Res Eng Design
DOI 10.1007/s00163-011-0116-7

ORIGINAL PAPER

An approach to the extraction of preference-related information from design team language

Haifeng Ji · Maria C. Yang · Tomonori Honda

Received: 4 November 2008 / Revised: 23 May 2011 / Accepted: 12 July 2011

Abstract The process of selecting among design alternatives is an important activity in the early stages of design. A designer is said to express design preferences when assigning priorities to a set of possible design choices. However, the assignment of preferences becomes more challenging on both a practical and theoretical level when performed by a group. This paper presents a probabilistic approach for estimating a team's overall preference-related information known as preferential probabilities that extracts information from the natural language used in team discussion transcripts without aggregation of individual team member opinions. Assessment of the method is conducted by surveying a design team to obtain quantitative ratings of alternatives. Two different approaches are applied to convert these ratings into values that may be compared to the results of transcript analysis: the application of a modified Logit model and simulation based on the principle of maximum entropy. The probabilistic approach proposed in the paper represents how likely a choice is to be "most preferred" by a design team over a given period of time. A preliminary design selection experiment was conducted as an illustrative case example of the method. Correlations were found between the preferential probabilities estimated from transcripts and those computed from the surveyed preferences. The proposed methods may provide a formal way to understand and represent informal, unstructured design information using a low overhead information extraction method.

Keywords Design preferences · Design decision-making · Concept selection · Design process

H. Ji
Yahoo! Inc., 701 First Ave, Sunnyvale, CA 94089, USA

M. C. Yang (✉)
Department of Mechanical Engineering and Engineering Systems Division, Massachusetts Institute of Technology, 77 Massachusetts Ave, 3-449B, Cambridge, MA 02139, USA
e-mail: mcyang@mit.edu

T. Honda
Department of Mechanical Engineering, Massachusetts Institute of Technology, 77 Massachusetts Ave, 3-446, Cambridge, MA 02139, USA

1 Introduction

Throughout the design process, design teams often face the task of making choices between possible design options. Designers may assign preferences, or priorities, to a set of possible design choices in order to facilitate the selection process. Engineering design research has focused on several methods to aid in making design decisions, including Pugh charts (Pugh 1991), Quality Function Deployment (QFD) (Hauser and Clausing 1988), Decision-based Design (DBD) (Hazelrigg 1998), and the Method of Imprecision (MoI) (Wood and Antonsson 1989; Otto and Antonsson 1991). Such approaches generally require a designer to state his or her explicit, quantitative preferences. However, when such methods are employed by a team of designers rather than by an individual, the process of determining preferences may become more complex. First, it can be difficult to elicit group opinion in a formal fashion. Group preferences may be obtained informally through methods, such as group voting, consensus building, or command decision-making (Thompson 2003), or more formally through surveys or questionnaires in which respondents are asked to rank their choices (Packard 1979; Brans and Vincke 1985), rate choices with values (von Neumann and

Morgenstern 1947; Tribus 1969; Keeney and Raiffa 1976; Thurston 1991; Scott and Antonsson 2005), or select a "most preferred" choice (Luce 1959; Fishburn 1978; Hanley et al. 2001; Busemeyer and Diederich 2002). However, it is not always practical to use such strategies in an engineering workplace environment as their administration involves considerable overhead and may be disruptive. A second issue is how to arrive at a single set of preferences to represent a range of group opinion. Aggregation of individual opinions may be performed informally through group decision-making such as voting or consensus building (Arnold 2001) or more formally through mathematical aggregation of preference values. However, Arrow has shown that the results of aggregation cannot be guaranteed to simultaneously fulfill a number of axioms that cover notions of consistency, fairness, and autonomy (Arrow 1970; Arrow and Raynaud 1986). A third issue is that design preferences can evolve over the course of a design process as a team obtains new insights about a design, resulting in iterative stages of concept convergence and divergence (Giffin et al. 2009). For example, an automotive design team might favor a particular type of engine for their next car model, but if new information about engine technology is discovered or government regulations on fuel efficiency change, the team might alter their design.

This study investigates two questions. First, Can a design team's preference-related information be determined in a way that does not require designers to explicitly state their preferences or to formally aggregate individual opinions? This paper builds on earlier work (Ji et al. 2007) and presents a probabilistic approach for estimating the preference-related information embedded in transcripts of team discussion. The method is called Preferential Probabilities from Transcripts, or PPT. A probabilistic method is chosen because it can provide a quantitative way of expressing a group's likely preference for a design alternative. A preferential probability of an alternative is the likelihood it will be preferred over all others, also known as its "most preferred" probability. PPT is an implicit approach that does not require designers to explicitly express their preferences and so minimizes intrusion on design teams and increases the usability of the method. The approach further represents a group's overall preferential probabilities by extracting this information from the whole collection of words uttered by the team, thus avoiding the need for aggregation of individual opinions. Finally, it can provide a time-based profile of preferential probabilities that offers insights into changes in a team's design choices over time.

On a broader level, this work is intended to serve as a way to bridge design decision-making theory with design practice. Reich (2010) observes that design research encompasses a number of contrasting approaches and worldviews to thinking about design, including design practice and design theory. It is hoped that PPT is a step toward linking the behavior of practicing design teams with more formal design theory through the medium of design language.

This paper also asks: How do the preferential probabilities obtained through extraction from transcripts compare with those obtained explicitly through more traditional means? This focuses specifically on surveys, as they are a common way to elicit preferences from individuals and teams. A traditional approach to aggregation of group opinion includes averaging individual ratings, but this violates Arrow's Theorem. The approach taken here is to assume a distribution of possible group rating values. These group rating values must then be converted into a form that can be compared to PPT values. Two approaches for conversion are proposed, one based on the Logit model and a new method based on the principle of maximum entropy.

This paper includes a limited, preliminary case example of a three-person design team working on two design selection tasks to illustrate how the method can be applied. PPT was used to extract preferential probabilities from team discussion transcripts. Each individual design team member was periodically asked to complete surveys of their preferences, and these results were compared with those found using PPT.

2 Related work

2.1 Preference elicitation, extraction and evolution

Several formal approaches exist for modeling preference in design decision-making. In the lottery method (Krantz et al. 1971), alternatives are preferentially scaled from 0 to 1. Pairwise comparisons are made to determine the relative desirability of the alternatives. A pairwise approach is also used in the Analytical Hierarchy Process (Saaty 2000). Other quantitative pairwise comparison approaches include those formulated by Wang (1997), who employs a fuzzy preference relationship to discriminate among three preference models, and Li and Jin (2006), who apply this fuzzy preference relationship to select among alternatives. Scott and Antonsson discuss multi-criteria decision-making (1999) and aggregation functions (1998) to formally calculate overall preferences using the Method of Imprecision. All of these methods explicitly elicit preferences from designers or assume a value for them.

