Research Article

Construction of Preferences by Constraint Satisfaction

Dan Simon,1 Daniel C. Krawczyk,2 and Keith J. Holyoak3

1University of Southern California; 2University of California, Berkeley; and 3University of California, Los Angeles

ABSTRACT—Participants were given a choice between two multiattribute alternatives (job offers). Preferences for the attributes were measured before, during, and after the choices were made. We found that over the course of decision making, the preferences shifted to cohere with the choice: The attributes of the option that was eventually chosen came to be rated more favorably than they had been rated initially, while the attributes of the rejected option received lower preference ratings than before. These coherence shifts were triggered by a single attribute that decisively favored one option (Experiment 1), and occurred spontaneously in the absence of a decisive attribute (Experiment 2). The coherence shift preceded commitment to choice. These findings favor constraint-satisfaction models of decision making.

A central tenet of classical theories of rational choice is that people harbor a stable, well-defined, and discernable order of preferences, and have computational skills that enable them to choose the courses of action that maximize their preferences. A paradigmatic example of a classical theory is multiattribute decision theory, which prescribes that the utility of a choice is equivalent to the sum of its preferences, that is, the sum of the weighted values of its attributes (Edwards & Newman, 1982; Keeney & Raiffa, 1976).

Research has challenged the axiom of preference invariance. Rather than being stable, well-defined, and discernable, preferences have been shown to be constructed (Markman, Zhang, & Moreau, 2000; Slovic, 1995); to some degree, they are labile, reversible, and obscure (for a review, see Payne, Bettman, & Johnson, 1992). Preference invariance is violated under different descriptions of essentially the same options (Tversky & Kahneman, 1986), when different modes of elicitation are invoked (Lichtenstein & Slovic, 1971), and when options are presented in different contexts (Shafr, Simonson, & Tversky, 1993). Various process theories claim that preferences are reconstructed to create dominance and thus bring the decision task to a point of commitment (Janis & Mann, 1977; Montgomery, 1983; Svenson, 1992; for a review, see Brownstein, 2003). These process theories are inconsistent with cognitive dissonance theory, which posits that changes in preferences are exclusively a postcommitment phenomenon (see Festinger, 1964, p. 153; also Simon & Holyoak, 2002).

Although we are in agreement with theories that posit restructuring for dominance, it must be acknowledged that they do not follow from a general psychological theory, and experimental evidence supporting them is rather limited. The best available evidence is derived from a methodology that measures values, but not weights, of attributes, and could be challenged for using sequential measurement that might interfere with the natural flow of the decision process (Russo, Medvec, & Meloy, 1996; Russo, Meloy, & Medvec, 1998). One general theoretical framework that could explain certain forms of preference construction is the connectionist approach to constraint satisfaction (Holyoak & Thagard, 1989; Read, Vanman, & Miller, 1997; Rumelhart & McClelland, 1986; Thagard, 1989). Connectionist networks perform constraint satisfaction by applying a relaxation algorithm that settles the network into a stable state in which asymptotic activation levels define a set of coherent variables. Bidirectional links enable units that “go together” to become highly active, and to collectively inhibit their rivals, thus becoming increasingly coherent with the emerging decision (coherence implies a state in which positively related variables are similarly activated). Evidence of constraint-satisfaction processing has been obtained in inference-based judgments (Holyoak & Simon, 1999; Read & Marcus-Newhall, 1993; Simon, Pham, Le, & Holyoak, 2001; Spellman, Ullman, & Holyoak, 1993), probabilistic judgments (Simon, Snow, & Read, in press), and analogical reasoning (Spellman & Holyoak, 1992). Work on legal decision making (Holyoak & Simon, 1999) has shown that coherence shifts in one decision task can trigger allied shifts in a subsequent decision task involving similar underlying issues.

The objective of the present study was to examine constraint-satisfaction processing in a realistic preference-based choice task, and to determine whether coherence (or dominance) is achieved by restructuring the preferences—their values, weights, or both. The experimental design allowed direct, within-subjects measurement of changes in preferences.

EXPERIMENT 1

Experiment 1 was designed to determine whether people’s assessments of a variety of attributes systematically shift to favor one choice over another. The experiment involved a choice between job offers,
and was designed to emulate the type of decision task to which multiattribute decision-making theory is most readily applied.

Method

Participants
Participants were 80 undergraduates at the University of California, Los Angeles (UCLA), who took part in the experiment to partially fulfill a course requirement for an introductory psychology class. Participants were run in groups ranging from 1 to 3 persons.

