control of value-based decision-making. bioRxiv, 837054.

Lichtenstein, S., & Slovic, P. (Eds.) (2006). *The construction of preference*. Cambridge, UK: Cambridge University Press.

Lim, S. L., O'Doherty, J. P., & Rangel, A. (2013). Stimulus value signals in ventromedial PFC reflect the integration of attribute value signals computed in fusiform gyrus and posterior superior temporal gyrus. Journal of Neuroscience, 33(20), 8729-8741.

Löffler, A., Haggard, P., & Bode, S. (2020). Decoding changes of mind in voluntary action—Dynamics of intentional choice representations. *Cerebral Cortex*, *30*(3), 1199-1212.

Montgomery, H., & Willen, H. (1999). Decision making and action: The search for a good structure. In P. Jusslin & H. Montgomery (Eds.), *Judgment and decision making: Neo-Brunswikian and process-tracing approaches* (pp. 147–173). Mahwah, NJ: Erlbaum.

Payne, J. W., Bettman, J. R., & Schkade, D. A. (1999). Measuring constructed preferences: Towards a building code. *Journal of Risk and Uncertainty*, *19*, 243–270.

Rieskamp, J., Busemeyer, J. R., & Mellers, B. A. (2006). Extending the bounds of rationality: Evidence and theories of preferential choice. *Journal of Economic Literature*, *44*(3), 631-661.

Russo, J. E., Carlson, K. A., Meloy, M. G., & Yong, K. (2008). The goal of consistency as a cause of information distortion. *Journal of Experimental Psychology: General*, *137*(3), 456–470.

Russo, J. E., & Chaxel, A.-S. (2010). How persuasive messages can influence behavior without awareness. *Journal of Consumer Psychology*, *20*(3), 338–342.

Russo, J. E., Medvec, V. H., & Meloy, M. E. (1996). The distortion of information during decisions. *Organizational Behavior and Human Decision Processes*, *66*, 102-110.

Simon, D., & Holyoak, K. J. (2002). Structural dynamics of cognition: From consistency theories to constraint satisfaction. *Personality and Social Psychology Review*, *6*, 283-294.

Simon, D., Krawczyk, D. C., Bleicher, A., & Holyoak, K. J. (2008). The transience of constructed preferences. *Journal of Behavioral Decision Making*, *21*, 1-14.

Simon, D., Krawczyk, D. C., & Holyoak, K. J. (2004). Construction of preferences by constraint satisfaction. *Psychological Science*, *15*, 331-336.

Simon, D., Pham, L. B., Le, Q. A., & Holyoak, K. J. (2001). The emergence of coherence over the course of decision making. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *27*, 1250-1260.

Simon, D., Stenstrom, D. M., & Read, S. J. (2015). The coherence effect: Blending hot and cold cognitions. *Journal of Personality and Social Psychology, 109,* 369–394.

Sims, C. A. (2003). Implications of rational inattention. *Journal of Monetary Economics*, *50*(3), 665-690.

Spellman, B. A., Ullman, J. B., & Holyoak, K. J. (1993). A coherence model of cognitive consistency. *Journal of Social Issues*, *4*, 147-165.

Svenson, O. (1992). Differentiation and consolidation theory of human decision making: A frame of reference for the study of pre- and post-decision processes. *Acta Psychologica*, *80*, 143-168.

Tversky, A. (1969). Intransitivity of preferences. *Psychological review, 76(1)*, 31.

Tversky, A., & Simonson, I. (1993). Context-dependent preferences. *Management Science*, *39*(10), 1179-1189.

