APPLIED COGNITIVE PSYCHOLOGY, VOL. 9, 459–484 (1995)

A Longitudinal Study of Real-Life Decision Making: Choosing a College[1]

KATHLEEN M. GALOTTI
Carleton College, Northfield MN, USA

SUMMARY

Three hundred and twenty two college-bound high school students described aspects of their college decision-making processes. Students listed their criteria as well as the alternatives (i.e., schools) under consideration, rated the importance of each criterion, and rated each alternative with respect to each criterion. They also gave their overall impressions of each alternative. Finally, students rated their comfort with the decision-making process and, at the conclusion of the study, reported on how many schools they had applied to, had or had not been accepted at, were waiting to hear from, or were waitlisted at. Students consider four or five alternatives, and use eight to ten criteria in evaluating them. These figures do not change appreciably over the course of the process, although only about half the criteria or/and slightly more than half of the schools considered at one time are considered again 6 months later, and there are several changes in the *kinds* of criteria considered at different points in time. There was a marginally significant trend for higher ability and average ability students to consider more criteria, more distinct types of criteria, and more alternatives than do lower ability students. There were no gender differences in this regard. Gender differences and academic ability group differences were apparent, however, in the types of criteria students reported. Participation in multiple sessions in this study had few reliable effects on decision-making performance. Students were given a list of 34 standard criteria at each session, and incorporated some of these into their own lists of criteria during subsequent sessions. However, there was no indication that repeated participation led students to adopt a more analytical strategy than they would have otherwise. Data were compared with three linear models of information integration. Models using data with multiple criteria better fit the students' data than did a model using only the most important criterion. Higher ability students were particularly better able to integrate information according to linear models.

How do people make real-life decisions? A vast literature documents people's biases and use of faulty heuristics on a number of laboratory decision-making tasks (see Slovic, Lichtenstein, and Fischhoff, 1988, for a review), and anecdotal evidence of

[1]This project was funded by The Spencer Foundation and by the National Institute of Child Health and Human Development (HD-28177-R15), both to Kathleen M. Galotti.

I gratefully acknowledge the assistance of Stephanie Aubry, Jessica Bresler, Laura Gibson, Julie Greene, Ruby Hou, Perrin Klumpp, Jodi Matthes, Jennifer McGregor, Nancy Michelsen, Terrance Overby, Amy Prullage, Melissa Schneider, Katherine Seiden, Daniel Simons, Meghan Wilhelm, and especially Melissa Mark and Sara Voelz for gathering and coding data. I am grateful for the efforts of the principals, guidance counselors, and teachers at the cooperating schools in helping us recruit students and schedule survey sessions. Lastly, I thank Jonathan Baron, Lloyd Komatsu, Larry Litten, Michael Pressley, and anonymous reviewers for comments on earlier versions of this paper.

Address for correspondence: K. M. Galotti, Department of Psychology, Carleton College, One North College St, Northfield MN 55057-4025. (email: kgalotti@carleton.edu).

CCC 0888–4080/95/060459–26

Received 2 November 1994
Accepted 5 December 1994

'botched' decisions and flawed processes abounds. Some existing literature describes real-life decision making, by focusing on the processes used by professionals in their domain of expertise (e.g. Hammond, Hamm, Grasia and Pearson, 1987; Lusk and Hammond, 1991). However, we still know surprisingly little about what non-experts do when they make major life decisions.

In large part, this gap in our knowledge may derive from the sheer difficulty of studying real-life decision making. Most real-life decisions of consequence are not on a predictable schedule, making recruiting and observation of research participants difficult. It is also not easy to determine what information people have available to them at what point(s) in the process. Moreover, it is very difficult to assess the optimality of the final decision, without knowing what the decision makers' own objectives and values are.

It is also possible that the processes of decision making change over time, especially as the decision point looms closer. Early in the process, people may spend most of their time identifying what sources of information relevant to the decision are available or screening the options (Beach and Potter, 1992). Later, they may try to identify alternatives or develop criteria against which to measure different alternatives. Wright and Kriewall (1980), for example, reported that mental processes used in a laboratory study of decision making appear to become more thorough as the time for making the final choice draws near.

There may also be individual differences in the style with which a decision maker acts. Some may have a very analytical approach from the beginning; others might rely most on 'gut-level' intuitions; others might switch from one style to another at different points in the process. This stylistic component of decision making may vary as a function of gender, socioeconomic status, level of education, intelligence, or any number of other factors.

This paper reports on a descriptive longitudinal study of how a particular real-life decision unfolds over the course of a year. The aim here is to describe some measures of decision-making performance, specifically how decision makers generate criteria, weigh the importance of those criteria, consider alternatives, and integrate information about alternatives and criteria. The real-life decision chosen for study was the choice of what college to attend. The decision makers studied in this investigation, then, were college-bound high school students.

The 'college decision'

The life decision examined here—choice of college—was selected for a variety of reasons. First, it is an important and difficult life decision, faced by many adolescents and their families. This decision has ramifications for family ties, friendships, and vocational and career plans. The choice of college is likely to determine whether the student will live away from home, will need loans, will maintain the same relationships with high school friends. Moreover, the choice of college may determine, or at least influence, the chances of success of other desired objectives—career opportunities, probability of admission to graduate or professional school, intellectual development, or social status.

This decision is one that occurs during a well-delimited time period and on a well-defined schedule. In very few other important life decisions can a large cohort of decision makers be found all of whom are a predictable period of time away from the final decision. This allows us to make better predictions about where in the process a student is likely to be at any given point. For example, colleges typically establish deadlines of late winter for applications and announce their admissions decisions in the spring. Thus, this particular decision has some convenient properties from a research standpoint.

For many students, this is the first major financial, educational, social, and vocational decision they have had much responsibility for and choice in. Thus, to a certain extent, we have a group of participants who are fairly homogeneous with respect to their level of 'practice' (i.e. little) with making important life decisions. This homogeneity is important in an exploratory study, because trends in the data will not be obscured by variance stemming from differential levels of experience.

Finally, like most complex decisions, the choice of college requires the student to seek out and integrate information from various sources. This decision can easily consume several dozen hours (or days or weeks) of a decision-maker's time. Information on colleges is readily available from such diverse sources as high school guidance counselors and the guidebooks, brochures and pamphlets (and, more recently, videotapes) produced by individual colleges and universities. Indeed, there is more information than any one individual can hope to process thoroughly in the year or so in which college decisions are typically made. Thus, this decision is one rich enough to support various approaches and strategies of information gathering and synthesis.

Adolescent decision making

Developmental psychologists see adolescents as developing the capabilities to make mature decisions. Adolescents are often described as possessing a plethora of newly emerging cognitive abilities, including the ability to think about possibilities, distinguishing the real from the possible; the capacity to think of hypotheses and how to test them; the capacity to think ahead, and to plan and anticipate consequences; the propensity and ability to reflect on their own thinking; and the ability to think beyond old limits, and to discover new horizons (Keating, 1980, 1990).

Decision making appears to be almost a defining activity of adolescence. Indeed, researchers studying identity development assess it in part by the degree to which the adolescent reports experiencing *conflict* (i.e. seeing more than one realistic and probably mutually exclusive alternative) and making *commitments* (i.e. choosing among alternatives in such domains as occupation, ideology, or romantic partner (Marcia, 1966; Rowe and Marcia, 1980, Scheidel and Marcia, 1985).

Much of the existing literature on adolescent decision making examines adolescents assessing the consequences of potentially risky behaviours, such as drinking or using drugs. In one such recent study (Beyth-Marom, Austin, Fischhoff, Palmgren, and Jacobs-Quadrel, 1993) adolescents and their parents were asked to consider the possible consequences of a teenager hypothetically either engaging in a risky behaviour or not engaging in it. Results showed similar response patterns for adults and adolescents, a finding that supports the contention of Ormond, Luszcz, Mann, and Beswick (1991) that, by about the age of 15, adolescents make decisions in similar ways to adults.

Assessing the quality of decision making

Existing literature on decision-making performance has documented a number of biases and other shortcomings (Arkes and Hammond, 1980; Baron, 1988;

Kahneman, Slovic, and Tversky, 1982; Slovic *et al.*, 1988). In particular, research has repeatedly shown that people too often stop thinking too early, ignore evidence, and integrate information in ways that favour their initial biases (Baron, 1988; Perkins, Allen, and Hafner, 1983). Moreover, much existing literature demonstrates that people are quite sensitive to the perceived effort required to integrate information (Bettman, Johnson and Payne, 1990; Creyer, Bettman, and Payne, 1990; Johnson, Payne, and Bettman, 1988).