Materials
Two instruments were used. The first was a baseline test that was presented before participants learned about the job offers. In this five-page instrument, called “Waiting For a Job Offer,” participants were told to imagine that they were about to graduate from college and were interviewing for a job in the field of marketing. While waiting to receive an offer, they were asked to state their evaluation of job-related features “that might be included in job offers.” The booklet contained 11 statements describing a variety of job attributes that participants were to judge on a 10-point scale from 1 (undesirable) to 10 (desirable). Our focus was on eight attributes, one high and one low, on four dimensions: commute, office, vacation, and salary. Each of the eight attributes appeared in one of the 11 statements (3 were used as distractor statements relating to aspects of job choice not tested in the second instrument of the experiment). After completing the desirability task, participants were asked to rate the importance of each of the four dimensions, assuming that they were included in a job offer. Each dimension was delimited by values that corresponded to its high and low attributes. The importance ratings were made on a 9-point scale ranging from 0 (no weight) to 8 (maximum weight). The order of statements for both the desirability and importance tasks was counterbalanced to control for order effects.

In the second instrument, called “Choosing Your Next Job,” participants were told that they had received job offers from two large department store chains, called Bonnie’s Best (“BB”) and Splendor. The companies were described as being similar in size, reputation, and opportunities for promotion. Participants were also informed that they had met with key personnel at the two companies and found them both to be stimulating and pleasant.

The jobs differed in several key aspects. Four attributes were described to each job such that each offer had two positive attributes and two negative attributes. The commute to Splendor was short (18 min), and Splendor offered a private office, but it also paid a low salary ($600 less than the industry’s average of $40,000) and offered minimal time off for vacation. The four attributes varied in the opposite manner for BB: It offered a higher salary ($40,300) and superior vacation package, but the commute to its offices was longer (40 min), and it offered only a cubicle. The eight attributes contained in the offers were the same ones that had been tested in the baseline measure. In an effort to manipulate participants’ decisions in favor of a particular choice, one job was described as being located in a fun part of town with good shopping and restaurants (“good-location” attribute), and the other was located in a dull, industrial part of town (“bad-location” attribute).

In the second instrument, participants were asked to report their choice between the offers and their confidence in that choice (on a scale from 0 to 5, with 5 representing maximal confidence), and were then asked for desirability evaluations for the eight attributes and importance evaluations for the four dimensions. The questions eliciting the preferences and weights were the same as those in the first instrument, except that they were worded in terms of the job offers. The order of questions was varied, as was the presentation order of the job offers.

Design and Procedure
Across two between-subjects conditions, we varied which company was associated with the good-location attribute. For half of the participants, BB had the good-location attribute whereas Splendor had the bad-location attribute, and for the other half of the participants, the reverse was true. All participants completed the experiment in two phases. In the first phase, they completed the baseline test. After this booklet was collected, they completed a 5- to 10-min unrelated reasoning task. In the second phase (the posttest), participants received a booklet containing the job offers and the second instrument.

Results
The data were initially analyzed to determine whether the manipulation of overall attribute goodness (inclusion of the good- or bad-location attribute) was predictive of the decision. The manipulation was effective, as 92% of participants in the good-Splendor condition decided in favor of the Splendor offer, and 72% of participants in the bad-Splendor condition decided in favor of the BB offer. \( \chi^2(1, N=80) = 33.80, p < .001 \). Confidence in the decisions was high, with means of 4.20 for BB choosers and 4.13 for Splendor choosers (with 5 representing maximum confidence). These means were not significantly different, \( p > .05 \). The phenomenon of high decision confidence despite the inherent ambiguity of the input parallels findings for legal decision making (Holyoak & Simon, 1999; Simon et al., 2001) and provides evidence for a constraint-satisfaction process.

To test whether participants’ assessments of the different attributes shifted to fit with their decision, we first analyzed the desirability data after linearly rescaling all desirability ratings to range from −1 to 1. (This transformation was performed so that we could later multiply the desirability ratings with the importance ratings using a comparable scale.) In order to measure participants’ overall evaluation of the desirability of the attributes of each job choice, we converted the ratings obtained into values that we call S scores, which provide an index of desirability for the Splendor job. The S score was computed by summing the average rating for the eight attributes (excluding location, which would create a confounding), reversing the scale for the attributes favoring BB (see Holyoak & Simon, 1999, for a similar coherence analysis). High S scores indicate strong preferences for Splendor’s positive attributes and low preference for BB’s positive attributes, and low S scores indicate the opposite.

