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

Email for correspondence:

DouglasGLee@gmail.com

Word count: 5,219

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. 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 (Busemeyer et al., 2019; Rieskamp, Busemeyer, & Mellers, 2006; Tversky & Simonson, 1993). Under this view, attribute values provide pre-existing and stable elements that serve as the basic inputs to the decision process.

An alternative view emphasizes the constructive nature of preference (Ariely, Loewenstein, & Prelec, 2006; Bettman, Luce, & Payne, 1998; Holyoak & Simon, 1999; Russo, Medvec, & Meloy, 1996). 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), and evaluating complex issues with moral implications (Holyoak & Powell, 2015; Spellman, Ullman, & Holyoak, 1993). 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 in measures of preliminary leanings, prior to making a firm decision or public commitment (Holyoak & Simon, 1999; Russo et al., 1996; 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. In particular, are all attribute values equally subject to coherence shifts, or are there systematic differences in the magnitude of such shifts? A separate line of research suggests that the way in which people process information during decision making is under metacognitive control. Mental effort is exerted to the degree necessary to reach some subjective threshold of confidence that the chosen option is indeed the best of the candidate set (Lee & Daunizeau, 2020a). According to their metacognitive control of decision-making (MCD) model, people process additional information about choice options up until the point where they are sufficiently distinguishable to provide a satisfactory level of confidence in the emerging choice. The MCD model, which complements work on constructed preferences (e.g., Bettman et al., 1998; Simon et al., 2001) and on rational inattention (Sims, 2003; Caplin & Dean, 2015), 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 choice confidence (Lee & Daunizeau, 2020a, 2020b).

Previous tests of the MCD model have measured changes in evaluations of the overall value of options, but 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. In accord with predictions of the MCD model, 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 individual attributes and overall value.

In addition, we examined the possibility that decision dynamics may also be influenced by differences in attribute composition. Consider two choices situations, both involving 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 an attribute is for 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, as the choice effectively becomes easier (Lee & Daunizeau, 2020a, 2020b).

The present study also attempted to identify the core component of the decision-making mechanisms that generates coherence shifts. Previous studies have assessed coherence shifts prior to making a firm and final decision by assessing preliminary leanings (e.g., Holyoak & Simon, 1999; Russo et al., 1996; Simon et al., 2001). However, even a request for a preliminary leaning necessarily creates a choice context (hence the common term, *choice-induced preference change*; Izuma et al., 2010; Voigt, Murawski, Speer, & Bode, 2019). It is possible, however, that coherence shifts can be produced *without* introducing an actual choice task. 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 any kind of choice (even a preliminary leaning). 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 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, 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). 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 item 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 1 shows examples of images of snacks, indicating their initial values on the two attributes of pleasure and nutrition.