However, many of these studies employed formal, well-defined decision-making tasks, for which correct solutions exist (Galotti, 1989). In contrast, when choosing a college there is no single, 'correct' choice to make. Instead, any number of choices might all be feasible or plausible, leaving the decision maker to find means by which to choose. Indeed, the college decision, unlike many laboratory decision-making tasks, requires students to seek out information, develop or discover their own criteria, assess different alternatives—all processes themselves that might require myriad decisions. Students do not have all of the relevant information with which to make the decision placed before them—indeed, they must decide when they have collected sufficient information. Most laboratory tasks present students with information of known (often perfect) reliability or validity (at least by hypothesis). In this real-life decision, the quality information collected (including impressions from visits, opinions of family and friends) must be assessed. These plausible differences between real-life and laboratory decision making raise the question of what yardsticks, if any, can be used to assess the process of making real-life, ill-defined decisions, such as choosing a college.

Payne (1982) has argued convincingly that decisions are contingent on task demands. In other words, the processes by which decisions are made are heavily influenced by factors associated with a particular decision. Indeed, Payne and his collaborators have recently argued that people often *generate* new strategies of decision making while in the midst of a decision-making task, in response, at least in part, to particular aspects of the specific decision (Payne, Bettman, Coupey, and Johnson, 1992; Payne, Bettman, and Johnson, 1993). These authors assert that constructive decision-making processes (i.e., strategies that are created anew in a course of a task) are more likely in task context of little prior knowledge, high stress and high complexity. Relative to typical laboratory gambles featured in decision-making research, the college decision does seem to meet all of these criteria, implying in turn that student decision-makers will create new strategies as they work on the problem, perhaps changing strategies over the course of the process.

In this investigation, different measures were used to describe and assess the processes used by the participants. One set of measures assessed the way students were structuring their decision. By 'structuring', I mean the number of kinds of criteria being generated, the number of alternatives being considered, and the ways in which those two sets of information are combined. In this sense, I am using the term 'structure' in somewhat different ways than others (e.g. Frisch and Clemen, 1994). To avoid confusion, let us call such a structure a 'decision map'.

The decision map that a student creates can vary in complexity in many ways. Two obvious ways have to do with the number of criteria students use, and the number of schools they consider. Obviously, the more of each of these two entities under consideration, the more information must be sought, evaluated, and integrated. Indeed, the first goal of the current study was to assess how students structured this important life decision, and how the structuring changed over time.

A second goal of the study was to examine the ways in which students integrated information during the course of the decision-making process. Specifically, the question at issue was how students put information about their criteria together with their impressions of alternatives with respect to those criteria. To assess this integration, models of information integration were needed.

One possible model was adopted from multi-attribute utility theory (MAUT). A large technical literature describing MAUT and other kinds of utility theories exists (Edwards, 1992a; Keeney and Raiffa, 1976; von Winterfeldt and Edwards, 1986). For our purposes, a brief and simplified explication will suffice. MAUT analyses consist of the following steps: (1) breaking down a decision into independent factors (e.g. cost, size of student body); (2) determining the relative weights of each factor; (3) listing the alternatives (e.g. different possible colleges to attend); (4) rating the alternatives on each factor (e.g. how does Timbuktu University rate on the factor of *cost*); (5) multiplying the ratings by the weightings to determine a final value for each alternative; and (6) choosing the alternative with the highest value.

MAUT, when properly applied and under certain assumptions, has been argued to be a normative model of decision making (Baron, 1988; Keeney, 1992). That is to say, people who follow MAUT maximize their own utility (roughly, personal satisfaction) in a way that is best for achieving all of their goals. If MAUT calculations are employed, the decision-maker can determine what the best options are and how they compare to each other. Decision-makers can also examine the reasons for the final overall rankings of alternatives by examining the weights of different factors and the rankings of different alternatives.

It is debatable whether MAUT is ever used spontaneously by people when making important decisions, especially if the information relevant to making the decision is extensive (Payne, 1976; Slovic *et al.*, 1988). Indeed, people's failure to honor the basic axioms of expected utility theory (the foundation of MAUT as well as other derivations) is well documented (Edwards, 1992b; Keller, 1992). Some have even questioned whether analytical decision strategies are ever heavily used (Mitchell and Beach, 1990). Thus, MAUT may not be a good descriptive model of what people actually do, although the techniques can be taught in an effort to aid people in making decisions (von Winterfeldt and Edwards, 1986). However, some researchers reject the idea of MAUT (or any subjective expected utility model) being considered the normative standard by which real-life decisions are judged (Frisch and Clemen, 1994).

For this reason, we compared the MAUT model to two others. The first one was called the Equally Weighted Criteria model. In this model, a subject's ratings of each alternative on each criterion are given equal weight and averaged. That is, the subjective importance weights students give for each criterion are simply ignored. Some evidence suggests that people are notoriously bad at providing their own meaningful weights (Reilly and Doherty, 1989), providing a good rationale for using this model. If the Equally Weighted Criteria model does as well as the MAUT model in fitting the data, it would suggest that student's importance weights play little role in their decision making.

A third and very simplified model was called the Top Criteria model. In this, the criterion to which the student gave the highest importance weighting was the only one used. If (as frequently happened), students gave two or more criteria the highest importance weightings, then the ratings of schools on those criteria were averaged.

Each model was used to compute, for each student, an 'expected value' for each school under consideration. Those values were then correlated with the student's 'overall impression', an intuitive, 'gut-level' comparative rating of each school under active consideration. Notice that all of these models fall into the category of linear models, argued to have been successfully used in a wide variety of decisions (Dawes and Corrigan, 1974; Dawes, 1982).

Individual differences in decision making were also assessed. In particular, gender and academic ability differences were examined. Because some previous work has suggested the existence of at least some gender differences in adolescents' thinking about commitments (Galotti, Kozberg, and Appleman, 1990), gender was included as a relevant independent variable. Academic ability seemed particularly relevant to this decision, as arguably those students with higher ability would have a greater range of options available from which to choose. Moreover, it might be the case that students of higher academic ability are more invested in the college decision, and expend more effort in making it. They might, for example, seek more information, evaluate the information they receive more critically, or integrate it more thoroughly. Another possibility is that students with higher academic ability are more able to deal with the complexity inherent in this real-life decision.

METHOD

Study design

College-bound high school students described their college decision making at one or more sessions during their junior and senior years of high school. This investigation employed a sequential longitudinal design, with up to three sessions per student. The first round of sessions took place in April of 1991, with a sample of high school juniors. Many of these students were reinterviewed in two subsequent sessions in their senior year of high school, in October of 1991 and April of 1992.

At each subsequent session, a new sample of students was also recruited, to assess any impact on decision making of participation in the study. The reason for this was a concern that participation in a previous session, during which participants were asked to generate and weight the importance of criteria, and then to generate and rate the goodness of alternatives with respect to each criterion, might lead subjects to adopt a more analytic strategy of decision making than they otherwise might during subsequent sessions. Moreover, because students were given a list of standard criteria at each session, there was reason to worry that subsequent participation might be affected.

At each session, students described their current thinking and experiences in choosing a college, both through open-ended questions and structured items. Specifically, students described the factors (e.g. cost, location) they were considering and their relative weights, the alternatives (e.g. colleges, universities, trade schools) they were considering, and how they rated each alternative on each factor. Next, they rated each alternative against a list of 34 standard factors, to which they also assigned importance weights. Students next ranked all alternatives under consideration for overall goodness of fit. Finally, they rated their overall level of satisfaction and comfort with (along with several other affective responses) the decision-making process.

Five major questions are addressed in this report. These are: (1) How do students structure this decision, in terms of the factors and alternatives considered at various points in the process? (2) How do students' decision maps change over time? (3) Does providing people with a set of criteria to consider affect their subsequent decision maps? (4) How do students integrate information and judgements about factors and alternatives? More specifically, to what degree can students' overall rankings of alternatives be predicted by formal linear models? (5) Does complexity of decision map, or correlation of decision maps with formal linear models, predict either greater satisfaction with the decision-making process, or greater success at the end of the process?

Cutting across all of the above questions is the issue of individual differences. Specifically, to what extent do students of different academic ability levels, who might be thought to have differential interest and/or investment in the college decision construct different decision maps? Do male and female students differ in either the structure or content of their decision maps? The data were also examined for practice effects, i.e. systematic differences in responses as a function of the number of times students participated in a survey session.

Subjects

A total of 322 high school students (88 males, 234 females)[1] from 19 high schools in southeastern Minnesota (18 public, 1 private) participated in one or more sessions. The breakdown of participation is as follows: One hundred and twenty four students (29 males, 95 females) were originally recruited in spring of their junior year of high school, and participated in the first round of sessions in April of 1991. Of these, 101 (26 males, 75 females) participated again in the second round held in October of the senior year (1991), when an additional sample of 99 students (22 males, 77 females) joined the study. Ninety of the original sample who had participated in Round 2 (24 males, 66 females), and 75 of the second sample (17 males, 58 females) again participated in the final round of sessions, held in April of the senior year of high school (1992), and were joined by a third sample of 99 subjects (37 males, 62 females) who participated in the last round only. Students who joined the study in the latter rounds were often, but not always, from the same high schools as those students who participated in all three rounds.