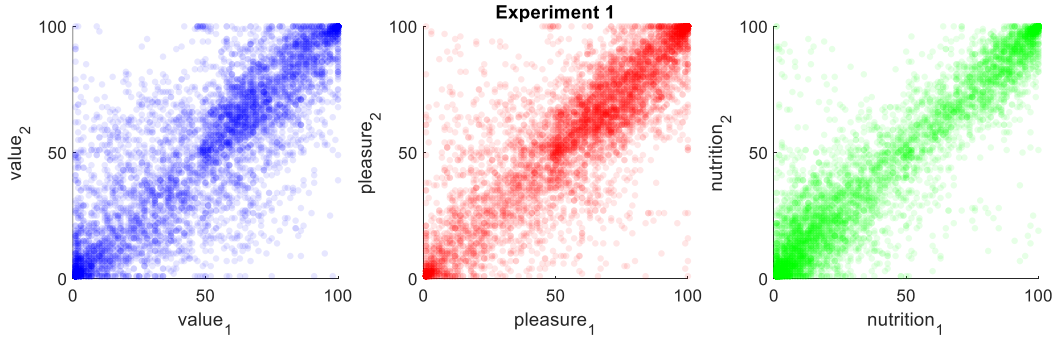
Voigt, K., Murawski, C., Speer, S., & Bode, S. (2019). Hard decisions shape the neural coding of preferences. *Journal of Neuroscience*, *39*(4), 718-726.

Warren, C., McGraw, A. P., & Van Boven, L. (2011). Values and preferences: Defining preference construction. *WIREs Cognitive Science*, *2*(2), 193–205.

Supplementary Material

Data Quality Checks

Before conducting our main analyses, we performed a number of simple quality checks to verify that our data were reliable. First, we assessed the test-retest reliability of overall value ratings. For each participant, we measured the correlation between first (R1) and second (R2) ratings (both experiments) and between first and third (R3) ratings (Experiment 2), across items. We found that ratings were generally consistent (in Experiment 1, median Spearman's rho for R1-R2 = 0.830; in Experiment 2, for R1-R2 = 0.857, and for R1-R3 = 0.840). We then performed the same tests for pleasure ratings, and found that ratings were generally consistent (in Experiment 1, median Spearman's rho for R1-R2 = 0.835; in Experiment 2, for R1-R2 = 0.877, for R1-R3 = 0.865). We then performed the same tests for nutrition ratings, and found that ratings were generally consistent (in Experiment 1, median Spearman's rho for 1 R1-R2 = 0.885; in Experiment 2, for R1-R2 = 0.886, for R1-R3 = 0.860). These correlations are presented in Figure S1.