Probabilistic methods to extract preferences have also been focused on in research. Methods based on factorial design and statistics are conjoint analysis (Green and

Srinivasan 1990) and discrete choice analysis (Hensher and Johnson 1981; Ben-Akiva and Lerman 1985). Conjoint analysis requires respondents to make trade-offs between choices to reveal preferences. Discrete choice analysis identifies patterns individual respondents make between different alternatives. Wassenaar and Chen have employed discrete choice analysis in a Decision-Based Design framework (2003). These probabilistic ways to extract preferences are also based on explicit surveys of respondents.

A lower overhead approach to extracting preferences is collaborative filtering (Kohrs and Merialdo 2000). Collaborative filtering assumes that individuals with similar profiles gravitate toward the same choices. However, it requires a large number of opinions to be effective, far more than on a typical design team.

This study presents an alternative approach to extracting preference-related information that does not require designers to explicitly state their preferences and is intended for small teams rather than large populations.

Furthermore, the above methods are better suited to representing preferences at a single point in time. In practice, preferences can change over time. Changes may be due to the introduction of new information or changes in the state of the decision-making entity (Luce 1959) or to the dynamic nature of the selection process in a group (Bockenholt 2002). Design may also be thought of as a continuous process of making iterative trade-offs, particularly as teams better understand their objectives, requirements, and constraints. In this sense, the design process can be thought of as an evolution of preferences. By taking an implicit approach, this research allows preference-related information to be extracted at varying intervals in time in a way survey- and questionnaire-based methods are not intended to.

2.2 Group preference aggregation

In most engineering activities, multi-disciplinary collaborative teamwork is essential (Geslin 2006). Much work has been conducted to investigate the aggregation of group preference that arises in such scenarios. Arrow's Theorem (Arrow 1970; Arrow and Raynaud 1986) demonstrates that there is no guarantee of consistency in a group. Dym et al. (2002) discuss pairwise comparison charts to aggregate individual orderings of design team members into "group decision"-like voting. Keeney and Raiffa (1976) employ cardinal utility functions to accumulate group preferences. Jabeur et al. (2004) and See and Lewis (2006) further assume unequal weights for the preferences of the group members.

This study takes an approach that does not require knowledge of an individual's preferences in order to formulate weightings nor does require aggregation of individual preferences. Rather, the group is treated as a single entity that generates information about the group's preferences during discussion. These group preferences are then extracted directly from transcripts of team discussion without aggregation.

2.3 Embedded design information

Underlying this work is the assumption that the language that designers use during design discussion may reflect their design process and can provide valuable insights into the design process (Yang et al. 2005; Dong 2006a, b). Preference-related information may be embedded in the qualitative design information that designers generate, such as logbooks (Brockman 1996), sketches (Shah et al. 2001; Yang 2003), prototypes (Yang 2005), and design documentation (Mabogunje and Leifer 1996; Song et al. 2003). Researchers have also examined the design process through the perspective of surveys (Brockman 1996), design journals (Jain and Sobek 2006), "story telling" (Song et al. 2003), word frequencies and information certainty analysis (Yang and Ji 2007), and language aggregation and accumulation (Dong 2006a, b). However, these forms may reflect the opinions of a single designer rather than a design team. Design team transcripts are of interest because they can include the opinions of multiple team members as well as back-and-forth discussion of issues, providing a rich source of group design rationale to draw upon.

3 Methods

In this study, preferential probabilities from transcripts (PPT) was used to extract preference information from a case study described in Sect. 3.3. Individual ratings were also collected through surveys during the case study and converted using two different methods described in Sect. 3.2. PPT was then compared with the survey-based methods to assess their consistency.

3.1 Preferential probabilities from design team transcripts

Preferential probabilities from transcripts is a method to extract preference information from design team discussion. It is based on the following set of assumptions about how design teams express their decisions verbally:

Assumption 1 Design team discussion is embedded with sufficient information to reflect group preferences.

Furthermore, what designers say to each other during the design process generally corresponds with what they think. This is also an assumption of protocol studies of designers (Cross et al. 1996).

Assumption 2 All major design alternatives are largely known a priori. While this may not be necessarily true for novel design problems, it is a reasonable assumption for incremental or redesign problems in which design alternatives have been considered in the past.

Assumption 3 A discussion can be divided into time intervals during which designers' preferences are assumed to be unchanged. A change in preference can only occur between consecutive time intervals.

Assumption 4 The most preferred alternative in one time interval is related to the most preferred alternative in the previous time interval. This relationship between the two can be represented probabilistically to describe how likely the design team is to keep the most preferred alternative or to select another.

Assumption 5 Designers tend to speak positively about the design alternatives they have a stronger preference for and negatively about those they prefer less. Within the same time interval, how often an alternative is mentioned positively or negatively relates to how much it is preferred. However, people occasionally do not say what they mean (Grefenstette 1993; Bertrand and Mullainathan 2001), so PPT employs a probabilistic formulation to account for such stochastic uncertainty in group discussion.

The basic steps of deriving PPT are as follows:

1. Collect the word occurrences of all design alternatives in a transcript of a design team's discussion. The collection of word occurrences is called *utterance data*. Synonyms for the same alternative are also counted as occurrences.
2. Build the Preference Transition Model to describe the relationship between preferences in two consecutive time intervals, along with the Utterance-Preference Model to describe the relationship between what designers say and what designers prefer within the same time interval. The parameters of the two models are unknown (details in Sect. 3.1.2).
3. Assign reasonable initial values to the parameters of these two models.
4. Apply both models to a transcript to predict preference data. The preference data will be used to describe the evolution of preference-related information over the design process (details in Sect. 3.1.3).
5. Update the parameters of these two models using a traditional expectation–maximum (EM) algorithm (Dempster et al. 1977) on the predicted preference data and the given utterance data (details in Sect. 3.1.4).
6. Repeat steps 4 and 5 until there is convergence on the hidden parameters of the models. Parameters converge because the EM algorithm is guaranteed to improve the probability of the occurrences of the utterance data at each iteration (Bilmes 1998).