Students were recruited through high school homeroom announcements. A letter described the study, and asked interested participants to provide us with their name and phone number. It also asked for written parental consent for students under the age of 18, and asked both the student and the parent to sign a release authorizing examination of schools records regarding transcripts, class rank, and standardized test scores (PSAT, SAT, and ACT). Students received \$5.00 for participating in each session, with a \$5.00/session bonus for participating in a second or third session.

All students who had ever participated in any session were sent a follow-up questionnaire in January of 1993, during their first year of college. Students who returned the questionnaire were sent \$5.00. Responses came from the following participants: Of the original sample of 124, 82 (70 females, 12 males) responded; of the second sample of 99, 62 (50 females, 12 males) responded, and of the final sample of 99, 63 (42 females, 21 males) responded. Of the total 207 students who responded to the questionnaire, 58 (48 female, 10 male) had also participated in all three

[1]Attempts to recruit more equivalent numbers of males and females was not successful; females in all high schools were disproportionately likely to agree to participate.

previous sessions. Three of the 207 respondents (all female; two from the original sample and one from the second sample) were not in a college or university in January of 1993.

Instruments and procedure

Students participated in 1-hour sessions scheduled at their high school, either in the evening or after school. Sessions included up to 12 students, but averaged approximately five. At least two research assistants were present at each session to hand out forms, answer questions, scan completed forms to detect errors or omissions, and to pay participants at the conclusion of the session. Participants completed six, seven, or eight (depending on the session) different colour-coded forms at their own pace. Some students took as little as 35 minutes, others took slightly over an hour, but most took about 55 minutes for their initial session, and about 45 minutes for subsequent sessions. Throughout the session, research assistants were available to answer questions about the meanings of instructions or other issues of clarification, but otherwise kept interaction with participants (or among participants) to a minimum. Students were not allowed to go back to change their responses to a previous instrument once it had been handed in.

In the order of administration, the forms relevant for present purposes[2] were:

Background information sheet
This was completed at the first session the student participated in. It asked for name, address, date of birth, year in school, and other personal information.

Self-generated factors weighting and ranking sheet
This instrument had several columns. In the first column, students were asked to list the factors (i.e., criteria) they were using in making decisions about college (e.g., cost, location, programme offerings). In the second column, students assigned each factor an importance weighting, on an integer scale from 0 to 10. The next seven columns provided spaces at the top for students to list schools under consideration. In the blanks underneath each listing, students rated *that* school on the corresponding factor listed in the first column, again using a 0 to 10-point integer scale. A second blank sheet was provided to encourage students to be complete; other blank sheets were also offered if needed (but were used rarely).

Standard factors weighting and ranking sheet
This instrument was identical to the one described above, except that the first column was filled in with a standard set of 34 factors, including such things as 'The admissions process', 'Financial aid availability', 'Computer facilities', 'Locale with respect to major city', 'Your parents' advice or preference'. The list of standardized factors was adapted from Brodigan (1984).

[2]Other forms asked students to describe the process in their own words, and asked them to rate the frequency with which they consulted various potential sources of information or engaged in different decision-making activities. A copy of all forms used is available upon request.

Overall impressions of schools sheet
Students listed each school under consideration and rated their overall feeling of how good a choice each school currently appeared to be for them, on an 11-point scale ('10' = best choice, '0' = worst choice).

Feelings about the decision-making process sheet
Using 7-point Likert scales ('1' = not at all; '7' = completely), students rated their current feelings about the process. Examples of items include: 'How stressful is it to make this decision?', and 'How comfortable are you with the way you are making this decision?'.

Applications results sheet
This was administered to all students participating in Round 3—April 1992—only. Students listed the colleges they had applied to, the colleges at which they were accepted, not accepted, waitlisted, and those from which they were waiting to hear.

Follow-up questionnaire: retrospective feelings about the decision-making process sheet
Using 7-point Likert scales ('1' = not at all; '7' = completely), students rated their feelings about the process. Examples of items include: 'How certain are you that your college decision was the right one?', and 'How satisfied do you feel with the way your decision has worked out?'.

RESULTS AND DISCUSSION

Demographic data for the entire sample of student participants, as well as those participants who completed three sessions of participation (henceforth referred to as the 'core' sample) are presented in Table 1. Very few differences were found in the three cohorts of students drawn, so unless otherwise noted, the results will be reported as averaged over all cohorts.

As might be expected for a college-bound sample, the level of academic ability was high: Mean scores for SATV + SATM and ACT composite were 1117.26 and 24.05, respectively, and the class rank (computed in percentiles at the end of junior year, based on GPA), was 73.55. Table 1 indicates that these measures took on similar mean values for the core sample.

Many of the analyses to be reported below used an independent variable called 'academic ability'. This was measured by equally weighted z-scores of eleventh grade GPA, eleventh grade class rank, and standardized test scores (ACT composite scores and SATV + SATM scores were converted to a common scale following a conversion table developed by Aslin (1971); students who took both tests had the higher score used). Ability scores were correlated with all of the academic measures, as expected. Moreover, ability was correlated with level of parental education ($r[316] = 0.29$ $p < 0.01$) and negatively with the number of siblings in the family ($r[316] = -0.19$, $p < 0.01$). Students were assigned to an ability group based on whether their ability score fell in the top, middle, or bottom third of the sample.[3] For descriptive

[3]The resulting data were analysed using mixed-design ANOVAs and MANOVAs. The disadvantage of this method of analysis stems from its lack of statistical power, relative to multiple regression analysis. The data were actually analysed both ways, with very similar results. For ease of exposition, the ANOVA results are presented here.

Table 1. Demographic characteristics of the sample

Measure	All participants (%) ($n=322$)	Core sample (%) ($n=90$)
Sex		
Male	27.6	27.7
Female	72.4	72.2
Race/ethnicity		
White/Caucasian	93.2	93.3
African–American	1.0	0.0
Latino/Latina	0.6	1.1
Asian/Asian–American	4.5	4.5
Other	0.6	1.1
Family income		
< 10K	3.5	1.3
10K–39K	22.6	22.1
40K–79K	43.9	41.6
80K–119K	19.9	20.8
120K–149K	6.3	9.1
> 150K	3.8	5.2
Averaged level of parental education		
High school or less	28.0	38.8
Some college or college graduate	29.2	23.3
Some postgraduate or more	42.8	37.8
ACT Composite Score	24.05 (sd = 4.28)	23.83 (sd = 4.38)
SAT Verbal + Math	1117.26 (sd = 164.74)	1141.91 (sd = 165.13)
Class Rank (percentile) (computed at end of junior year)	73.55 (sd = 22.77)	73.96 (sd = 20.79)

purposes, I refer to these groups as the 'higher', 'average', and 'lower' academic ability groups, although it must be kept in mind that all three groups display a good deal of academic ability. Table 2 shows the means of all of these measures for the three ability groups. Only one gender difference was found in the demographic variables: males outperformed females on the SATV + M measure ($M = 1184.61$ versus 1081.76; $t[111] = -0.33$, $p < 0.001$).

To facilitate exposition, the remaining presentation of results will be organized according to the questions presented earlier:

How do students structure this decision (i.e., what are their decision maps)?

At each session, students listed criteria they were using to make their college decision, and also the alternatives (schools) they were currently considering. From these lists, three distinct measures that reflect the structure of decision making were adopted from previous work (Galotti and Kozberg, 1987; Galotti *et al.*, 1991), based on measures used in the divergent thinking literature (Torrance, 1972; Wallach and Kogan, 1965). The first is *number of criteria*, and relates to a measure Torrance and others call 'fluency'. The second measure is derived from the first, by classifying all

Table 2. Mean ability measures by academic ability group

Group	GPA[a]	Class Rank[a]	ACT Composite[b]	SAT V+M[b]
Lower ability ($n = 105$)				
M	2.60	47.87	15.75	955.00
SD	0.42	17.13	8.53	149.27
Average ability ($n = 108$)				
M	3.34	78.13	22.54	1047.50
SD	0.25	9.72	5.68	168.60
Higher ability ($n = 105$)				
M	3.85	94.69	24.90	1187.97
SD	0.15	3.75	9.35	124.98

[a]Computed at the end of junior year. GPA is reported on a 4-point scale; Class Rank is percentile rank.
[b]The SAT was taken by only 113 of the sample, compared with 281 for the ACT.

criteria listed by all subjects into a taxonomy developed by the author and research assistants. Called *flexibility of criteria*, this measure is the number of different categories the criteria are coded into, and reflects the number of distinct *types* of factors considered (for a fuller description of types used, see Galotti and Mark, 1994). The third measure used, *number of alternatives* was the number of schools under active consideration.