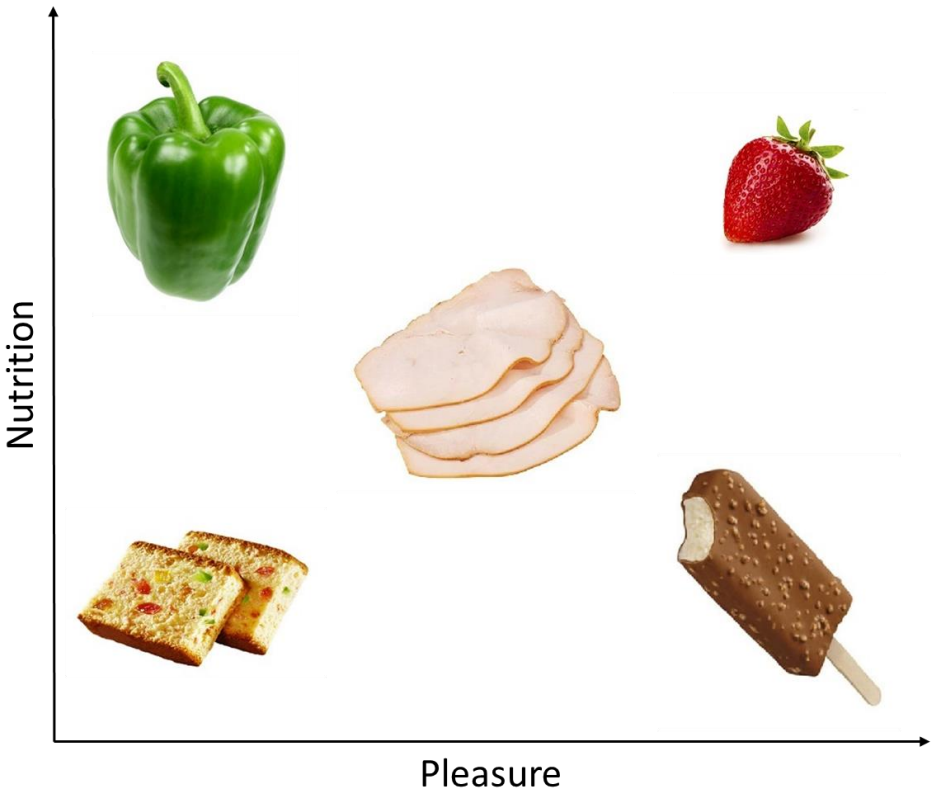

Figure 1. An example of snack foods included in the stimulus set, plotted according to their respective ratings along the dimensions of pleasure and nutrition.

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. 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. 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

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 mean SoA of 3.70 in overall value across all items ($p < .001$), which is comparable in the magnitude observed in previous studies (Lee & Daunizeau, 2020b; Voigt et al, 2019; Izuma et al, 2010).

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, 2020b), $|dV_1|$ had a reliable negative correlation with overall SoA and a reliable positive correlation with choice confidence. We ran general linear regression models (GLM) at the participant level, after converting all variables to *z*-scores within participant, regressing $|dV_1|$ on overall SoA and on confidence, and tested the beta weights across participants using two-sided *t*-tests. 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 ($p < .001$).

Lee and Daunizeau (2020b) 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 is larger when $|dV|$ is calculated using post-choice ratings (i.e., $|dV_2|$) rather than pre-choice ratings. This finding 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). This beta weight was reliably greater than that obtained using $|dV_1|$ (p = .003).

Lee and Daunizeau (2020b) 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 (Lee & Daunizeau, 2020b). 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).

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 (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. 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]'):

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

Using this formal definition of disparity, we performed GLM regressions at the participant level (after converting all variables to *z*-scores within participant) 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 and choice confidence. 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, $p = .013$) (see Figure 2).

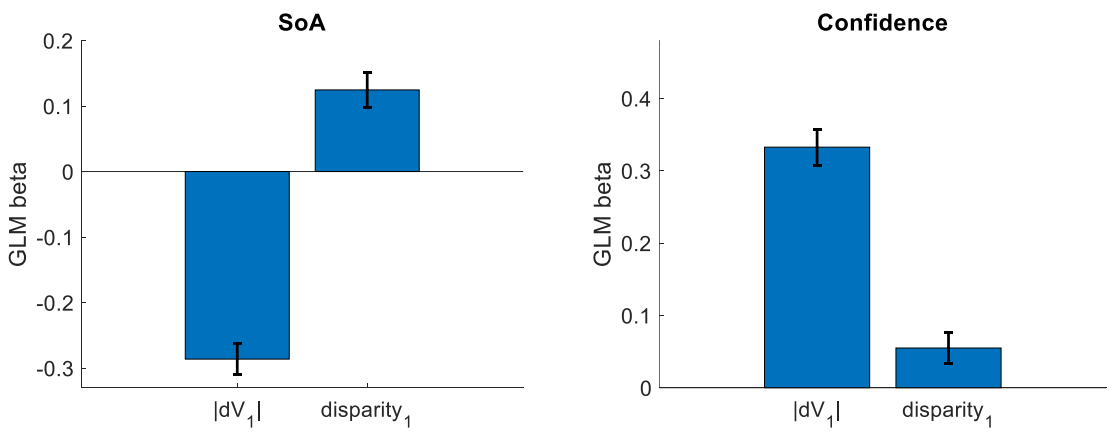

Figure 2. Impact of $|dV_1|$ and $disparity_1$ on SoA (left) and on confidence (right). Bar heights represent cross-participant mean GLM beta weights; error bars represent standard error of the mean.

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 tested our predictions that $|dP_1|$ would predict SoA_P and that $|dN_1|$ would predict SoA_N by running GLM regressions at the participant level, after converting all variables to *z*-scores within participant. 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 a GLM regression of both $|dP_1|$ and $|dN_1|$ on choice confidence. Across participants, beta weights were positive and significant for both independent variables (mean $|dP_1|$ beta = 0.241, $p < .001$; mean $|dN_1|$ beta = 0.081, $p < .001$). 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, $p = .01$), but that for $|dN_2|$ did not differ from that for $|dN_1|$ (mean beta difference = -0.012, $p = .432$) (see Figure 3).