3.1.1 Notation

N	total number of possible design alternatives
T	total number of time intervals in the design process
a_m	m-th alternative
A	the set of all alternatives $\{a_1, a_2, \ldots, a_N\}$
π_i	the alternative that is preferred to all others in Time Interval i, i.e., the most preferred alternative
ε_i	an alternative uttered during Time Interval i of the design process
σ_i	the sequence of design alternatives uttered in Time Interval i. For example, if the design alternatives are uttered in Time Interval 2 as a_2, a_2, a_1, a_1, a_1, a_2, a_3, a_1,a_1, a_3, then $\sigma_2 = \{a_2, a_2, a_1, a_1, a_1, a_2, a_3, a_1, a_1, a_3\}$
$P\ (\pi_i = a_m)$	the probability that Alternative a_m is preferred to all other alternatives (i.e., Alternative a_m is "most preferred") in Time Interval i. If the preference value of Alternative a_m in Time Interval i is represented by $\mu_i\ (a_m)$ on a scale from 0 to 1, then $\pi_i = a_m$ is equivalent to $\mu_i(a_m) \geq \mu_i(a_n)$ for all $1 \leq n \leq N$
$P\ (\pi_i = a_m \vert \pi_j = a_n)$	the probability that Alternative a_m is preferred to all other alternatives in Time Interval i, given that Alternative a_n is preferred to all other alternatives in Time Interval j
$P\ (\varepsilon_i = a_m)$	the probability that an uttered alternative is Alternative a_m in Time Interval i
$P\ (\varepsilon_i = a_m \vert \pi_j = a_n)$	the probability that an uttered alternative is Alternative a_m in Time Interval i given that Alternative a_n is preferred the most in Time Interval j

3.1.2 Preference transition model and utterance-preference model

There are two models employed in PPT: the Preference Transition Model (PTM) and the Utterance-Preference Model (UPM). Both of these model a team's "most preferred" alternative. The Preference Transition Model relates the design team's preference in the current time interval to that in the next. It is the mathematical implementation of Assumption 4. At each transition between intervals, the design team can either (1) keep the most preferred alternative from the previous time interval or (2) change the most preferred alternative to another alternative. The transition relationship between one interval and the next depends on the preference strengths of the most preferred alternative and the less-preferred alternatives. In this study, a preliminary relationship is approximated in which all less-preferred alternatives in the current time interval are equally likely to become the most preferred alternative in the next time interval.

An alternative may be "most preferred" in two cases. The first is that the alternative is most preferred in both the previous and current time intervals. The second is that the alternative transitions from less-preferred to most preferred. The Preference Transition Model is expressed in Eq. 1.

$$P(\pi_{i+1} = a_n | \pi_i = a_m) = \begin{cases} p & \text{when } n = m \\ \frac{1-p}{N-1} & \text{when } n \neq m \end{cases} \tag{1}$$

where $0 \leq p \leq 1$ is a hidden parameter, which is the probability that the most preferred alternative is kept unchanged from one time interval to the next. The larger p is the more consistent preferences are over the design cycle; and the smaller p is the more likely preferences change.

The *Utterance-Preference Model (UPM)* relates the team's preference to the utterance of alternatives within the same time interval. It tries to approximate what designers think with what designers say. This model is the mathematical implementation of Assumption 5.

In this study, a design alternative can be uttered in a transcript in either a positive or negative sense. When a negative word (e.g., "no," "not," "hardly") appears near an alternative in a transcript, this utterance is regarded as negative, otherwise it is positive. One strategy would be to subtract the negative utterances from the count of positive utterances, but this could lead to negative sums. Instead, a negative utterance is counted as a positive utterance for all other alternatives. Since the model is probabilistic, it inherently considers cases when an alternative is mentioned that is not preferred the most. The model establishes the general pattern of how often a design team mentions the alternative they prefer the most and how often they mention the less-preferred alternatives. Similarly, a preliminary model assumes that less-preferred alternatives are equally likely to be uttered by designers within the same time interval. The Utterance-Preference Model is expressed in Eq. 2:

$$P(\varepsilon_i = a_n | \pi_i = a_m) = \begin{cases} q & \text{when } n = m \\ \frac{1-q}{N-1} & \text{when } n \neq m \end{cases} \tag{2}$$

where $0 \leq q \leq 1$ is a hidden parameter. This is the probability that the most preferred alternative is uttered in the discussion. In protocol studies of designers (Cross et al. 1996), it was assumed that what designers say generally corresponds with what they think. In this study, it is assumed that designers say what they prefer in most cases, i.e., $q > \frac{1-q}{N-1}$. The reasonable value range for q is $\frac{1}{N} < q \leq 1$.

3.1.3 Calculation of preferential probabilities

Once utterance data and preferential probabilities in the current time interval have been obtained, the preferential probabilities in the next time interval are calculated by the law of total probability:

$$P(\pi_i = a_k | \sigma_i, \sigma_{i-1}, \sigma_{i-2}, \ldots, \sigma_1) = \sum_{1 \leq m \leq N} P(\pi_i = a_k | \sigma_i, \sigma_{i-1} \ldots, \sigma_1, \pi_{i-1} = a_m)$$
$$P(\pi_{i-1} = a_m | \sigma_{i-1}, \sigma_{i-2}, \ldots, \sigma_1) \tag{3}$$

By Bayes' theorem,

$$P(\pi_i = a_k | \sigma_i, \sigma_{i-1}, \ldots, \sigma_1, \pi_{i-1} = a_m) = \frac{P(\sigma_i | \pi_i = a_k, \sigma_{i-1}, \ldots, \sigma_1, \pi_{i-1} = a_m) P(\pi_i = a_k | \sigma_{i-1}, \ldots, \sigma_1, \pi_{i-1} = a_m)}{\sum_{1 \leq n \leq N} P(\sigma_i | \pi_i = a_n, \sigma_{i-1}, \ldots, \sigma_1, \pi_{i-1} = a_m) P(\pi_i = a_n | \sigma_{i-1}, \ldots, \sigma_1, \pi_{i-1} = a_m)} \tag{4}$$

For convenience, Eq. 4 is simplified to Eq. 5 by eliminating utterance and preference terms in the historical intervals. This simplification is based on two assumptions: (a) the current utterance depends only on current preferences and (b) the current preference depends only on most recent preceding $(i-1)$-th preferences and current utterances. This is a reasonable assumption because historical utterance information is effectively captured in the preferences of the most recent preceding interval

$$P(\pi_i = a_k|\sigma_i, \sigma_{i-1}, \ldots, \sigma_1,\ \pi_{i-1} = a_m) = \frac{P(\sigma_i|\pi_i = a_k)P(\pi_i = a_k|\pi_{i-1} = a_m)}{\sum_{1 \le n \le N} P(\sigma_i|\pi_i = a_n)P(\pi_i = a_n|\pi_{i-1} = a_m)} \quad (5)$$

Substituting Eq. 5 back into Eq. 3 gives Eqs. 6 and 7. When $i \ge 2$,

$$P(\pi_i = a_k|\sigma_i, \sigma_{i-1}, \ldots, \sigma_1) = \sum_{1 \le m \le N} \frac{P(\sigma_i|\pi_i = a_k)P(\pi_i = a_k|\pi_{i-1} = a_m)}{\sum_{1 \le n \le N} P(\sigma_i|\pi_i = a_n)P(\pi_i = a_n|\pi_{i-1} = a_m)} P(\pi_{i-1} = a_m|\sigma_{i-1}, \sigma_{i-2}, \ldots, \sigma_1) \quad (6)$$