The first analyses used all three of the aforementioned measures collected from the third (April 1992) session. A major purpose of this analysis, which included the broadest coherent subset of the sample, was to assess the effects of repeated testing, as students from the first cohort would have participated in this task three times, students in the second cohort, twice, while students in the third cohort would be seeing the instruments for the first time. A 3 (cohort) × 2 (gender) by 3 (ability group) × 3 (measure: number of criteria, flexibility of criteria, number of alternatives) ANOVA, with repeated measures on the last factor, revealed no overall main effects for cohort, gender, or ability, nor were any interactions among these variables statistically significant. There was a main effect for measure ($F[2, 488] = 104.07$, $MS\ error = 9.47$, $p < 0.001$), with the overall means being 8.85, 7.71, and 4.24, for number of criteria, flexibility of criteria, and number of alternatives, respectively. *Post-hoc* Tukey HSD tests revealed that the all means differed reliably from one another ($p < 0.01$).

Having established the lack of a repeated testing effect on the measure, I next examined the data of the 90 'core' participants, who had participated in all three sessions. A 2 (gender) by 3 (ability group) × 3 (measure: number of criteria, flexibility of criteria, number of alternatives) × 3 (session) ANOVA, with repeated measures on the last two factors, once again did not show any main effects for gender or ability, although this time a significant interaction between the two emerged ($F[2, 84] = 3.12$, $MS\ error = 30.27$, $p < 0.05$). Overall means for females in the lower, average, and higher ability groups, respectively, were 6.14, 7.91, and 7.76; while the corresponding means for males were 7.97, 6.46, and 8.59. *Post-hoc* Tukey HSD tests did not indicate any reliable difference between any pair of means. However, the ANOVA indicated that the overall effect for ability group was marginally significant ($F[2, 84] = 2.69$, $MS\ error = 30.27$, $p < 0.08$).

Again, there was a main effect of measure, with the overall means being 9.36, 7.94, and 4.21, for number of criteria, flexibility of criteria, and number of alternatives, respectively ($F[2, 169] = 89.04$, $MS\ error = 13.98$, $p < 0.001$). *Post-hoc* Tukey tests revealed that all means differed reliably from one another ($p < 0.01$). No other effects or interactions reached significance.

What kinds of criteria did students generate, and did this vary by time? We categorized the criteria (factors) students listed on the self-generated factors weighting and ranking sheet into 23 categories. Table 3 presents the mean number of factors generated per category, by time of interview, for the 90 core subjects.[4] The number of factors generated were subjected to a 2 (gender) × 3 (ability group) × 23 (category) × 3 (session) mixed ANOVA, with repeated measures on the last two factors. The analysis indicated a main effect for category ($F[22, 1848] = 18.16$, $MS\ error = 0.51$, $p < 0.001$), and an interaction between category and session ($F[44, 3696] = 1.44$, $MS\ error = 0.23$, $p < 0.05$). Table 3 presents the means for this interaction. *Post-hoc* Tukey HSD tests (run at the $p < 0.01$ level, because of the number of comparisons involved), indicated a significant difference in generation of factors as a function of time for the following categories: an increase for 'Majors/programmes offered', 'Campus atmosphere', 'Dorms/residence halls', 'Extracurricular programmes', and 'Location'; a decrease for 'Admissions requirements', 'Course offerings/curriculum'; and a non-linear change in the categories 'Class size/student-faculty ratio', 'Size' and 'Financial aid'.

This analysis also revealed statistically significant interactions of category with both gender ($F[22, 1848] = 3.28$, $MS\ error = 0.51$, $p < 0.001$) and ability group ($F[44, 1848] = 1.71$, $MS\ error = 0.51$, $p < 0.01$). Means for both of these interactions are presented in Table 4. *Post-hoc* Tukey HSD tests (run at the $p < 0.01$ level, because of the number of comparisons involved), indicated a significant gender difference in generation of factors for the following categories: females higher for 'Type of school', 'Campus appearance', 'Dorms/residence halls', and 'Size'; males higher for 'Academic challenge', 'Admissions requirements', 'Reputation/accreditation', 'Success of graduates' and 'Facilities'. Similar *post-hoc* Tukey HSD tests indicated a number of reliable ability group differences: those that increased with academic ability level: 'Majors/programmes offered', 'Type of school', 'Size', 'Financial aid', and 'Distance from home'; those that decreased with academic ability level: 'Reputation/accreditation'; and others: 'Admissions requirements', 'School policies/regulations', 'Campus appearance', 'Class size', 'Dorms/residence halls', and 'Facilities'. No other significant effects or interactions emerged.

These analyses suggest that college-bound high school students report considering about nine criteria, about eight distinct *types* of criteria, and approximately four different schools as they face this decision. Interestingly, these figures do not change over the course of the year, as students move closer to the decision. There is a trend (that reaches only marginal statistical significance) for higher and average ability students to report more criteria, more distinct types of criteria, and more schools

[4]Another analysis examined the data from all subjects from the April 1992 sessions, and found results similar to those described below. However, there were no significant effects or interactions with cohort, suggesting that repeated participation in our study did not affect the content of decision maps.

Table 3. Factors students report considering by time of interview

Factor	April 1991	October 1991	April 1992
Academic factors			
Academic challenge	0.31	0.23	0.26
Admissions requirements	0.39[a]	0.34[a]	0.20[b]
Course offerings/curriculum	0.32[a]	0.21[b]	0.22[b]
Faculty quality/credentials	0.16	0.21	0.11
Majors/programmes offered	0.93[a]	1.03[b]	1.04[b]
Reputation/accreditation	0.40	0.42	0.38
School policies/regulations	0.17	0.14	0.10
Success of graduates	0.34	0.33	0.27
Types of school (e.g. public/private)	0.54	0.62	0.46
Institutional factors			
Campus appearance	0.24	0.21	0.29
Campus atmosphere	0.26[a]	0.51[b]	0.57[b]
Class size/student/faculty ratio	0.36[a]	0.26[b]	0.42[a]
Dorms/residence halls	0.23[a]	0.33[b]	0.41[b]
Extracurricular programmes	0.58[a]	0.69[b]	0.63[b]
Facilities	0.24	0.30	0.37
Location	0.56[a]	0.62[a,b]	0.69[b]
Physical setting (e.g. rural/urban)	0.36	0.37	0.33
Size	0.64[a]	0.76[b]	0.67[a,b]
Financial factors			
Cost	0.80	0.87	0.82
Financial aid	0.41[a]	0.70[b]	0.57[c]
Personal/social factors			
Distance from home	0.38	0.33	0.33
Parents'/friends' advice	0.12	0.11	0.11
Peers/friends at school	0.23[a]	0.12[b]	0.17[a,b]

Numbers indicate the mean number of factors generated that fell into a category from the cohort of students who participated in all three interviews. Numbers within a row with different superscripts differ at the 0.01 level, by a *post-hoc* Tukey HSD test.

than lower ability students. Males and females do not differ in the complexity of their self-reported decision maps.

However, males and females *did* differ in the kinds of criteria they generated, as did students of different academic ability. This suggests that the content, if not the structure, of decision maps *does* vary by individuals belonging to different demographic groups. Moreover, the emphases on different kinds of criteria for the group of students as a whole also changed over time.

How did the reported decision maps change during the course of the decision-making process?

How consistent were the criteria or the schools under consideration over time for those students who participated in more than one round of interviews? One measure of consistency is the percentage of overlap in the criteria listed from one round to the next. This measure is calculated by dividing the number of criteria listed on two

Table 4. Criteria students report considering by gender, and by ability group

Factor	Sex		Ability Group		
	Female	Male	Higher	Average	Lower
Academic factors					
Academic challenge	0.24[a]	0.35[b]	0.28	0.26	0.27
Admissions requirements	0.26[a]	0.43[b]	0.38[a]	0.24[b]	0.32[a,b]
Course offerings	0.28	0.19	0.24	0.28	0.24
Faculty quality/credentials	0.17	0.12	0.15	0.18	0.15
Majors/programmes offered	1.02	0.96	1.38[a]	0.97[b]	0.79[c]
Reputation/accreditation	0.33[a]	0.59[b]	0.30[a]	0.41[a]	0.56[b]
School policies/regulations	0.12	0.17	0.13[a,b]	0.21[a]	0.06[b]
Success of graduates	0.21[a]	0.45[b]	0.25	0.27	0.31
Types of school (e.g. public private)	0.70[a]	0.14[b]	0.82[a]	0.53[b]	0.37[b]
Institutional factors					
Campus appearance	0.28[a]	0.13[b]	0.24	0.32[a]	0.18[b]
Campus atmosphere	0.45	0.43	0.51	0.47	0.38
Class size/student/faculty ratio	0.35	0.32	0.32	0.43[a]	0.29[b]
Dorms/residence halls	0.36[a]	0.24[b]	0.25[a]	0.47[b]	0.26[a]
Extracurricular programmes	0.61	0.71	0.64	0.61	0.65
Facilities	0.27[a]	0.39[b]	0.32	0.40[a]	0.21[b]
Location	0.63	0.63	0.58	0.61	0.67
Physical setting (e.g. rural/ urban)	0.38	0.37	0.42	0.42	0.31
Size	0.74[a]	0.55[b]	0.86[a]	0.72[b]	0.55[c]
Financial factors					
Cost	0.84	0.81	0.83	0.83	0.82
Financial aid	0.57	0.53	0.82[a]	0.46[b]	0.47[b]
Personal/social factors					
Distance from home	0.36	0.34	0.42[a]	0.41[a]	0.26[b]
Parents'/friends' advice	0.09	0.17	0.11	0.09	0.14
Peers/friends at school	0.16	0.21	0.10	0.24	0.17