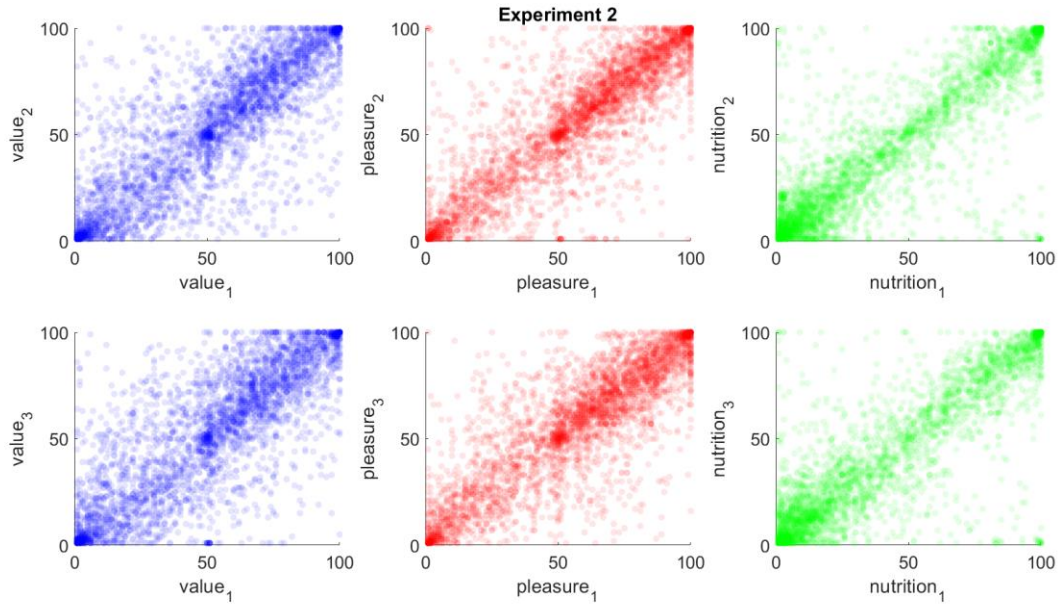


Figure S1. Test-retest reliability for ratings across experimental task sections. Left plots show overall value ratings (blue), middle plots show pleasure ratings (red), and right plots show nutrition ratings (green).

Beyond test-retest reliability, we also showed that the distributions for overall ratings, for pleasure ratings, and for nutrition ratings did not change between experimental task sections (see Figure S2). This is important, because if participants changed how they used the rating scale as they rated more options, a mere effect of habituation to the rating scale could have caused the observed spreading of alternatives effect. However, we see no evidence that this was the case in either experiment.

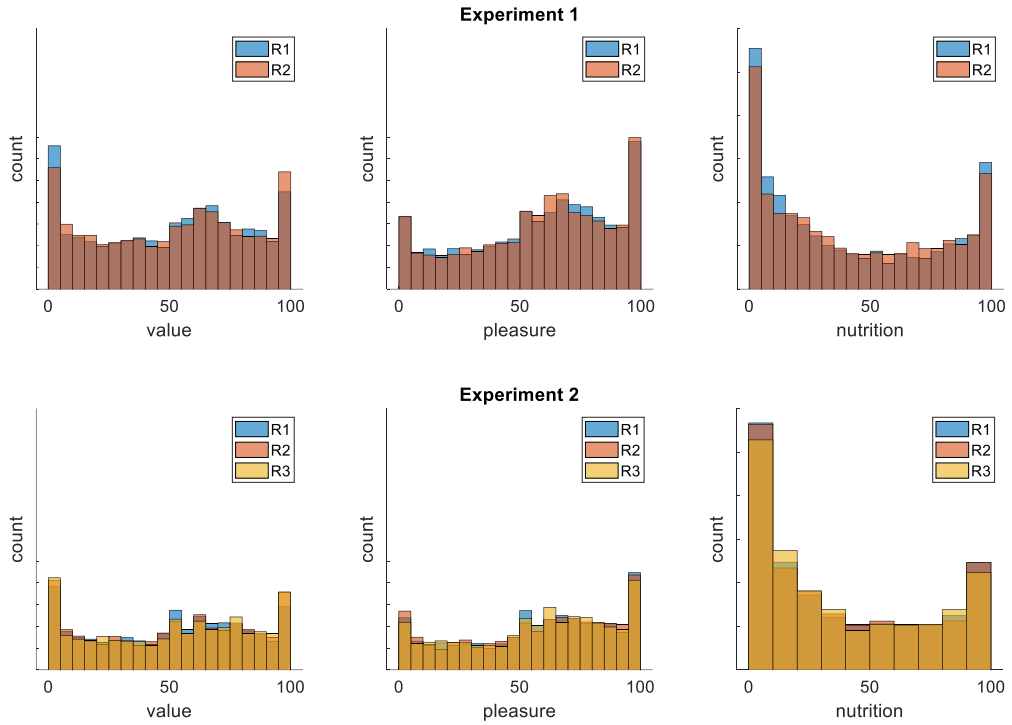


Figure S2. Histograms of ratings (overall value, pleasure, and nutrition) pooled across participants, separately for Experiment 1 (top row) and Experiment 2 (bottom row). R1 (blue) is for the first rating task, R2 (red) is for the second rating task, R3 (yellow) is for the third rating task. Notably, there is no difference in any of the distributions across tasks.