When $i = 1$,

$$P(\pi_1 = a_k|\sigma_1) = \sum_{1 \le m \le N} \frac{P(\sigma_1|\pi_1 = a_k)P(\pi_1 = a_k|\pi_0 = a_m)}{\sum_{1 \le n \le N} P(\sigma_1|\pi_1 = a_n)P(\pi_i = a_n|\pi_0 = a_m)} P(\pi_0 = a_m) \quad (7)$$

Suppose there are w_i utterances of design alternatives in the ith time interval, as $a_i^{(1)}, a_i^{(2)}, a_i^{(3)}, \ldots, a_i^{(w_i)}$, which are all in Alternative Set A. If it is assumed that designer's preference is constant within the time interval, then

$$P(\sigma_i|\pi_i = a_k) = \prod_{u=1}^{w_i} P(\varepsilon_i = a_i^{(u)}|\pi_i = a_k) \quad (8)$$

Preferential probabilities in the next time interval are recursively calculated from the preferential probabilities in the current time interval using Eqs. 6, 7, and 8. In order to complete the recursive calculation, two pieces of information are needed:

- The initial preferential probabilities of all alternatives before the first time interval.
- The hidden parameters p and q of the Preference Transition Model and Utterance-Preference Model.

Initial preferential probabilities The first piece of information may be obtained by: (1) conducting surveys of designers before the start of the design process, (2) collecting preference-related information from an earlier design process, (3) analyzing preferences from the design of similar products, or (4) establishing an unbiased starting point which assumes a uniform distribution of alternatives.

Hidden parameters The second piece of information may be estimated using a more detailed approach. If the PTM and UPM models are known, it is feasible to calculate the preferential probabilities for each design alternative for every time interval and then plot the evolution of preferences over the life of the design discussion. However, the parameters of these two models are initially unknown. In this situation, utterance data are observable but preference data are not, and the models are incomplete because of the hidden parameters. An EM algorithm (Dempster et al. 1977) is often used in statistics for finding maximum likelihood estimates of parameters in probabilistic models where the models depend on unobserved hidden variables. In this study, it is applied to seek the values of the two hidden parameters of the two models.

An EM algorithm has two steps, the E-step and the M-step. The E-step estimates unobservable data. It can be accomplished by Eqs. 6 and 7. The M-Step computes the maximum likelihood estimates of the parameters by maximizing the expected likelihood found on the E-step. In this study, it corresponds with estimating the values of p and q, which make the utterance sample of the design alternatives occur in the discussion with the maximal likelihood.

From Eq. 1, it is known that $P(\pi_{i+1} = a_k|\pi_i = a_k)$ is independent of k and i. It means that no matter what time interval it is in, no matter which alternative designers prefer the most in the previous time interval, designers have a fixed probability of not changing the most preferred alternative.

Using the maximum likelihood (Fisher 1922), $P(\pi_{i+1} = a_k|\pi_i = a_k)$ can be estimated as

$$P(\pi_{i+1} = a_k|\pi_i = a_k) = \frac{\sum_{1 \le i \le T-1} \sum_{1 \le m \le N} C(\pi_{i+1} = a_m, \pi_i = a_m)}{\sum_{1 \le i \le T-1} \sum_{1 \le n \le N} \sum_{1 \le m \le N} C(\pi_{i+1} = a_n, \pi_i = a_m)} \quad (9)$$

where $C(\pi_{i+1} = a_n, \pi_i = a_m)$ is a fractional count that tracks the cases that a_n is most preferred in the current time interval, while a_m is most preferred in the previous time interval. $C(\pi_{i+1} = a_n, \pi_i = a_m)$ can be calculated as follows.

$$C(\pi_{i+1} = a_n, \pi_i = a_m) = P(\pi_{i+1} = a_n|\sigma_{i+1}, \sigma_i, \ldots, \sigma_1) P(\pi_i = a_m|\sigma_i, \sigma_{i-1}, \ldots, \sigma_1) \quad (10)$$

Fractional counts are fractional numbers whose values are proportional to the integral numbers which count the cases when $\pi_{i+1} = a_n$ and $\pi_i = a_m$. Thus, fractional counts can be used in Eq. 9 to estimate the parameter p.

Similarly, from Eq. 2, $P(\varepsilon_i = a_k|\pi_i = a_k)$ is independent of i and k. It means that in a certain time interval, designers have a fixed probability that they will utter the same alternative as the one they prefer the most.

Assuming the maximum likelihood, $P(\varepsilon_i = a_k|\pi_i = a_k)$ can be estimated as

$$P(\varepsilon_i = a_k | \pi_i = a_k) = \frac{\sum_{1 \le i \le T} \sum_{1 \le m \le N} C(\varepsilon_i = a_m, \pi_i = a_m)}{\sum_{1 \le i \le T} \sum_{1 \le n \le N} \sum_{1 \le m \le N} C(\varepsilon_i = a_n, \pi_i = a_m)} \quad (11)$$

where $C(\varepsilon_i = a_n, \pi_i = a_m)$ is also a fractional count, which counts the number of cases that a_n is uttered, while a_m is most preferred within the same time interval. It can be calculated as follows:

$$C(\varepsilon_i = a_n, \pi_i = a_m) = C(\varepsilon_i = a_n) P(\pi_i = a_m | \sigma_i, \sigma_{i-1}, \ldots, \sigma_1) \quad (12)$$

where $C(\varepsilon_i = a_n)$ is the number of utterances of Alternative a_n in the time interval i.

Equations 11 and 12 calculate q based on the samples of alternative utterances and preferences. When using the above procedure to calculate q, it should be noted that the value of q should be more than *1/N*.

Because the EM algorithm is guaranteed to improve the probability of the sample of alternative occurrences at each iteration, p and q will converge to values which try to maximize this probability (Bilmes 1998). These converged values can be regarded as the parameters for the Preference Transition Model and Utterance-Preference Model. The shortcoming of the EM algorithm is that it may converge to a local optimum rather than a global one. The use of multiple initial estimates may help avoid being trapped in local optima. Simulated annealing can also be combined with the EM algorithm to overcome the local optima problem (Ueda and Nakano 1998; Ueda et al. 2000).

3.1.4 Time intervals

For the sake of simplicity, it is assumed that designers do not change their preferences for design alternatives within a single time interval. Preferences can only change at the transitions between time intervals. The length of each interval depends on the desired granularity of preference evolution. Intervals may be of fixed- or variable length. In this study, interval changes occur at approximately 10-min intervals.