Numbers indicate the mean number of factors generated that fell into a category from the cohort of students who participated in all three interviews. Numbers within a section within a row with different subscripts differ at the 0.01 level, by a *post-hoc* Tukey HSD test.

rounds by the total number of criteria listed on the first round. A similar measure of consistency can be calculated comparing the percentage of overlap of schools listed from one session to the next.

Because these data require participation in multiple sessions, only the data from the 'core' sample were analysed. A 2 (gender) × 3 (ability group) × 2 (measure: percentage overlap of criteria, percentage overlap of schools) × 3 (interval: April 1991–October 1991; April 1991–April 1992; October 1991–1992) mixed ANOVA, with repeated measures on the last two factors, was performed.

The analysis showed no main effect or interaction with gender or ability group, nor did either of these factors interact with the other factors. There was a main effect of interval ($F[2, 166]=16.76$, $MS\ error=430.52$, $p<0.001$). However, this effect must be interpreted in light of a significant interaction between measure and interval

(F[2, 166] = 7.60, *MS error* = 336.26, $p < 0.001$). Overall, the mean percentage overlap of criteria was 48.82 between April 1991 and October 1991, 52.06 between October 1991 and April 1992, and 47.96 between April 1991 and April 1992. Overall, the mean percentage overlap of schools was 47.64 between April 1991 and October 1991, 69.05 between October 1991 and April 1992, and 42.75 between April 1991 and April 1992. *Post-hoc* Tukey HSD tests indicated a significant ($p < 0.01$) difference in means for the overlap in criteria vs. schools for the interval October 1991–April 1992 only.

Which criteria remained on subjects' lists from one session to the next? Predictably, those criteria to which subjects had assigned higher importance weightings were more likely to be re-generated a subsequent session. Mean importance weightings for re-generated versus non-regenerated criteria were computed, and analysed by a 2 (gender) × 3 (ability group) × 2 (type: regenerated versus not regenerated) × 3 (interval: April 1991–October 1991, October 1991–April 1992, April 1991–April 1992) mixed ANOVA with repeated measures on the last two factors. This analysis revealed a significant effect of type, with the mean importance rating of regenerated criteria ($M = 7.29$) significantly higher than the mean importance rating of non-regenerated criteria ($M = 6.94$; F[1, 83] = 8.60, *MS error* = 3.68, $p < 0.01$). No other significant effects or interactions emerged.

What were the effects (if any) of providing students with a standard list of criteria?

Recall that immediately after completing the 'self-generated factors weighting and ranking sheet', participants were given a similar instrument, one that contained a list of 34 criteria (the 'standard factors weighting and ranking sheet'). It might be argued that the provision of these criteria might affect participants' decision making, or at least their performance in subsequent sessions. It might first be noted that the number of criteria on the standard sheet (34) greatly exceeded the number of self-generated criteria (around nine) for every subject, a finding quite reminiscent of those reported by Beyth-Marom *et al.* (1993), looking at the number of consequences adolescents and adults listed for risky behaviours (around seven) relative to a list of consequences described in other literature (over 50).

The first relevant analyses examined the number of self-generated factors that overlapped with the list of standard factors. The data from the April 1992 session was examined first, to assess the effects of repeated testing. Recall that students in the first cohort would have been seeing both ranking and weighting sheets for the third time; those in the second cohort, twice, and those in the third cohort, once. A 3 (cohort) × 2 (gender) × 3 (ability group) ANOVA performed on the percentage of overlapping criteria did yield a main effect of cohort (F[2, 244] = 48.35, *MS error* = 7.64, $p < 0.01$). The mean number of self-generated overlapping criteria was 8.51, 7.34, and 7.20 for students in the first, second, and third cohorts, respectively. *Post-hoc* Tukey HSD tests indicated that the first mean differed reliably from the other two, which did not differ ($p < 0.05$). No other effects or interactions reached significance.

Recognizing the effects of practice, I went on to examine the data from the 90 core subjects, analysing the number of overlapping criteria in a 2 (gender) × 3 (ability group) × 3 session mixed ANOVA, with repeated measures on the last factor. This analysis indicated an overall main effect of ability group (F[2, 84] = 3.92, *MS error* = 16.91, $p < 0.05$). The mean number of overlapping criteria was 6.88, 8.23, and

8.94 for students in the lower, average, and higher ability groups. *Post-hoc* Tukey HSD tests indicated that the first and third means differed reliably from each other, but not from the second ($p < 0.01$).

Not surprisingly, given the previous analysis, there was also a significant main effect for sessions. The mean number of overlapping criteria was 7.17 in the April 1991 session, 7.99 in the October 1991 session, and 8.51 in the April 1992 ($F[2, 168] = 3.60$, *MS error* = 3.61, $p < 0.05$). *Post-hoc* Tukey HSD tests indicated that all means differed reliably from the others ($p < 0.01$).

Recall that participants gave importance weights to the criteria they generated, and to the standard list of factors on a scale of 0–10. Mean importance weights of overlapping versus non-overlapping criteria (the overlap here is with the standard list) were subjected to a 2 (sex) × 3 (ability group) × 3 (session) × 2 (overlap status) mixed ANOVA with repeated measures on the last two factors for data from the 90 core subjects. The analysis revealed a main effect for overlap status, with all participants giving higher importance weights ($M = 7.66$) to overlapping criteria than to criteria they generated which did not overlap with the standard list of criteria ($M = 6.47$; $F[1, 73] = 100.53$, *MS error* = 1.54, $p < 0.001$). Males ($M = 6.47$) gave significantly lower importance ratings to all factors than did females ($M = 7.18$; $F[1,73] = 11.51$, *MS error* = 4.09, $p < 0.001$). Finally, there was a main effect for session, with overall importance weights declining over sessions, with mean importance weightings 7.24, 6.91, and 6.81 for the April 1991, October 1991, and April 1992 sessions, respectively ($F[2,146] = 3.66$, *MS error* = 0.84, $p < 0.05$). *Post-hoc* Tukey HSD tests showed that the second and third means differed significantly from the first ($p < 0.01$).

Taken together, the results of these analyses suggest that repeated exposure to the standard list of criteria *did* affect students' future generation of criteria, although the effects, in terms of number of additional overlapping criteria generated, were small. Higher ability students generated lists of criteria that overlapped more with the standard list of criteria than did lower ability students, and this effect was present even in the first session.

To what degree can students' rankings of potential schools be predicted by formal linear models?

Students making the college decision have a great deal of information to integrate. The information focused upon in this study included the student's criteria, the relative importance of the criteria, the alternatives (schools) under consideration, and the student's perception of how well each alternative was assessed with respect to each criterion.

Even given the restriction of focus in this study to this limited set of information, we may still well ask how students are able to integrate all of it. To what extent are students' overall intuitive impressions of schools a reflection of this information integration? One class of ways of integrating information, well studied in the decision-making literature, are formal linear models (Dawes and Corrigan, 1974). Three different formal linear models of integrating the information described above were considered. Each model provided an 'expected value' for each school—a measure of overall goodness a given school 'ought' to have, given a particular way of combining that school's ratings on the various criteria given. These 'expected values' were then correlated with the student's 'overall impressions' of the set of schools under consideration.

The three models of information integration were the following:

Full multi-attribute utility theory model (FM)

This model incorporated all of the information a subject provided on the last three instruments described in the method section. The 'expected value' for a school was computed by multiplying the importance weighting of each factor listed by the subjective ranking of *that* particular school on that particular factor.[5] Thus, each school listed received one summary score. Next, these scores were correlated with the ranks of students' overall impression rating of each school given on the *overall impressions of schools sheet*. (Recall that students could not look back on previously completed instruments while giving these overall impressions.) Positive correlations indicate 'better' integration of information, at least by MAUT criteria.