Mean desirability ratings were compared using a 2 (decision group) \( \times 2 \) (test phase) mixed-model analysis of variance (ANOVA), which revealed a highly significant interaction between decision group and phase, \( F(1, 78) = 18.73, p < .001 \). The rated desirability of the composite of Splendor attributes was higher at the posttest than the baseline test for Splendor choosers, whereas the rated desirability of the composite of Splendor attributes was lower at the posttest than the baseline test for BB choosers.
We then tested for differences between the importance ratings for the attributes before and after the decision. Prior to running statistical tests, we rescaled the importance data such that the data ranged from 0 to 1, with 0 representing lowest possible importance. This was done in order to compute a composite utility measure, as described in the next paragraph. The analysis of the importance ratings was performed by running a separate mixed-model ANOVA for each attribute, with decision group as a between-subjects variable and phase as a within-subjects variable. These analyses revealed significant interactions for the office attribute, $F(1, 78) = 57.42, p < .01$; the commute attribute, $F(1, 78) = 511.15, p < .01$; and the vacation attribute, $F(1, 78) = 6.56, p < .05$. Post hoc Newman-Keuls tests revealed that ratings of the attributes that were favorable for Splendor choosers (office and commute) differed reliably between the two decision groups at the post-tests ($p < .01$), but not at the baseline tests. Overall, the results indicate that for all attributes except salary, the importance ratings of the two decision groups diverged from the baseline to the posttest, with means generally increasing for those traits that were desirable attributes of the chosen job and decreasing for those traits that were undesirable attributes of the chosen job (see Fig. 1).

To provide an integrated measure of coherence using both desirability of attributes and importance weights, we computed the product of attribute desirability (scaled from $0$ to $1$) and attribute importance (scaled from $0$ to $1$) to make an integrated $S$ score. This integrated $S$ score provides an analogue of utility as it is conventionally defined in multiattribute utility theory. A mixed-model ANOVA with decision group as a between-subjects variable and test phase as a within-subjects variable revealed a highly significant interaction between decision group and phase, $F(1, 78) = 26.80, p < .001$. This analysis of integrated $S$ scores yielded the same basic trends as were observed for the desirability scores alone, and the differences were somewhat more reliable (see Fig. 2).

Finally, we determined whether attribute assessments changed as a coherent set by looking for positive correlations among the various attribute shifts that accompanied the decision. A correlational analysis using Cronbach’s alpha was carried out on the $S$ scores (not integrated $S$ scores) for the four critical attributes and the decision (represented by 0 for BB choosers and 1 for Splendor choosers). A Cronbach’s alpha of near 0 would indicate no correlation between the decision and the desirability of the set of attributes, and an alpha of 1 would indicate the highest possible correlation. This analysis revealed that the overall positive correlations were low at the baseline test ($Cronbach’s \alpha = .22$), but they were considerably higher at the posttest ($Cronbach’s \alpha = .48$). Thus, it appears that the overall mean shifts discussed earlier were accompanied by correlational shifts among the attributes.

**EXPERIMENT 2**

The goal of Experiment 2 was to assess whether the coherence shifts observed in Experiment 1 are a postcommitment phenomenon, or whether they occur prior to the point of commitment to a final decision. It has often been claimed that the point of commitment (i.e., a declaration of a “final decision”) is psychologically critical to triggering dissonance reduction. According to Festinger (1964), “dissonance-reduction processes do not automatically start when a decision is made. Simply making a decision is not enough. The decision must have the effect of committing the person” (p. 42; see also Gollwitzer & Bayer, 1999; Kiesler, 1971; Wicklun & Brehm, 1976). In contrast, constraint-satisfaction models imply that coherence shifts...
begin prior to the point of commitment, and in fact drive the decision-making process.

Method

Participants
Participants were 80 UCLA undergraduates who were recruited as in Experiment 1.

Materials, Design, and Procedure
The baseline test instrument was presented in a five-page booklet, and was identical to that used in Experiment 1, with the exception that the long commute was changed from 40 to 35 min (the same change was made in the other instruments).