Next, we checked whether the range of values for both attributes (pleasure and nutrition) was the same for each participant. This would be important because we relied on GLM regression to estimate the attribute importance weights upon which our disparity calculations were based. If the range of one attribute was substantially different than that for the other, then the beta weights obtained by GLM regression might not actually be indicative of how much each attribute contributes to overall value. We found that most participants used the full range of possible

values for both pleasure and nutrition, and that the full pleasure-nutrition space was sufficiently covered (see Figure S3). Our techniques reported in the main manuscript were thus acceptable.

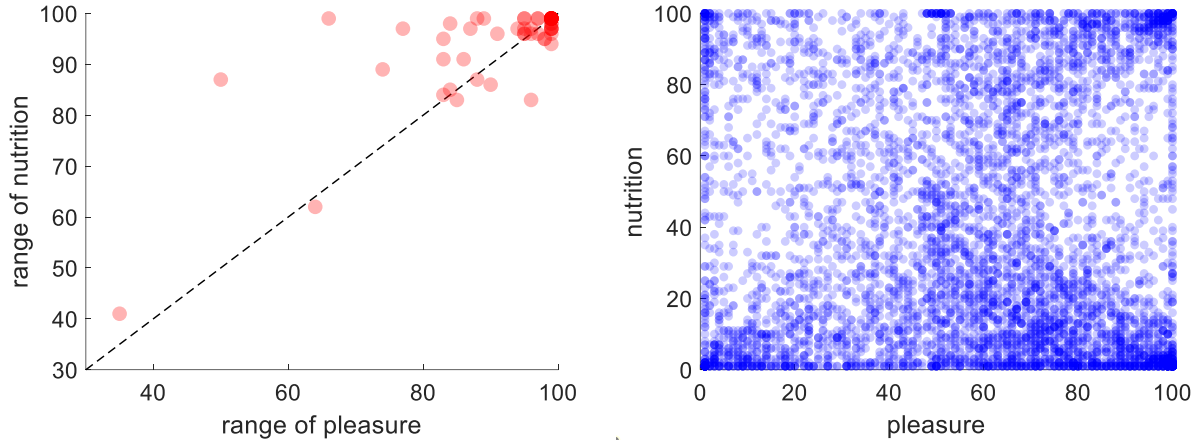


Figure S3. Across participants, pleasure and nutrition ratings occupied the full range of the scale. The left plot shows the range nutrition ratings against the range of pleasure ratings (each red dot indicates one participant). The right plot shows that the full pleasure-nutrition space was sufficiently covered, across participants (each blue dot represents one item). Here we show only the data for Experiment 1; the data for Experiment 2 were similar.

Because many of our analyses relied on the |dV| and disparity variables, we checked to ensure that each of these variables was sufficiently well distributed across a wide-enough range of values. We found that the distributions were similar, and that the |dV|-disparity space was sufficiently covered (see Figure S4).

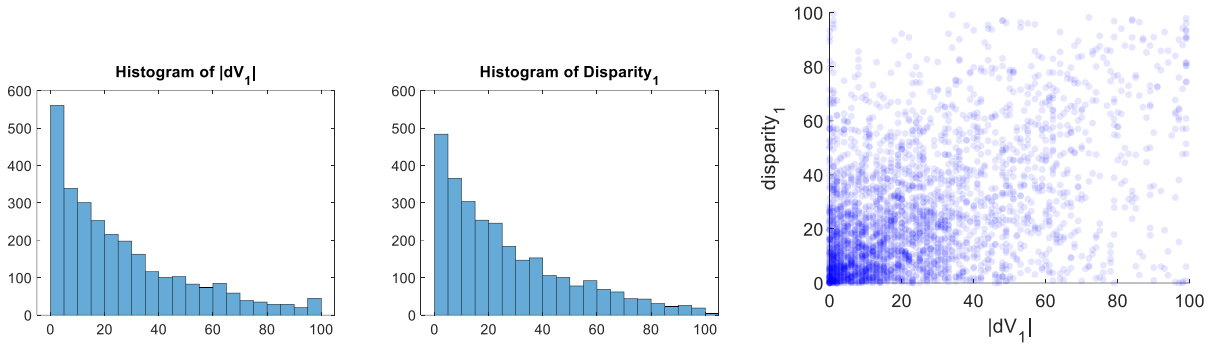
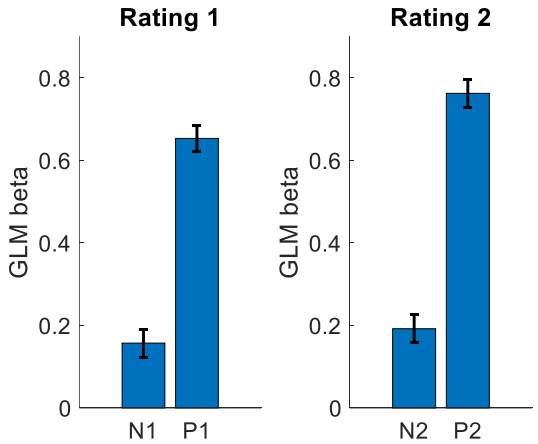