A similar kind of analysis was carried out using responses to the *standard factors weighting and ranking sheet*. The rationale for having both weighting and ranking sheets was as follows: the self-generated factors sheet (always administered first) allowed subjects to list their own criteria. The standard factors sheet provided all subjects with a larger set of criteria that included criteria they might not have spontaneously thought of earlier. Thus, Full MAUT expected values could be calculated both ways.

The Full MAUT model makes the questionable assumption that students' assessments of weights and ratings of alternatives make use of true ratio scales. That is, it assumes that a student who gives the criterion of 'cost' an importance weighting of '4' and 'location' a '1', treats the first criterion as exactly four times as important as the second. To assess this assumption, we compared the performance of the MAUT model with two other, more simple linear models of information integration. Both of these models, like MAUT, assumed an analytical and specifically linear approach to the decision. The other two models relaxed one or more of the MAUT assumptions.

Equally weighted criteria model (EW)

With this model, the 'expected value' for each school was computed by taking the average of the ratings of each school on each factor, ignoring the subjective importance weightings. Again, expected values were correlated with the ranks of students' overall impression rating of each school given on the *overall impressions of schools sheet*, both for the *self-generated factors weighting and ranking sheet*, and the *standard factors weighting and ranking sheet*.

Top criteria model (TC)

A yet simpler way of calculating the expected value of a school is to include only the student's most important criterion(a). Therefore, expected values under this model were calculated by using the ratings *only* on the criterion to which a student had given the highest importance weighting. If a student had given more than one criterion the highest weight, then the average ratings on all these criteria were

[5] To meet the requirements of a MAUT analysis, we checked to be sure that all the factors listed were 'psychologically independent' (Baron, 1988: p. 312).

computed to calculate the expected value of that school. Again, expected values were correlated with the ranks of students' overall impression rating of each school given on the *overall impressions of schools sheet*, both for the *self-generated factors weighting and ranking sheet*, and the *standard factors weighting and ranking sheet*.

Correlations[6] between expected values and subjects' overall impressions were calculated from the April 1992 sessions, to ensure the broadest coherent subset of subjects. These correlations (computed within each subject) were analysed with a 3 (ability group × 2 (sex) × 3 (cohort) × 3 (model: Full MAUT, Equally Weighted Criteria, Top Criteria) × 2 (type of correlation–using Self-Generated Factors, using Standard Factors) mixed ANOVA, with repeated measures on the last two variables.[7]. The analysis revealed, first, an overall main effect for model, with mean correlations with students' overall impressions of schools being 0.69, 0.67, and 0.53, when expected values of schools using the Full MAUT, Equally Weighted Criteria, and Top Criteria models were used ($F[2, 324] = 30.05$, *MS error* $= 0.10$, $p < 0.001$). Tukey HSD *post-hoc* tests indicated that the first two means differed significantly from the third ($p < 0.01$).

The ANOVA also yielded a number of interactions. One of the most interpretable was between model and ability group ($F[4, 324] = 6.71$, *MS error* $= 0.10$, $p < 0.01$). The means for this interaction are presented in Table 5. *Post-hoc* Tukey (HSD) tests showed that for the higher and average ability groups, the correlations with the Full MAUT and Equally Weighted criteria models were significantly higher than were the correlations computed with the Top Criteria model ($p < 0.01$). For the lower ability students, all three models yielded similar correlations.

Some interactions with cohort as a factor were also statistically significant, although no clear interpretable pattern emerged. First was an interaction between cohort and model ($F[2, 324] = 2.59$, *MS error* $= 0.10$, $p < 0.05$). *Post-hoc* Tukey HSD tests indicated a statistically significant difference only between the first and second cohorts in correlations with the equally weighted model ($M = 0.74$ versus 0.60; $p < 0.01$). Moreover, this two-way interaction occurs in light of a significant three-way interaction, among cohort, ability group, and model ($F[8, 324] = 2.85$, *MS error* $= 0.10$, $p < 0.01$), interpretation of which proved elusive.

A number of other interactions, most of them of higher order, were also found. In many of these cases, however, the cell sizes were very unbalanced, with some cell sizes containing fewer than 10, rendering interpretation all but impossible. Among these were: cohort × sex × model ($F[4, 324] = 2.94$, *MS error* $= 0.10$, $p < 0.05$), cohort × ability group × sex × model ($F[8, 324] = 2.34$, *MS error* $= 0.10$, $p < 0.05$), and cohort × ability group × sex × model × measure ($F[8, 324] = 2.01$, *MS error* $= 0.06$, $p < 0.05$).

Perhaps surprisingly, there was no significant main effect for measure, meaning that correlations computed with the self-generated list of factors were as strong as

[6]The correlations reported are Pearson correlation coefficients performed on values participants actually gave. Spearman correlation coefficients, performed on ranks, yielded very similar results. Correlations were computed within subjects, and subjects with only one or two schools listed were therefore necessarily excluded from this analysis, leaving 42 subjects in these analyses.

[7]Another analysis for data from the core subjects included time of interview as a factor, and showed a main effect for this factor, with correlations becoming higher over the three survey periods. However, because of missing data, the results were based on the data of only 25 subjects. For this reason, I report only analyses of the data of all those subjects who participated in the last, April 1992 survey, where data were available for more participants.

the correlations computed with the standard list of factors. There were only two interactions with measure as a factor. The first was the uninterpretable five-way interaction described above. The second was a three-way interaction with ability group, model, and measure ($F[8, 324]=3.14$, *MS error*$=0.06$, $p<0.05$). *Post-hoc* Tukey HSD tests indicated that correlations differed between self-generated and standard list of factors only under the 'top criterion' model, and only for the highest ability group ($p<0.05$).

The results of this analysis suggest several things about the ways in which students' integration of information can be modeled. First, linear models generally perform well in predicting students' overall intuitive rankings of schools, although none of the models tested was even close to perfect. However, there was no perceptible difference between the correlations generated with the full MAUT model (including subjects' importance weights of the criteria), and a model in which all listed criteria were given equal weights. Moreover, no strong differences were found in the correlations as a function of whether or not the students' own list of criteria were used or the more inclusive list of standard factors was used.

Interestingly, linear models were especially good at predicting the correlations for higher and average ability students. Relative to the single ('top') criterion model, correlations generated from the full MAUT and the equally weighted criteria model were significantly higher, but only for the two 'higher ability' groups of students. This may suggest that those students do a better job of integrating information. Alternatively, the results may mean that lower ability students integrate information non-linearly, although in other equally complex ways yet to be specified.

Does complexity of decision map, or correlation of decision maps with formal linear models, predict either greater satisfaction with the decision-making process, or greater success at the end of the process?

Recall that at each survey, and again in the follow-up questionnaire, participants were asked to respond to several rating scales, reporting (among other things) their certainty in the decision, their level of comfort with the process, and their own retrospective satisfaction with the process. In the April 1992 session, participants also provided a list of schools they were accepted at, waitlisted at, not accepted to, or waiting to hear from.

Table 5. Mean MAUT correlations between overall impressions and expected values, by academic ability and model of information integration[1]

	Model of information integration		
Academic ability group	Full MAUT	Equally weighted	Top criteria
Higher ability	0.71	0.73	0.49
Average ability	0.73	0.71	0.54
Lower ability	0.59	0.59	0.56

[1]Correlations were computed within subjects, i.e. a different correlation was computed for each subject. Subjects with two or fewer alternatives listed were thus excluded from the analysis, leaving data of 42 subjects for the final ANOVA.

amount of information to be gathered, evaluated, and integrated at any given time. Payne's (1976) laboratory studies of people selecting hypothetical apartments showed that people's decision-making strategies were less compensatory and less complex when the number of alternatives is large; the data reported here seem to provide a real-life replication.

Interestingly, the overall structure of decision maps, in terms of the *number* of criteria or alternatives considered, does not change over the course of the decision-making process. What does change is the *content* of the decision map: the specific criteria considered. Content changes in decision maps as a function of both gender and academic ability were quite apparent as well.

Women, more than men, listed criteria pertaining to 'climate'—residential facilities, campus appearance, size and type of school—while males were more concerned with factors such as academic challenge, success of graduates, or reputation/accreditation. One speculation is that differential socialization causes women to care more about their surroundings and the people in them, and men to focus more upon factors that might enhance future individual success.

Another aspect of the decision was that, over the two 6-month intervals studied (April 1991–October 1991; October 1991–April 1992), students are only likely to re-list about half of the criteria on the second occasion as they do on the first. There was somewhat greater consistency for number of alternatives re-listed in the second interval (October–April), although it ought to be kept in mind that the number of criteria listed was always about double the number of alternatives listed.