3.2 Methods to evaluate PPT

A simple approach to assessing the validity of PPT is to look for consistency between the original transcript and the preferential probabilities extracted from it (Ji et al. 2007). Comparison between PPT results and a qualitative reading of the transcript show a similar pattern of preferences for alternatives over time. However, more quantitative approaches for evaluation of important for two reasons. First, both the extracted preferential probabilities and qualitative readings are drawn from a single source—the discussion transcript. Second, a qualitative assessment cannot describe to what degree the preferential probabilities extracted with PPT correlate with the team's intentions.

To evaluate whether or not PPT is a viable method for extracting preferential probabilities from transcripts, this study turns to ratings and rankings drawn from surveys of designers as they discuss design choices. Surveyed preference information addresses both of the issues described above. However, such numerical ratings that cannot be directly compared with preferential probabilities and must be converted into a more appropriate form. This section formulates two approaches for translating survey results: one based on the Logit model and a new one based on simulation using maximizing information entropy that is called preferential probabilities from surveys (PPS).

3.2.1 Approach 1: Translating survey data into preferential probabilities using a modified Logit model

In the fields of transportation, economics, and engineering design, the "most preferred" choice is often modeled using utility functions. These can be grouped into two approaches. One is based on constant utility functions (Luce 1959). It assumes that the utilities of alternatives are constant and that the preferential probabilities are defined by a probability distribution function parameterized by these utilities. Another approach relies on random utility functions (Manski 1977). It takes into account inconsistencies in individual choice behavior by assuming that a utility function consists of a deterministic component and a random disturbance. When preference ratings are regarded as utilities, this model can be applied to convert preference ratings into preferential probabilities.

Note that there are some limitations on using these models to convert survey ratings into preferential probabilities. First, there may be unknown parameters in these models. For example, the Logit model (Ben-Akiva and Lerman 1985; Wassenaar and Chen 2003; Wassenaar et al. 2005), a widely used random utility model, is sensitive to different values of its scale factor. Second, these models assume a priori that the shapes of the distribution are the same even with different stated ratings, but because of the boundary constraints of typical ratings schemes, e.g., [0, 1], the shape of the distribution may be distorted at different rating values, especially near the boundaries. The third is that these models assume that the ratings are in terms of utilities which are additive, while preference ratings in engineering design may have different properties (Otto and Antonsson 1993). In this paper, probabilistic preferences for all alternatives are calculated using the Logit model as shown below:

$$P(\pi_i = a_m) = \frac{\lambda e^{\mu_i(a_m)}}{\sum_{k=1}^{N} \lambda e^{\mu_i(a_k)}} \tag{13}$$

where λ ($\lambda = 0$) is the scale factor of the Logit model. The Logit model assumes the utilities or ratings are logistically distributed (similar to a normal distribution) because of uncertainty. These lambda values are found using the probability function:

$$F(\varepsilon) = \frac{1}{1 + e^{-\lambda\varepsilon}} \tag{14}$$

where ε is the difference between two ratings. In this study, ratings range from [0, 1], so the possible range for ε is $[-1, 1]$, which differs from its original use with utilities in which $\varepsilon \in [-\infty, \infty]$.

The positive scale parameter λ implies how large the tail of the logistical distribution is. When it becomes infinity, all ratings are deterministic, and preferential probabilities are either 0 or 1. When it becomes zero, all choices are equally likely to be selected as the most preferred. Therefore, the values of the preferential probabilities translated from ratings with the modified Logit model are highly dependent on λ. Given the limited sample size in this study, it considers a most probable range for the scale factor λ for comparison with PPT, rather than determining the most appropriate value for the scale factor. For this study, the following two questions must be answered for finding a range for λ.

1. Given that the preferential probability of one alternative over another is 0.5 when the difference between two ratings is 0, what is a reasonable preferential probability if the difference between the ratings for those two alternatives is very small but greater than 0? For example, suppose the aggregated rating for one alternative is 0.50 and 0.49 for another. What is the probability that the team actually prefers the first alternative over the second?
2. Given that the preferential probability for one alternative over another is 1 when the difference between two ratings is of the maximal possible difference (1 in this study), what is a reasonable preferential probability if the difference is very big but less than 1? For example, suppose the aggregated rating for one alternative is 0.95 and 0.05 for another. What is the probability that the team actually prefers the first alternative over the second?

Answers to these two questions can be used to estimate a reasonable range, though not an exact value, for λ, which can help determine the possible range for the converted preferential probabilities.

3.2.2 Approach 2: preferential probabilities translated from surveys with simulation (PPS)

The second approach for translating preference ratings gathered from surveys into preferential probabilities is a new simulation approach that draws on the principle of maximum entropy (Jaynes 1957; Jaynes 1968). The principle of maximum entropy is chosen because it gives the least biased distribution with the given information. This method does not assume a distribution a priori. The distribution and the parameters are calculated while maximizing the information entropy so that there are no unknown parameters such as λ in the Logit model. This approach also considers boundary constraints while applying the principle of maximum entropy, which generates distinctive distributions for different stated ratings.

This new method, preferential probabilities from surveys (PPS), assumes preference ratings can be random for both individuals and the team and applies the principle of maximum entropy to both individual survey ratings and group ratings. A simulation is run to collect statistical results for estimating the preferential probabilities. PPS includes three main parts:

1. Construct a probability distribution for each individual rating preference for each alternative;
2. Construct a probability distribution for the group rating preference for each alternative;
3. Generate group preferential probabilities through simulation.

3.2.2.1 Part 1: Construction of a probabilistic distribution for individual preferences A design team's preferences may not always be consistent and is one reason preferences can be challenging to study formally. In research on how choices are made, individuals may not always select the same alternative when faced with the same situation more than once (Manski 1977; Ben-Akiva and Lerman 1985). In this work, a distribution was constructed to map individual ratings into a range of possible values to help account for this potential variability in individual preferences. The assumption of the mapping is that the rating a designer gives is the expected value of the distribution and is one of the conditions that this distribution needs to satisfy. In this study, the distribution of ratings is determined by maximizing the entropy of the distribution with the given data. In surveys, all the rating values are bounded to represent design preference rather than a utility value (see Otto and Antonsson 1993). In this work, ratings ranged from [0, 1] where 1 correlates with the highest preference rating for an alternative. Suppose l is the lower bound, u is the upper bound, and r is the expected value (average) of the distribution. Let $f(x)$ be the distribution function, the entropy of the probability distribution function is:

$$-\int_{l}^{u} f(x)\ln(f(x))\mathrm{d}x \tag{15}$$

with the constraints

$$\int_{l}^{u} f(x)\mathrm{d}x = 1 \tag{16}$$

$$\int_{l}^{u} xf(x)\mathrm{d}x = r \tag{17}$$

Equation 16 guarantees that the total probability between the bounds adds up to 1, and Eq. 17 means that the stated rating value given by the designer is exactly the expected value of the distribution.