Figure S4. |dV| and disparity were similarly and sufficiently well distributed. Left and middle plots show histograms of |dV| and disparity (here shown for R1, across participants). Right plot shows that the full |dV|-disparity space was sufficiently well covered. Here we show only the data for Experiment 1; the data for Experiment 2 were similar.

Attribute Importance Weights

Because our experimental design has never before been used (to our knowledge) with attribute-specific ratings, we wanted to verify that the attribute ratings we collected were truly related to overall value as we expected them to be. To do this, we first regressed (GLM) the pleasure and nutrition attribute ratings (R1) of the items onto the overall value ratings, separately for each participant. Across participants in both experiments, there was a reliable contribution of both attributes to value ratings (Experiment 1: mean pleasure beta = 0.653, $p < .001$; mean nutrition beta = 0.157, $p < .001$; Experiment 2: mean pleasure beta = 0.660, $p < .001$; mean nutrition beta = 0.118, $p < .001$). These effects were stronger when we used the second round of attribute and value ratings (R2) instead of the first (Experiment 1: mean pleasure beta = 0.763, $p < .001$; mean nutrition beta = 0.192, $p < .001$; Experiment 2: mean pleasure beta = 0.733, $p < .001$; mean nutrition beta = 0.154, $p < .001$). The difference in the beta weights using R2 vs R1

ratings was significant for pleasure (Experiment. 1: $p < .001$; Experiment 2: $p < .001$), but only moderately significant for nutrition (Experiment 1: $p = .096$; Experiment 2: $p = .029$). For Experiment 2, these effects were stronger when we used the third round of attribute and value ratings (R3) instead of the second (mean pleasure beta = 0.780, $p < .001$; mean nutrition beta = 0.189, $p < .001$). The difference in the beta weights using R3 vs R1 ratings was significant for both pleasure ($p = .039$) and nutrition ($p = .047$). These results imply that participants had a more precise sense of their subjective valuations (in terms of attribute contribution to overall value) for the set of items after deliberation (for both similarity judgment and explicit choice) relative to before (see Figure S5).