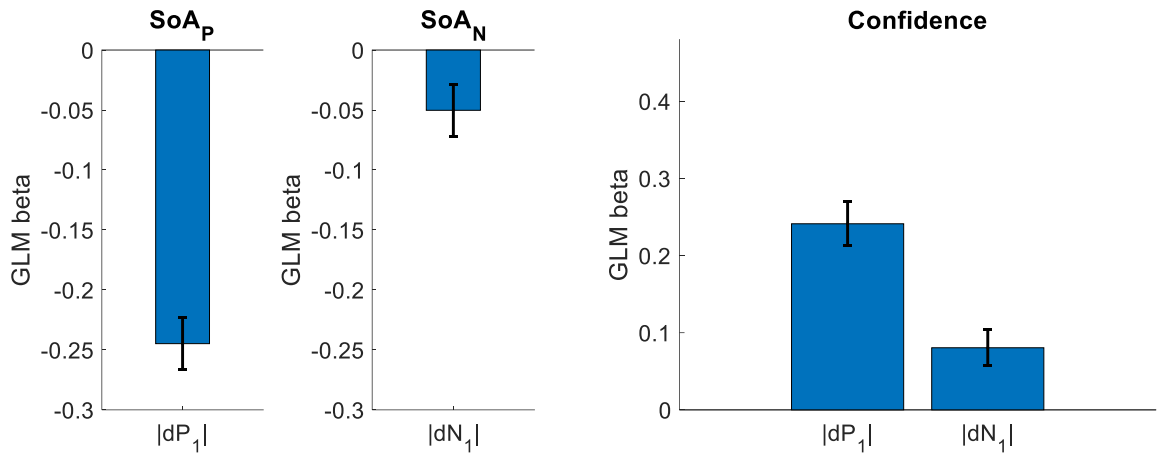

Figure 3. Impact of $|dP_1|$ on SoA_P and $|dN_1|$ on SoA_N (left); and of both $|dP_1|$ and $|dN_1|$ on confidence (right). Bar heights represent cross-participant mean GLM beta weights; error bars represent standard error of the mean.

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). 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 $SoA_1 = 2.46$, $p < .001$).

We then ran GLM regressions of $|dV_1|$ on SoA_1 and on eventual choice confidence. Both 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). We then performed a GLM regression of $|dV_1|$ and SoA_1 on confidence. The mean beta weight for $|dV_1|$ was 0.438 (p < .001) and for SoA_1 was 0.215 (p < .001). We also conducted GLM regressions of $|dV_2|$ on confidence. The beta weight for $|dV_2|$ was significant and in the predicted direction (mean = 0.458, p < .001). Replicating the comparable finding in Experiment 1, the beta weight for $|dV_2|$ on confidence was significantly higher than for $|dV_1|$ (mean difference = 0.073, p = 0.001) (see Figure 4). Thus, value differences after coherence shifts were better predictors of choice confidence than were the initial value differences.

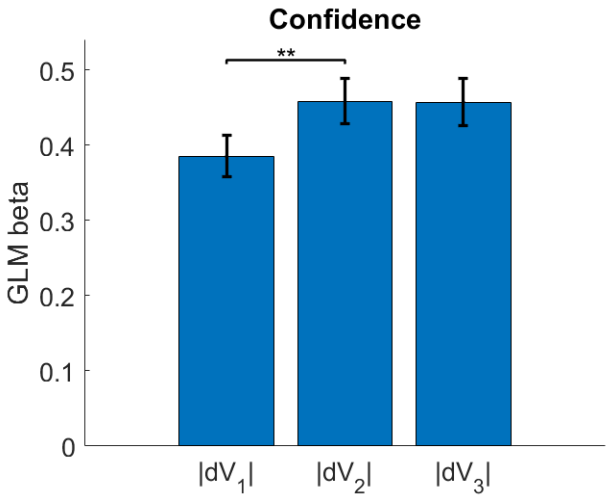

Figure 4. Impact of |dV| on choice confidence, separately for each round of ratings. Bar heights represent cross-participant mean GLM beta weights; error bars represent standard error of the mean.

Finally, we ran GLM regressions of $|dV_1|$ and $disparity_1$ on SoA_1 and choice confidence. Across participants, all beta weights were significant and in the predicted direction (mean $|dV_1|$

beta for SoA_1 = -0.246, $p < .001$; mean $disparity_1$ beta for SoA_1 = 0.130, $p < .001$; mean $|dV_1|$ beta for confidence = 0.368, $p < .001$; mean $disparity_1$ beta for confidence = 0.075, $p = .002$) (see Figure 5).