Examination of models of information integration suggest that students do in fact consider multiple criteria in forming their overall impressions of a particular alternative. However, the relatively equal fit of the Full MAUT and the Equally Weighted Criteria models suggest in contrast that students did not make use of the importance weights they gave to different criteria. There were noticeable differences in this aspect of performance as a function of academic ability. In particular, those models involving multiple criteria fit better the data of more academically able students relative to the data of lower ability students. This suggests that linear models of information integration, widely cited in the decision-making literature, differentially fit the data of particular students.

There are at least possible explanations of the differences in performance as a function of academic ability. The first is that students of higher academic ability perform better (i.e. tend to generate more complex decision maps, integrate information better) because they have a greater investment in this particular decision. More able students are likely to be those to whom higher education matters more. A second, and somewhat related possibility, is that higher academic ability students perceive themselves as having more options in the first place than do students of lower academic ability. Students of lower ability might perceive their choices as already very restricted. In contrast, higher ability students may have found themselves already bombarded with information from counsellors, family members, and friends, such that the initial 'environmental press' to manage information occurs very early for them, and that by the beginning of this study, they were already developing more complex information-management strategies. A third possibility is that more academically able students simply have more cognitive resources with which to make any decision, and/or to respond to this particular task, which asked students to reflect quite a bit on their own information processing.

From this, a very rough measure of 'success' with the process was calculated, by dividing the number of acceptances by the number of places applied to.

These measures were entered as dependent variables in multiple regression analyses that included the following independent variables: academic ability, family income, gender, cohort, decision map complexity, and correlation of impressions with linear models. 'Decision map complexity' was computed by multiplying the number of self-generated criteria by the number of reported alternatives during the April 1992 session (to maximize the number of subjects included in the analysis). 'Correlation of impressions with linear models' was measured using the correlations between overall impressions and the equally weighted model predictions from the self-generated factors.

Very few of these analyses yielded any statistically significant results. In the few that did, only one predictor variable was entered into the final equation. As expected, academic ability was a significant predictor of success (beta = 0.34; R squared = 0.12; $F[1, 171] = 22.85$, $p < 0.001$). Somewhat unexpectedly, cohort was a significant predictor of reporting feeling comfortable with the decision-making process, in April 1992 (beta = −0.16; R squared = 0.03; $F[1, 177] = 4.92$, $p < 0.05$), but not in January 1993. Apparently, students who joined the study earlier had better feelings about the process, although the magnitude of the effect is small.

The lack of results in these analyses suggests that, however else they differ, students of differing levels of ability, family income, and genders do not feel differently about the process, either in process or retrospect. For the most part, students' feelings of stress, enjoyment, and certainty of their decision is independent of academic ability, family income, or gender. Moreover, students who created a more complex decision map by considering more alternatives and/or criteria did not express noticeably more (or less) confidence in the process, nor did they look back on their decisions with greater certitude. They also received no better 'payoff' in terms of higher rates of acceptance among the set of schools to which they applied.

GENERAL DISCUSSION

How do high school students go about the process of making a decision about college? At any given point in the process, students appear to consider four or five alternatives, and report using between eight to ten criteria in evaluating these alternatives. Higher ability students tend to consider more criteria, more distinct types of criteria, and more schools than do lower ability students, although the effect is small and only marginally significant. This trend suggests that higher ability students construct a slightly more complicated decision 'map' for themselves, complicating the task before them by using more criteria with which to make a decision. Note, however, that *all* students might be faulted for failing to consider a number of other potential criteria. For example, there were 34 factors listed on the standard form; most students gave each one a moderate or high importance weighting, despite the fact that even the higher ability students only spontaneously listed at most a third of these. In general, then, all participants could be described as under-exploring the decision.

It may be that students' restricted decision maps represent an adaptive information management strategy, as the restrictions help to keep in bounds the

Alternatively, it might be argued that the fit of data with linear models is *not* an especially good benchmark for information integration. Perhaps lower ability students also integrate information very well, but according to some non-linear, non-additive model. It is impossible to rule out this possibility, although its plausibility rests on an identification of such a non-linear model.

Interestingly, neither the complexity of students' decision maps, nor the degree to which students 'followed' linear models in their overall intuitive assessments showed any 'payoff' to the student in either satisfaction or 'success', as crudely measured here. That is, students with 'better' decision-making performance (more complex maps, more linearly predictable overall ratings) were not feeling especially better about the decision-making process as it was unfolding, did not feel better about it in retrospect, and did not select a final 'pool' of alternatives at which their final rate of acceptance was higher. At least in the measures reported, better cognitive performance did *not* yield better outcomes. Obviously, this preliminary finding needs more stringent testing in future work.

Given the somewhat arduous nature of this task (listing and weighting criteria, listing and ranking all current alternatives on all previously listed criteria, then listing and ranking all current alternatives on a set of 34 standard criteria, after weighting the importance of each), it is somewhat surprising that few practice effects emerged. No cohort effects were observed for the number of criteria or schools listed, nor in the type of criteria listed, and only a few were found for fit of linear models to subjects' data (and these were only higher order interactions). Predictably, subjects who had previously seen the standard list of 34 criteria were likely to list more of them spontaneously in a subsequent session, but the magnitude of the effect was slight.

What overall conclusions can be drawn about students' reports of how they are facing this decision? Students *do* find this a stressful decision to make, and they appear to restrict the set of information they actively consider at any point in time to about nine criteria and about four alternatives. There is little change in the *structure* at different points in the process, although the specific criteria and alternatives *do* change—about half of each are replaced during each 6-month period. Students' intuitions are correlated with linear models, but only moderately. The fit of linear models to students' impressions is better for academically able students, suggesting individual differences in the approach taken to this decision-making task. Finally, few practice effects were observed in decision-making performance for students who participated in multiple sessions.

One caveat to the above set of results is that the data gathered probably are *reports* of what is under consideration, and not a direct measure of it. Some of the questions asked of students may have assumed that they had greater access to their assessments of criteria and/or alternatives. The data on importance weights, for example, suggests that students' responses may not have been meaningfully related to the actual importance students placed on the criterion; i.e. that students did not have fully articulated values for these items (Fischhoff, 1991). It may be that different students are differentially able to report on less subtle aspects of their decision maps as well (e.g. what criteria and/or alternatives really were under consideration) and that this differential accounts for the significant group differences reported (or, alternatively, masked many results that would have turned out to be significant).

Students' reports of the major aspects of their decision maps, however, do seem plausible given other literature on adolescent decision making. The number of criteria they list is comparable, for example, to the number of consequences thought of by the adolescent (and adult) subjects of Beyth-Marom *et al.* (1993). If students are restricting the number of criteria or alternatives simply to manage information, then providing them with alternative information-management strategies is likely to improve the thoroughness with which they explore the decision space. Recent work suggests that peoples' decision making on laboratory problems *can* be improved even after only brief training sessions (Larrick, Morgan and Nisbett, 1990); it remains to be seen as to whether similar effects could be produced for real-life important decisions, and also what the degree of persistence of the effects would be.

Several other questions and new directions are raised by this research. One of the most provocative is the idea that the 'college decision' is really a series of interleaved decisions: Is college a possibility, either financially or academically? If so, should college be considered for immediately after graduation? Who or what sources should be consulted in thinking about these questions? How are conflicts among different sources resolved? When and how are alternatives 'put on a list' for active consideration, and just what does it mean to 'consider' an alternative? For example, is there a pre-decision screening phase to this process (Beach, 1993), and if so, what is its nature? How do students come to add a criterion or an alternative to their lists in the first place? How do they assign ranks to the schools under consideration with respect to one or more criteria? Under what circumstances do they make revisions to their criteria, their subjective importance weightings, or their rankings? How do they know when they have gathered enough information? At what point do others (e.g. parents, teachers, counsellors) assume decision-making responsibility, and how does this change the decision-making process?

Newer models of decision making, focusing less on linear and compensatory models, and more on intuitive, holistic processes have recently been sketched (e.g. Beach and Mitchell, 1987; Frisch and Clemen, 1994). Because these models as yet do not provide as clear or as testable predictions, they have not been assessed in this study. However, the lack of perfect fit of students' data with traditional linear models does suggest room for different models of decision making. Tests of such other models, of course, requires further theoretical development.

This study focused upon a particular real-life decision, which, as described earlier, has several unique properties. It is not clear how well performance on this decision would compare with performance on other real-life decisions of magnitude. The existence of college guidebooks, computer programs, and the recent emergence of freelance 'college counsellors', who help students and parents select colleges at which to apply, may represent a wealth of information unavailable in other real-life decisions. Thus, the findings in this study must not be blindly extrapolated to other real-life decisions.

An obvious next step, then, is to extend the study of real-life decision making to other naturally occurring decisions. To what degree are the findings reported here replicated in studies of, for example, the decision to have a child (or to have more children), to purchase a house, to develop a financial portfolio, to change careers? To what degree do existing laboratory models of decision making fit these such decisions? If differences in kind emerge for the different decisions, what factors account for them?