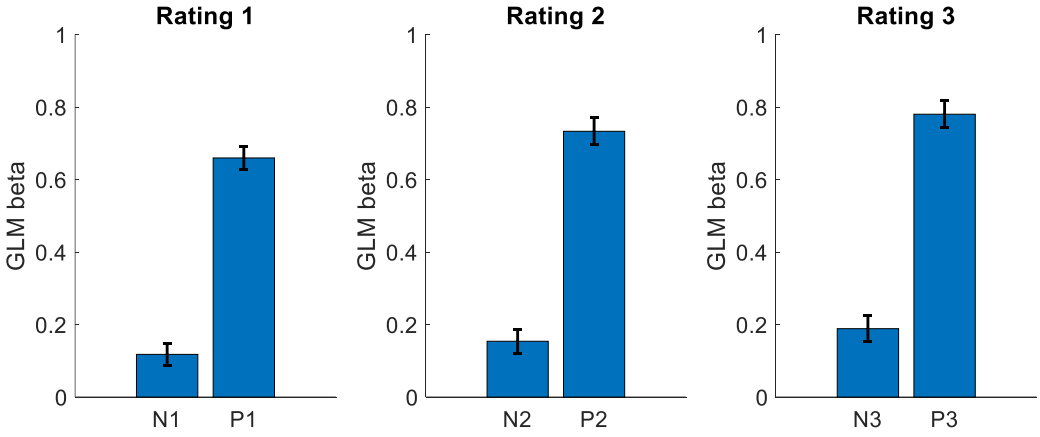
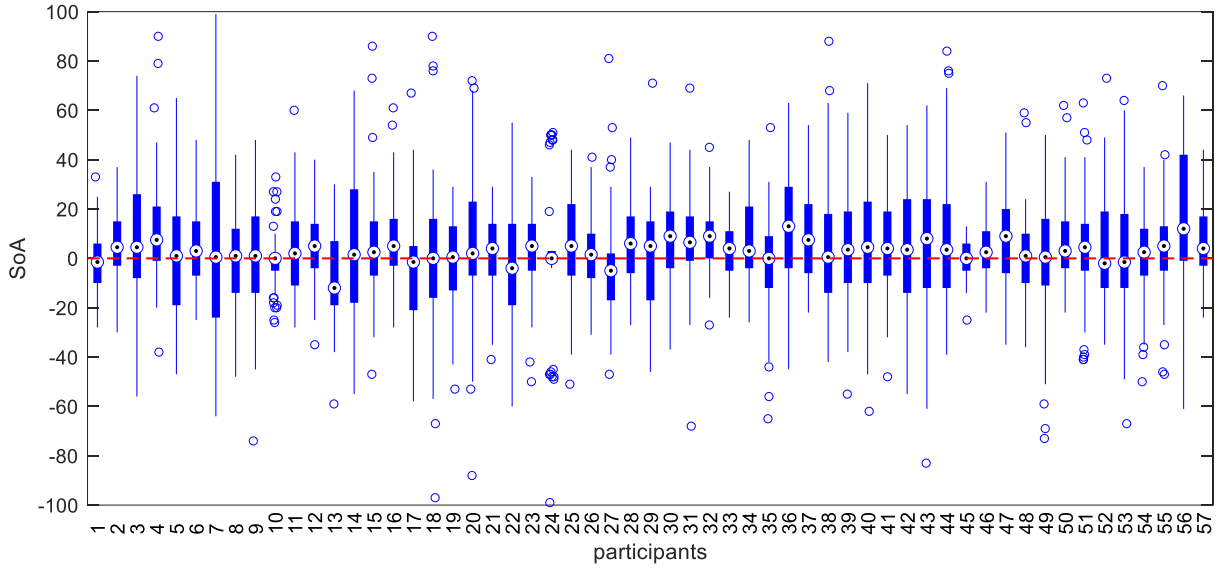
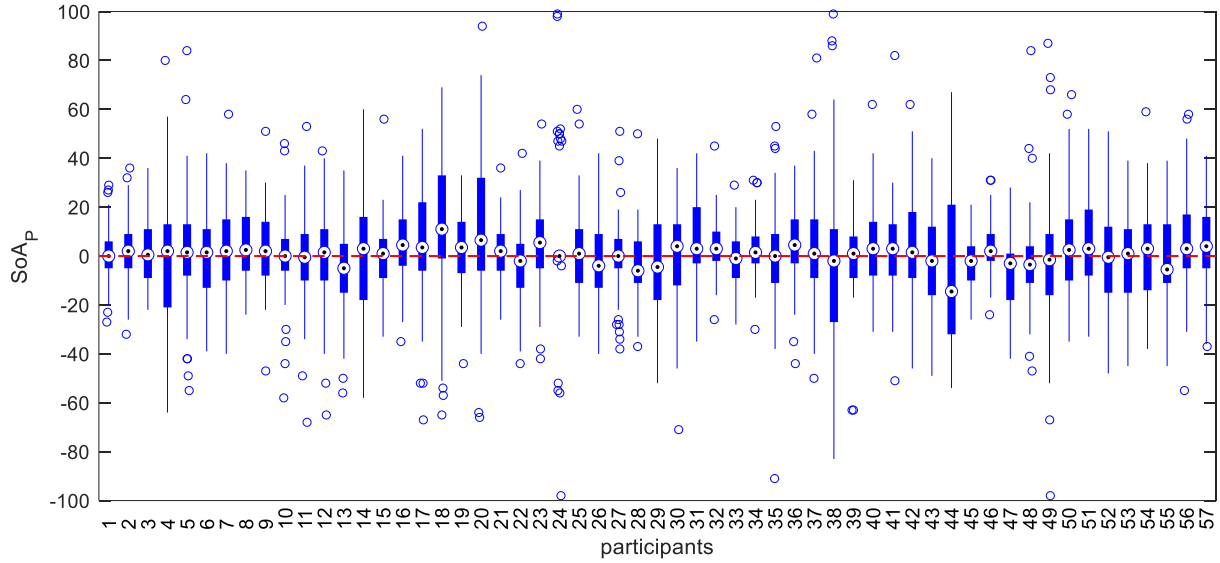