The second and third instruments were presented together in a single 10-page booklet. In order to determine whether participants would demonstrate coherence shifts in the absence of any manipulation that would lead them to choose one option over the other, we removed the location manipulation used in Experiment 1. The other main difference from Experiment 1 was the addition of a section of text presented alongside the description of the job offers. This text, designed to delay the point of commitment, stated that a large corporation named Punch was considering buying out either BB or Splendor, and a buy-out would eliminate all new jobs in the respective company. The instructions advised the participants to consider the offers, but to delay their final decision until the situation was clarified. After the descriptions of the job offers were presented, the instructions read: “At this stage you are still waiting to hear about the decision by Punch and are suspending your own decision. However, you may have a feeling about the aspects included in the two job offers.” Ratings of desirability and importance were then obtained with the questions framed in terms of the job offers. Finally, at the end of this second instrument (interim phase), the instructions stated: “Even though you are still waiting to learn about Punch’s decision, you may have a preliminary leaning toward either one of the two offers. Please indicate your preliminary leaning. Remember that you will still be free to make any decision you like after you are told about Punch’s plans.”

After indicating their preliminary leaning, participants were asked to point of commitment, and in fact drive the decision-making process.

Results
The data were initially tabulated to determine the number of participants who had decided in favor of each of the two offers, and whether participants had switched decisions from the interim to final decision. A total of 44 participants tentatively chose BB in the interim phase, whereas 36 chose Splendor. Four participants who were leaning toward BB in the interim phase switched their final decision to Splendor and were dropped from all subsequent analyses. The confidence in the decisions was generally high, with means of 3.48 for tentative BB choosers, 3.44 for tentative Splendor choosers, 3.70 for final BB choosers, and 3.50 for final Splendor choosers. None of these means differed reliably.

We then tested for coherence shifts using the desirability scores. As in Experiment 1, this analysis was performed after first converting all of the ratings to a standard range of -1 to 1. S-score data were analyzed using a 2 x 3 mixed-model ANOVA, which revealed a significant interaction between decision and test phase, \( F(2, 146) = 9.62, p < .001 \). Post hoc Newman-Keuls tests revealed that for BB choosers, both the interim mean S score (\( M = - .22 \)) and the posttest score (\( M = - .27 \)) were significantly lower than the baseline score (\( M = - .07 \)), but the interim and posttest scores did not differ reliably. These differences suggest that the BB choosers shifted toward liking BB more and Splendor less both prior to and after they had committed to a decision, but not before the offers were made. The pattern revealed by Newman-Keuls tests was less clear for Splendor choosers. For them, mean S scores increased from the baseline test (\( M = .06 \)) to the interim test (\( M = .07 \)) to the posttest (\( M = .11 \)); however, these differences did not reach significance, \( p > .05 \). Further Newman-Keuls tests revealed that the two decision groups differed significantly at all phases of the experiment.

As in Experiment 1, we tested for differences among the importance ratings in each phase of the experiment by running a separate mixed-model ANOVA for each attribute, with decision as a between-subjects variable and phase as a within-subjects variable. These analyses revealed a significant interaction for the commute attribute, \( F(2, 148) = 5.80, p < .01 \). The office attribute approached, but did not reach, significance, \( F(2, 148) = 2.72, p = .06 \). There were no significant interactions for either the salary or the vacation attribute. Overall, the results of Experiment 2 were less robust than those of Experiment 1. Nonetheless, the general trends of the importance ratings indicated that the ratings of the two decision groups gradually diverged from the baseline to the interim to the posttest, with means generally increasing for those traits that were desirable attributes of the chosen job and generally decreasing for those traits that were undesirable attributes of the chosen job, though these patterns were absent for BB choosers on the salary and vacation attributes.

To further assess coherence shifts, we again calculated integrated S scores by taking the product of desirability and importance. Figure 3 depicts the resulting mean integrated S scores. A 2 x 3 mixed-model ANOVA revealed a significant interaction between decision group and test phase, \( F(2, 148) = 14.49, p < .001 \). Post hoc Newman-Keuls tests identified patterns similar to those observed for the desirability scores alone. For BB choosers, both the interim mean S score (\( M = - .23 \)) and the posttest score (\( M = - .27 \)) were significantly lower than the baseline score (\( M = - .11 \)), but the interim and posttest scores did not differ reliably. These differences suggest that the BB choosers shifted toward valuing BB more and Splendor less both prior to and after they had committed to a decision, but not before the offers were made. For Splendor choosers, the posttest score (\( M = .11 \)) was significantly higher than both the baseline and interim scores; however, the change from the baseline test (\( M = .07 \)) to the interim test (\( M = .08 \)) was not reliable. As in the analysis of desirability alone, the two decision groups differed significantly at each phase of the experiment.