Using a Lagrange multiplier and Euler–Lagrange equation, the maximization of the entropy with the above two constraints becomes:

$$\frac{\delta}{\delta f}\{-f(x)\ln(f(x)) + \lambda_a f(x) + \lambda_b x f(x)\} = 0 \tag{18}$$

From Eq. 18, $f(x)$ can be represented in an exponential form as shown in Eq. (19).

$$f(x) = \lambda_0 e^{\lambda_1 x}, \quad (l \le x \le u) \tag{19}$$

where λ_0 and λ_1 are the parameters of the distribution, which are determined by the boundary constraint and the average value.

Substituting $f(x)$ in Eqs. 16 and 17 with Eq. 19, then λ_0 and λ_1 can be solved from the set of Eqs. 16 and 17 using Newton's method (Kelley 2003). In this study, MATLAB was employed to solve this set of equations.

When $r < (l + u)/2$, i.e., the expected rating is smaller than the middle of the range, λ_1 will be negative, and the distribution of the rating is a truncated exponential distribution decaying from the lower bound to the upper bound; when $r > (l + u)/2$, λ_1 will be positive, and the distribution becomes a mirrored truncated exponential distribution decaying from the upper bound to the lower bound. There are three more extreme cases: (1) when $r = (l + u)/2$, λ_1 will become 0, and the distribution is reduced to a uniform distribution between the lower and upper bounds; (2) when $r = l$, λ_1 will be negative infinity, and the distribution is reduced to a Dirac delta distribution at the lower bound, which means the alternative is rejected; (3) when $r = u$, λ_1 will be positive infinity, and the distribution is reduced to a Dirac delta distribution at the upper bound, which means the alternative is accepted. Figure 1 shows instances of the distributions for the stated rating 0.1, 0.2, and 0.7 when the bounded range is [0, 1].

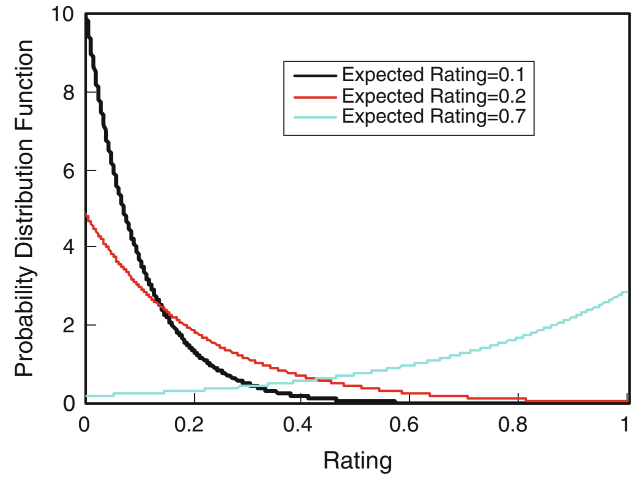

Fig. 1 Instances of rating distribution with maximization of entropy

This study also considers the case where ratings are given in relative terms rather than absolute. For example, the sampling from the distribution function may have an implicit constraint that the sampling ratings from the probability distribution should sum up to a certain value, such as when designers have 10 points to allocate among alternatives.

Suppose an individual designer is given W points to allot among N alternatives. The relative ratings the designer assigns are $r_1, r_2, \ldots r_N$. The possible relative values for Alternative i are in the range $[l_i, u_i]$. Let $x_1, x_2, \ldots x_N$ be the sampling variable for the relative ratings for each alternative. The joint distribution function can be represented by

$$f(x_1, x_2, \ldots, x_N)$$

With the constraint that $x_1 + x_2 + \cdots + x_N = W$, the function can be reduced to

$$f(x_1, x_2, \ldots, x_{N-1})$$

Similarly, by maximizing the entropy of the joint distribution function with the constraint that the expected value on variable x_i is r_i, $f(x_1, x_2, \ldots, x_{N-1})$ can be represented as in Eq. 20.

$$f(x_1, x_2, \ldots, x_{N-1}) = \begin{cases} \lambda_0 e^{\lambda_1 x_1 + \lambda_2 x_2 + \ldots + \lambda_{N-1} x_{N-1}} & \text{if } W - u_N \le \sum_{k=1}^{N-1} x_k \le W - l_N \\ & \text{and } l_i \le x_i \le u_i, \quad \forall\, i \in [1, N-1] \\ 0 & \text{otherwise} \end{cases} \tag{20}$$

The above distribution function shows that the joint exponential distribution is only meaningful when all the sample variables are in the possible ranges, otherwise it is zero.

$\lambda_0, \lambda_1, \ldots, \lambda_{N-1}$ can be solved from the following N equations.

$$\int_{l_1}^{u_1}\int_{l_2}^{u_2}\cdots\int_{l_{N-1}}^{u_{N-1}} f(x_1, x_2 \ldots, x_{N-1}) = 1$$
$$\int_{l_1}^{u_1}\int_{l_2}^{u_2}\cdots\int_{l_{N-1}}^{u_{N-1}} x_1 f(x_1, x_2 \ldots, x_{N-1}) = r_1$$
$$\vdots$$
$$\int_{l_1}^{u_1}\int_{l_2}^{u_2}\cdots\int_{l_{N-1}}^{u_{N-1}} x_{N-1} f(x_1, x_2 \ldots, x_{N-1}) = r_{N-1}$$

The first equation guarantees that the total probability is 1 integrated over the possible rating ranges for the joint distribution, and the next $N - 1$ equations set the requirements for the expected values for variable x_1 to x_{N-1}. The expected value for variable x_N is met tacitly because $E(x_N) = E(W - x_1 - x_2 - \cdots x_{N-1}) = W - r_1 - r_2 - \cdots r_{N-1} = r_N$.

Like Fig. 1, Fig. 2 shows a ratings distribution for three alternatives with expected ratings of 0.1, 0.2, and 0.7, but a constraint holds the sum of the sampled ratings to 1. In comparing Fig. 2 with Fig. 1, it is observed that in Fig. 2, the distribution drops down near the upper bound and is especially obvious for distributions with high expected values. In this case, the sampled rating for one alternative is determined by the sampled ratings for all other alternatives. If the sum of the ratings for all other alternatives is greater than 1, then the set of samples is invalid and has to be disregarded because it does not meet the constraint. With this constraint, points with higher values in the distribution are more likely to be ignored in sampling.