Figure S5. Importance weights of attributes contributing to overall value for Experiment 1 (top row) and Experiment 2 (bottom row), calculated using GLM regression of attribute ratings onto overall value ratings. The magnitude of the beta weights for both nutrition (N) and pleasure (P) increased from first to second rating task, and from second to third rating task, suggesting that deliberation (in the choice or similarity judgment tasks between each rating task) refined the subjective mapping from attribute values to overall value.

Spreading of Alternatives

In the Results section of the main manuscript, we report group averages for the mean and median within-participant spreading of alternatives (SoA). SoA was highly variable across trials for every participant, implying that much of the observed SoA effect in our data is driven by either noise or some unknown factor. Nevertheless, SoA is reliably greater than zero on average (both mean and median) for most participants. This holds true for both overall value SoA, and for attribute-specific SoA (SoA_P and SoA_N; see Figure S6).


SoA
participants


SoA_P
participants

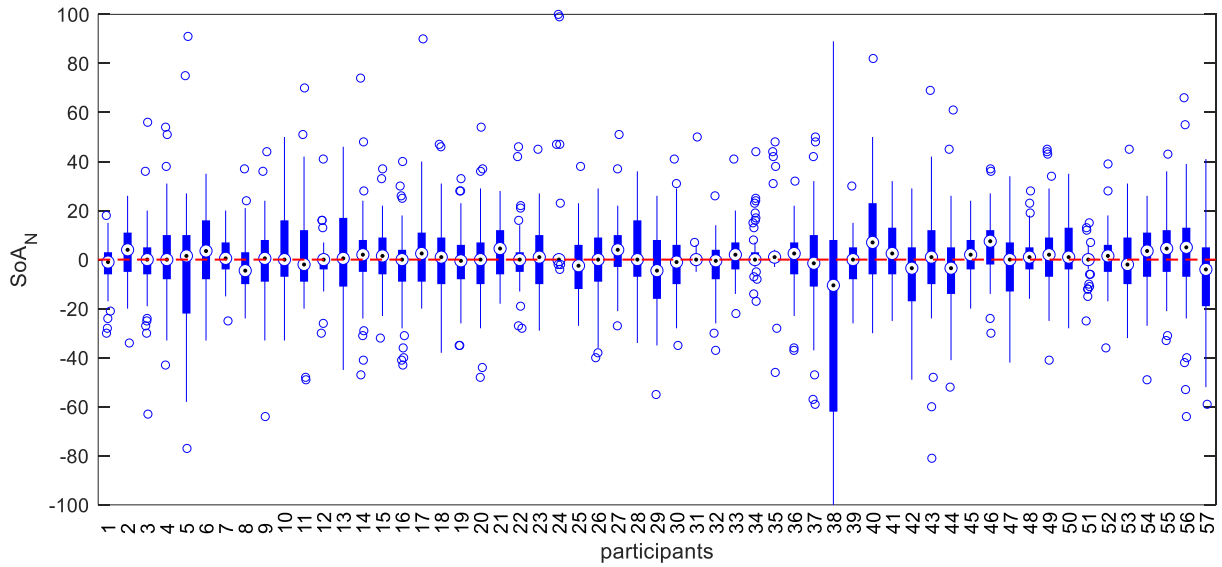


Figure S6. Within-participant distributions of spreading of alternatives for overall value (SoA, top plot), pleasure (SoA_P, middle plot), and nutrition (SoA_N, bottom plot). For each box, the central mark is the median, the edges of the box are the 25th and 75th percentiles, the whiskers extend to the most extreme data points the boxplot algorithm considers not to be outliers, and the outliers are plotted individually. Here we show only the data for Experiment 1; the data for Experiment 2 were similar.