As in Experiment 1, a correlational analysis was carried out using Cronbach’s alpha to determine whether there were positive associations between participants’ ratings of desirability and importance.
correlations among the decision and the supporting attributes. At the baseline test, there were few positive correlations (Cronbach's $\alpha = -0.58$). At the interim test, however, this changed, and the correlations overall became positive (Cronbach's $\alpha = 0.13$). At the posttest, the correlations increased again (Cronbach's $\alpha = 0.26$), indicating that increased coherence among the desirability scores accompanied the overall attribute shifts.

**GENERAL DISCUSSION**

The present study demonstrates that preferences used in decision making are not fixed, as assumed by classic theories of rational choice, but rather are reconstructed in the course of decision making (cf. Janis & Mann, 1977; Montgomery, 1983; Russo et al., 1996, 1998; Svenson, 1992). Our principal finding is that as people processed the decision task, their preferences for the attributes of the alternative that was ultimately chosen increased, while their preferences for the attributes of the to-be-rejected choice decreased. These results constitute a violation of the axiom of preference invariance (though they do not provide evidence of preference reversals). In general, both the reported values of the attributes (ratings of desirability) and their weights (ratings of importance) shifted to make one alternative dominate the other. The most robust coherence shifts involved the product of these two factors (a close analogue to the overall utility of a choice as defined in multiattribute decision theory). Other dynamic models of decision making (e.g., decision field theory; Busemeyer & Townsend, 1993) predict divergence of the choice options, but not changes in assessments of the attributes underlying them. The present evidence that the attributes themselves are reevaluated (before the point of commitment) thus provides evidence for constraint satisfaction over other dynamic accounts. These findings cannot be attributed to differences in methods used to elicit or describe the options, nor to variations in context (cf. Slovic, 1995; Tversky & Kahneman, 1986). Rather, the reconstruction of preferences seems to be the natural outcome of the very process of decision making.

As we have found in previous studies that tested decisions involving high-level reasoning (Holyoak & Simon, 1999; Simon et al., 2001), evidence integration (Simon et al., in press), and social reasoning (Read, Snow, & Simon, 2003), decisions were accompanied by coherence shifts and high levels of confidence. This suggests that constraint-satisfaction processing provides a good explanation for a broad range of decision-making tasks. Experiment 1 demonstrated that the introduction of one piece of strong evidence (the good-location/hard-location manipulation) triggers changes in the evaluation of unrelated attributes (cf. Holyoak & Simon, 1999, Experiment 3). This finding is reminiscent of classical perceptual phenomena associated with the ambiguous Necker cube, for which a shift in the interpretation of one portion of the figure causes shifts in the interpretation of all other portions. Our results indicate that a variation in one attribute not only determines the decision (which would be entirely rational), but also causes a global coherence shift involving changes in preferences for other logically unrelated attributes. Experiment 2 provided evidence that preference changes occur both before and after the point of commitment, as has been found for other types of decisions (Phillips, 2002; Simon et al., 2001) and using different methodological designs (Russo et al., 1996, 1998).

The magnitudes of the shifts observed in these studies were smaller than those observed in our previous work with legal cases (Holyoak & Simon, 1999; Simon et al., 2001). One reason for this difference is that the arguments that figured in the studies of legal decision making were more abstract and ambiguous. Moreover, many of the legal issues involved competing attributes (e.g., alternative precedents), either of which might reasonably be viewed as superior. In contrast, the attributes involved in the present study varied monotonically in goodness (e.g., a higher or a lower salary), thus effectively blocking cardinal reversals of preference.

Although our findings challenge the descriptive validity of formal decision-making theories, they do not challenge their normative or prescriptive value. Following the procedure implied by multiattribute decision theory may assist decision makers in gaining insights into their values and goals, help people communicate about their respective values and goals (Baron, 2000; Edwards & Fasolo, 2001), and perhaps serve as a benchmark for identifying whether and how far one's preferences actually shifted in the course of making a decision. At the same time, constraint satisfaction provides an adaptive psychological mechanism that enables people to construct dominance and thus reach confident decisions in complex cases.

**Acknowledgments**—This research was supported by National Science Foundation Grants SES-0080424 and SES-0080375.

**REFERENCES**


(Received 11/20/02; Revision accepted 2/26/03)