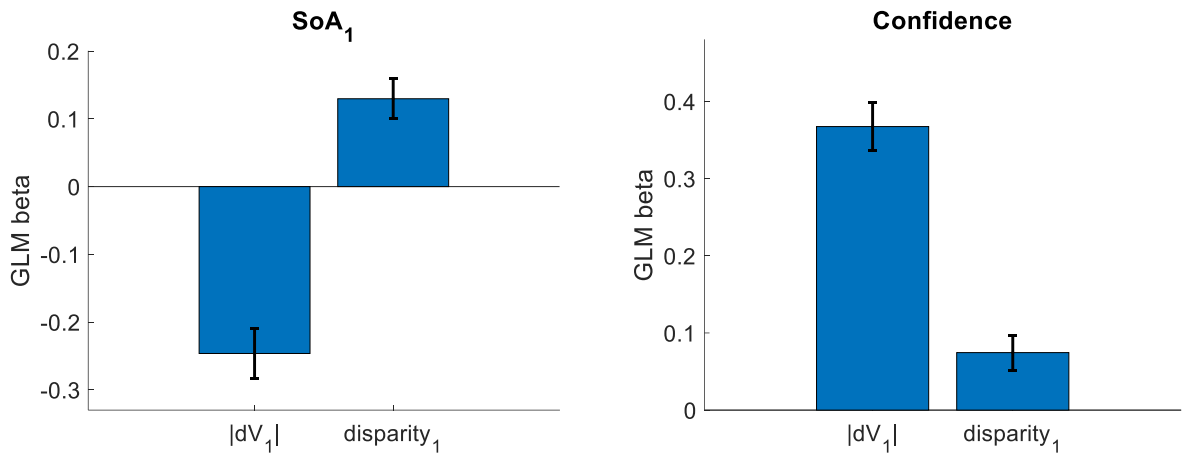

Figure 5. Impact of $|dV_1|$ and $disparity_1$ on SoA_1 (left) and on confidence (right). Bar heights represent cross-participant mean GLM beta weights; error bars represent standard error of the mean.

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 SoA_1 was 2.46, SoA_2 (i.e., the additional shift from phase 2 to final phase 3) was 1.07, an increment that fell just short of statistical significance ($p = .07$). Similarly, while $|dV_2|$ was a more accurate

predictor of confidence than $|dV_1|$ (see above), $|dV_3|$ was no more accurate as a predictor than was $|dV_2|$ (mean difference = -0.003, p = .95) (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

The present study provides several new insights 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 (Voigt et al, 2019; Lee & Coricelli, 2020; Lee & Daunizeau, 2020a, 2020b), but also for each of the two individual attributes that determined value (cf. Simon et al., 2004). Moreover, as predicted by the MCD model (Lee & Daunizeau, 2020a), which postulates metacognitive influences on choice deliberation, 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, with confidence being more accurately predicted by value difference measured by ratings obtained after the choice.

Another novel contribution of the present study was the demonstration that coherence shifts are predicted not only by the overall value difference between options, but also by the pattern of attribute composition across options. 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 in the eventual choice.

Experiment 2 provided what to the best of our knowledge is the most direct evidence yet that coherence shifts occur prior to choice and guide its outcome, rather than being post-decisional rationalizations that reduce cognitive dissonance for a decision already made (Festinger, 1957). We hypothesized that the core component of the decision process that drives coherence shifts is an active comparison of the values of the options. 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 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.

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, Dolan, & Behrens, 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. The MCD model (Lee & Daunizeau, 2020a) provides a computational account of why and how coherence shifts can be generated during intra-decisional processing. Future studies using multiple methodologies will hopefully provide a deeper understanding of the timing of coherence shifts at the attribute level.

Acknowledgements

Preparation of this paper was supported by NSF Grant BCS-1827374.

References

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

Bettman, J. R., Luce, M. F., & Payne, J. W. (1998). Constructive consumer choice processes. *Journal of Consumer Research*, *25*(3), 187-217.

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.

Busemeyer, J. R., Gluth, S., Rieskamp, J., & Turner, B. M. (2019). Cognitive and neural bases of multi-attribute, multi-alternative, value-based decisions. *Trends in Cognitive Sciences*, *23*(3), 251-263.

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.

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.

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

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

Lee, D., & Daunizeau, J. (2020b). 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.

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.

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., 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.

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.