Similarity Judgments

We realize that the phrasing of our similarity judgment question in Experiment 2 ("How similarly would you like these items as a daily snack?") was a bit unnatural. It was therefore possible that participants might have been confused about how to properly respond to such a question. It was our intention that participants would assess the similarity of the options in terms of their subjective overall values. We thus checked to see how well the reported similarity

judgments for the choice pairs correlated with their differences in overall value (as reported in R1). We found that, as expected, the two measures were highly negatively correlated across participants (mean Spearman's rho = -0.410, $p < .001$; see Figure S7). Also as expected, the magnitude of the correlation increased when using post-judgment ratings (R2) (mean Spearman's rho = -0.489, $p < .001$), and the difference in the two correlations was significant ($p < .001$).

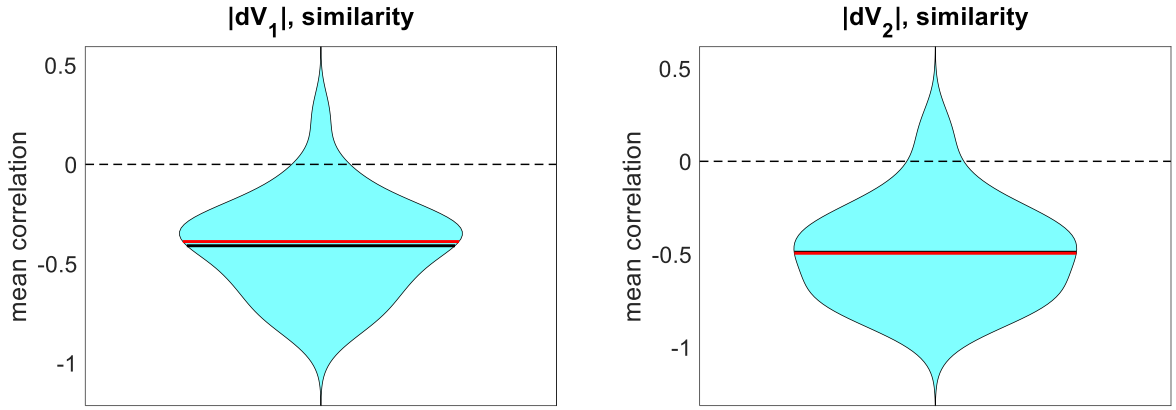


Figure S7. Similarity judgments in Experiment 2 capture value difference, as intended. Overall value difference for choice pair options (|dV|) correlated negatively with subjective reports of similarity for the same choice pair options. Left plot shows the correlation between similarity and pre-judgment ratings; right plot shows the correlation between similarity and post-judgment ratings.