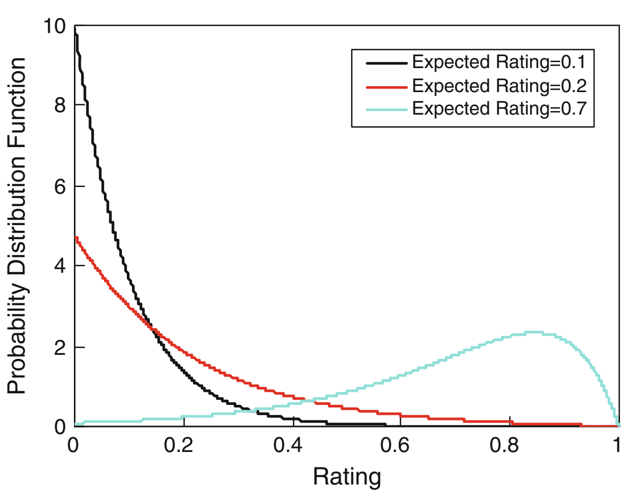

Fig. 2 Instances of rating distribution (the sum of three sampled ratings is 1)

3.2.2.2 Part 2: Construction of a probabilistic distribution for group preferences This paper considers two of the challenges in determining a team rating from individual ratings. One challenge is that any individual rating can be uncertain, either due to fuzziness ("does the person think of a rating of 0.3 or 0.4 as roughly the same?") or simple human error ("I meant 0.3 but marked 0.4 by accident"). A second challenge is the role of team organizational issues and social dynamics. Individuals may start group discussion with their own preferences, but both the content of group discussion and the nature of group dynamics may influence individuals and teams to change these preferences over time.

In this study, it is assumed that group ratings are bounded somewhere between the highest individual rating and the lowest individual rating. The construction of the distribution for team ratings is similar to the approach described in Part 1 and includes information about the lower bound, the upper bound, and the weighted average rating. These three values, average weights, upper bound, and lower bound, are used to establish the likely range and shape of the possible values for group opinion. The weighted average rating can be estimated in several ways. If there is no hierarchy and no apparent leader in the team, the weightings can simply be assumed equal or proportional to an individual's utterances in the discussion. Weightings may be adjusted to reflect information such as an individual's leadership, expertise, and member importance (Jabeur et al. 2004; See and Lewis 2006).

3.2.2.3 Part 3: Computation of a group's preference-related information Monte Carlo simulation is used to determine the chances that one alternative has a higher group rating than any other alternative. In each round of simulation, individual ratings are sampled from the probabilistic distribution constructed in Part 1, and then group ratings for each alternative are found from the probabilistic distribution in Part 2. By comparing the group ratings of the alternatives, the most preferred alternative is determined. The preferential probability for an alternative can be estimated from statistical simulation by calculating the proportion of rounds when this alternative has the highest simulated group rating.

When there is no constraint on the sum of the sampled ratings on all alternatives, the steps of simulation are as follows:

1. Construct a distribution for each individual member's rating for each alternative, as described in Part 1;
2. From each distribution, randomly select a sample as the "true" individual rating;
3. For each alternative, based on the sampled individual ratings, construct a distribution for the possible group rating, as described in Part 2;

4. Sample to get the group rating for each alternative;
5. Compare the group ratings to determine the most preferred alternative;
6. Repeat these five steps until the predefined maximum number of simulation runs is reached. The group's preferential probabilities can be estimated from the statistical simulation results.

If there is a constraint that the total sampled ratings on all alternatives are fixed (e.g., designers allot a fixed number of points to the alternatives), then the joint distributions for both individual and group ratings are used as described in Part 1, and only *(N-1)* alternative ratings are to be sampled, the left one is determined by subtracting the sampled ratings from the predetermined sum, e.g., 1 when the ratings are normalized values as shown in this study.

In the simulation for this study, the rejection method (Ross 2006; Press et al. 2007) can be employed to generate the samples for the distribution functions. The rejection method can generate sampling values from an arbitrary probability distribution function.

3.3 Methods to compare preferential probabilities from PPT and surveys

This section describes two measures to compare group preferential probabilities translated from surveys with those extracted from the design discussion transcript (PPT). In each of these methods, the set of preferential probabilities is represented in vector form, with each element representing a specific preferential probability for an alternative at a specified interval. These vectors are V and W, with n elements in each vector.

$$V = (v_1, v_2, \ldots, v_n)$$

$$W = (w_1, w_2, \ldots, w_n)$$

where $v_i\,(1 \le i \le n)$ is the i-th preferential probability extracted from the transcript with PPT and $w_i\,(1 \le i \le n)$ is the ith preferential probability converted from rating values from surveys with PPS.

3.3.1 Geometric distance

Geometric distance indicates how far two groups of preferential probabilities are from one another. In this study, the L_2 norm (Euclidean metric) as shown in Eq. 21 is employed to measure the distance between two groups of preferential probabilities. The smaller the distance, the smaller the average difference between these two groups of preferential probabilities is.

$$L_2(V, W) = \|V - W\| = \sqrt{\sum\nolimits_{i=1}^{n} (v_i - w_i)^2} \tag{21}$$

3.3.2 Pearson correlation

The Pearson product-moment correlation coefficient is calculated between surveyed and extracted values. Correlation coefficients can range from +1 to -1. The closer the correlation coefficient is to +1/-1, the more two variables are correlated positively/negatively. The Pearson product-moment correlation coefficient is calculated by Eq. 22.

$$r(v, w) = \frac{\sum\nolimits_{i=1}^{n} (v_i - \bar{v})(w_i - \bar{w})}{(n-1)s_v s_w} \tag{22}$$

where $\bar{v}$ and $\bar{w}$ are the sample means of v and w, s_v and s_w are the sample standard deviations of v and w, and n is the sample size.

4 Case study

This section presents a preliminary case study of a coffee maker design to illustrate the application of the PPT extraction method, its expected outcomes, and comparisons with other strategies for eliciting preference information, such as surveys. This case study is limited in its scope and is intended to offer anecdotal evidence of the method's value, rather than conclusive validation.

4.1 Coffee maker design selection task

A small design team consisting of a group of three graduate students in engineering at a US university was given a design selection task. The team was asked to decide on two-component selection design problems (a carafe and a filter for a coffeemaker), each of which had three candidate alternatives. Total cost for these two components together could not exceed $35:

"Imagine you are a retired person who is a coffee connoisseur. Your day cannot begin until you make coffee each morning for you and your spouse. You are in good health but are not as strong or mobile as you were when you were younger. As a connoisseur, you prefer fresh ground coffee to instant coffee like Folger's, and you are well informed about the various types of gourmet coffee available, as well as the tools and equipment to prepare it. However, you are now on a fixed income and are conscious about how you spend your money which is why you make coffee at home rather than visit Peet's every morning."