Taken as a whole, this study provides a picture of how an important, potentially life-framing, decision unfolds for adolescents over the course of a year. The study documents that students' untutored intuitions measure up, at least to some degree, against at least one normative model of decision making, although there is certainly a great deal of room for improvement. Refinements in the measures used are clearly needed, and investigation of how different 'episodes' of decision making fit together to form a larger whole are certainly indicated. Future work must address the typicality of this as a real-life decision, so that an assessment of how well the conclusions presented here apply to other real-life decisions.

REFERENCES

Arkes, H. R. and Hammond, K. R. (Eds.) (1980). *Judgment and decision making*. Cambridge: Cambridge University Press.

Aslin, A. W. (1971). *Predicting academic performance in college: selectivity data for 2300 American colleges*. New York: Free Press.

Baron, J. (1988). *Thinking and deciding*. Cambridge: Cambridge University Press.

Beach, L. R. (1993). Broadening the definition of decision making: The role of prechoice screening of options. *Psychological Science*, **4**, 215–220.

Beach, L. R. and Mitchell, T. R. (1987). Image theory: principles, goals, and plans in decision making. *Acta Psychologica*, **66**, 201–220.

Beach, L. R. and Potter, R. E. (1992). The pre-choice screening of options. *Acta Psychologica*, **81**, 115–126.

Bettman, J. R., Johnson, E. J., and Payne, J. W. (1990). A componential analysis of cognitive effort in choice. *Organizational Behavior and Human Decision Processes*, **45**, 111–139.

Beyth-Marom, R., Austin, L., Fischhoff, B., Palmgren, C., and Jacobs-Quadrel, M. (1993). Perceived consequences of risky behaviors: adults and adolescents. *Developmental Psychology*, **29**, 549–563.

Brodigan, D. L. (1984). *Admissions contact questionnaire*. Unpublished questionnaire. Carleton College, Northfield, MN.

Creyer, E. H., Bettman, J. R., and Payne, J. W. (1990). The impact of accuracy and effort feedback and goals on adaptive decision behavior. *Journal of Behavioral Decision Making*, **3**, 1–16.

Dawes, R. M. (1982). The robust beauty of improper linear models in decision making. In D. Kahneman, P. Slovic, and A. Tversky (Eds.), *Judgment under uncertainty: heuristics and biases*, (pp. 391–407). New York: Cambridge University Press.

Dawes, R. M. and Corrigan, B. (1974). Linear models in decision making. *Psychological Bulletin*, **81**, 95–106.

Edwards, W. (Ed.) (1992a). *Utility theories: measurements and applications*. Boston, MA: Kluwer Academic Publishers.

Edwards, W. (1992b). Toward the demise of economic man and women; bottom lines from Santa Cruz. In W. Edwards (Ed.), *Utility theories: measurements and applications* (pp. 253–267). Boston, MA: Kluwer Academic Publishers.

Fischhoff, B. (1991). Value elicitation: is there anything in there? *American Psychologist*, **46**, 835–847.

Frisch, D. and Clemen, R. T. (1994). Beyond expected utility: rethinking behavioural decision research. *Psychological Bulletin*, **116**, 46–54.

Galotti, K. M. (1989). Approaches to studying formal and everyday reasoning. *Psychological Bulletin*, **105**, 331–351.

Galotti, K. M. and Kozberg, S. F. (1987). Older adolescents' thinking about academic/vocational and interpersonal commitments. *Journal of Youth and Adolescence*, **16**, 313–330.

Galotti, K. M., Kozberg, S. F., and Appleman, D. (1990). Younger and older adolescents' thinking about commitments. *Journal of Experimental Child Psychology*, **50**, 324–339.

Galotti, K. M., and Mark, M. C. (1994). How do high school students structure an important life decision? A short-term longitudinal study of the college decision-making process. *Research in Higher Education*, **35**, 589–607.

Hammond, K. R., Hamm, R. M., Grasia, J., and Pearson, T. (1987). Direct comparison of the efficacy of intuitive and analytical cognition in expert judgment. *IEEE Transactions on Systems, Man, and Cybernetics*, **17**, 753–770.

Johnson, E. J., Payne, J. W., and Bettman, J. R. (1988). Information displays and preference reversals. *Organizational Behavior and Human Decision Processes*, **42**, 1–21.

Kahneman, D., Slovic, P., and Tversky, A. (Eds.) (1982). *Judgment under uncertainty: heuristics and biases*. Cambridge: Cambridge University Press.

Keating, D. P. (1980). Thinking processes in adolescence. In J. Adelson (Ed.), *Handbook of adolescent psychology* (pp. 211–246). New York: Wiley.

Keating, D. P. (1990). Structuralism, deconstruction, reconstruction: the limits of reasoning. In W. F. Overton (Ed.), *Reasoning, necessity, and logic: developmental perspectives* (pp. 299–319). Hillsale, NJ: Lawrence Erlbaum Associates.

Keeney, R. L. (1992). On the foundations of prescriptive decision analysis. In W. Edwards (Ed.), *Utility theories: measurements and applications* (pp. 57–72). Boston, MA: Kluwer Academic Publishers.

Keeney, R. L. and Raiffa, H. (1976). *Decisions with multiple objectives: preferences and value tradeoffs*. New York: Wiley.

Keller, L. R. (1992). Properties of utility theories and related empirical phenomena. In W. Edwards (Ed.), *Utility theories: measurements and applications* (pp. 3–23). Boston, MA: Kluwer Academic Publishers.

Larrick, R. P., Morgan, J. N., and Nisbett, R. E. (1990). Teaching the use of cost–benefit reasoning in everyday life. *Psychological Science*, **1**, 362–370.

Lusk, C. M. and Hammond, K. R. (1991). Judgment in a dynamic task: microburst forecasting. *Journal of Behavioral Decision Making*, **4**, 55–73.

Marcia, J. E. (1966). Development and validation of ego-identity status. *Journal of Personality and Social Psychology*, **3**, 551–558.

Mitchell, T. R. and Beach, L. R. (1990). '. . . Do I love thee? Let me count . . .' Toward an understanding of intuitive and automatic decision making. *Organizational Behavior and Human Decision Processes*, **47**, 1–20.

Ormond, C., Luszcz, M. A., Mann, L., and Beswick, G. (1991). A metacognitive analysis of decision making in adolescence. *Journal of Adolescence*, **14**, 275–291.

Payne, J. W. (1976). Task complexity and contingent processing in decision making: An information search and protocol analysis. *Organizational Behavior and Human Performance*, **16**, 366–387.

Payne, J. W. (1982). Contingent decision behavior. *Psychological Bulletin*, **92**, 382–402.

Payne, J. W., Bettman, J. R., Coupey, E., and Johnson, E. J. (1992). A constructive process view of decision making: Multiple strategies in judgment and choice. *Acta Psychologica*, **80**, 107–141.

Payne, J. W., Bettman, J. R., and Johnson, E. J. (1993). *The adaptive decision maker*. New York: Cambridge University Press.

Perkins, D. N., Allen, R., and Hafner, J. (1983). Difficulties in everyday reasoning. In W. Maxwell (Ed.), *Thinking: the expanding frontier*. (pp. 177–189). Philadelphia, PA: The Franklin Institute.

Reilly, B. A. and Doherty, M. E. (1989). A note on the assessment of self-insight in judgment research. *Organizational Behavior and Human Decision Processes*, **44**, 123–131.

Rowe, I. and Marcia, J. E. (1980). Ego identity status, formal operations, and moral development. *Journal of Youth and Adolescence*, **9**, 87–99.

Scheidel, D. G. and Marcia, J. E. (1985). Ego identity, intimacy, sex role orientation, and gender. *Developmental Psychology*, **21**, 149–160.

Slovic, P., Lichtenstein, S., and Fischhoff, B. (1988). Decision making. In R. C. Atkinson, R. J. Hernstein, G. Lindzey, and R. D. Luce (Eds.), *Stevens' handbook of experimental psychology, vol. 2. Learning and cognition (2nd. ed)* (pp. 673–738). New York: Wiley.

Torrance, E. P. (1972). Directions manual and scoring guide. *Torrance Tests of Creative Thinking*. Bensenville, IL: Scholastic Testing Service.

von Winterfeldt, D. and Edwards, W. (1986). *Decision analysis and behavioral research.* Cambridge: Cambridge University Press.

Wallach, M. and Kogan, N. (1965). *Modes of thinking in young children.* New York: Holt, Rinehart, and Winston.

Wright, P. and Kriewall, M. A. (1980). State-of-mind effects on the accuracy with which utility functions predict marketplace choice. *Journal of Marketing Research*, **17**, 277–293.