Before the experiment, each team member was given a think-aloud training exercise to practice speaking each alternative with its proper name rather than an ambiguous pronoun ("this" or "that") in order to facilitate the tracking of design alternatives in the transcript. For example, a "glass carafe" can only be called "glass carafe", "glass

Table 1 Features and specifications of case study coffee carafe

Name/ID	Glass carafe	Stainless steel carafe	Plastic carafe
Photo			
Description	Glass with warming plate	Thermal insulated stainless steel	Thermal insulated plastics (inside glass)
Cost	$10.00	$20.00	$15.00
Warming plate cost	$5.00	0	0
Footprint size	Big	Small	Small
Fragility	Fragile	Strong	Fragile material inside
Durability (reliability)	Durable	Durable	Less durable
Heat retention	Good with heating plate	OK with double layers of steel	Good with mirror glass inside
Weight	Light	Heavy	Light
Portability	Not portable	Portable	Portable
Easy to clean	Easy to clean	Not easy to clean	Not easy to clean
Style and esthetic value	Moderate attractive	Very attractive	Not attractive
Capacity	Can be designed as wanted Available for 2 cups and 6 cups	Can be designed as wanted Available for 2 cups and 6 cups	Can be designed as wanted Available for 2 cups and 6 cups
Spout	Not dribbles after pouring	Dribbles after pouring	Dribbles after pouring
Can tell how much coffee is left	Yes	No	No

Table 2 Features and specifications of case study coffee filter

Name/ID	Gold tone filter	Paper filter	Titanium filter
Photo			
Description	Permanent gold tone filter	Disposable paper filter	Titanium permanent coffee filter
Cost	$9.99	$3.99/100	$19.99
Durability	Permanent use	Disposable	Permanent use
Styling	Neutral golden color	N/A	Fashionable color
Cleanability	Clean after use and will never stain	Disposable, single use	Clean after use and may stain
Portability	Not foldable	Collapsible/foldable, easily portable	Not foldable
Easy to remove from the cone	No	No	Yes, with the handle
Coffee aroma retention	Can absorb and retain coffee aroma	N/A	Good at absorbing and retaining coffee aroma with continuous coffee cooking

pot", "glass coffee carafe", "glass coffee pot", "carafe A", "pot A", or "glass alternative." During the experiment, they freely discussed their preferences and rationale with each other until a consensus was reached. This discussion was audio- and video-recorded and transcribed.

Table 1 lists the three carafe alternatives (glass, stainless steel, and plastic). Table 2 lists the three filter alternatives (gold tone, paper, and titanium). The designers were provided additional features and specifications that might play a role in their preferences for the carafe and filter in Tables 1 and 2.

4.1.1 Design team discussion and surveys

During the design decision-making task, participants were surveyed to elicit their preferences. Surveys are used in fields such as marketing to elicit individual and group preferences. One of the issues in studying preferences is that they may be ambiguous. Hey (1998) as well as Kulok and Lewis (2005) note that human designers may not be consistent when they state their preferences explicitly. This has the potential to make quantitative analysis of surveyed preferences difficult. The approach taken in this paper is to examine overall trends in preferences across a number of design alternatives, rather than assume that the findings for one alternative at one point in time are correct.

Studies of team discussion suggest that team members enter into discussion armed with only partial, independent knowledge of a topic, and group discussion can play a role in eliciting this partial knowledge to form more complete, shared group knowledge so that better decisions may be made (Gigone and Hastie 1997). In order to encourage discussion among the participants and to better simulate a real-world team experience, information about the design choices was provided in the following ways:

- Individuals were provided detailed information regarding only one of the three alternatives (for example, only the glass carafe) in order to simulate a more realistic, partial knowledge scenario. Team members would then discuss product features in order to uncover information about the other alternatives (the stainless steel and plastic carafes).
- Paper-based surveys were completed individually, and participants had no knowledge of their teammates' responses.
- Individuals were encouraged to provide a brief rationale for their rating and ranking to decrease the possibility of arbitrary ratings.

During the task, participants were asked to complete surveys with their preference ratings for design choices at ~10-min intervals. The experiment lasted 50 min, including 10 min for instruction and training, and 8 min for filling out 5 surveys during the session. The three designers (X, Y, and Z) were each given 10 points to allot to three alternatives, with a higher number representing a higher preference. In this way, relative preferences were obtained that would permit comparison with the relative preference information obtained with PPT. Table 3 details designers' normalized survey ratings. Interval 0 marks the time before the design process had started.

4.2 Transcript analysis using PPT

Table 4 shows the number of times each of the six alternatives was uttered in 10-min intervals, including time for completing surveys. Note that designers discussed the carafe selection problem throughout experiment, but did not start discussing the filter until Time Interval 2.

In this example, initial values of the hidden parameters in Eqs. 1 and 2 were randomly chosen to be $p_1 = 0.4$ and $q_1 = 0.5$ for both component selection problems. Any initial values for p and q were appropriate so long as $0 < p < 1$ and $1/3 < q < 1$, as values would be updated in future iterations). The initial preferential probabilities at the beginning of the design discussion could be given in several ways, as described in Sect. 3.1.3. In the case where no survey information is available (i.e., Time Interval 0 before team discussion begins), equal likelihoods can be assigned for initiating PPT. However, if there is survey rating information available, PPT can be initialized with

Table 3 Collected survey ratings

Interval	Designer	Carafe			Filter		
		Glass	Steel	Plastic	Gold tone	Paper	Titanium
0	X	0.5	0.3	0.2	0.2	0.3	0.5
	Y	0.4	0.5	0.1	0.2	0.7	0.1
	Z	0.4	0.5	0.1	0.3	0.6	0.1
1	X	0.5	0.3	0.2	0.3	0.5	0.2
	Y	0.5	0.4	0.1	0.2	0.7	0.1
	Z	0.6	0.4	0	0.3	0.6	0.1
2	X	0.5	0.2	0.3	0.3	0.5	0.2
	Y	0.5	0.4	0.1	0.2	0.7	0.1
	Z	0.6	0.4	0	0.2	0.8	0
3	X	0.5	0.3	0.2	0.2	0.5	0.3
	Y	0.6	0.3	0.1	0.2	0.7	0.1
	Z	1	0	0	0	1	0
4	X	0.5	0.3	0.2	0.3	0.2	0.5
	Y	0.6	0.3	0.1	0.3	0	0.7
	Z	1	0	0	0	0.3	0.7

Table 4 Number of times alternatives were uttered during case study

Interval	Carafe			Filter		
	Glass	Steel	Plastic	Gold tone	Paper	Titanium
1	13	8	7	0	0	0
2	12	7	6	3	5	2
3	11	6	1	8	15	10
4	9	3	0	5	6	14

probability values translated from survey ratings in Interval 0 in order to ease comparison of PPT with survey results. In this case study, initial values in Table 8 are used for initializing PPT. Applying PPT to this case study, p (from the Preference Transition Model) and q (from the Utterance-Preference Model) converged to 0.530 and 0.920, respectively, after 23 iterations for the carafe and converged to 0.487 and 0.446 after 8 iterations for the filter. Note that the convergence rate depends on initial values chosen for p and q. When p and q are chosen from a uniform random distribution in a feasible range, the number of iterations required for convergence with a tolerance of 0.001 has the following statistics: 1) a mean of 25 iterations and a standard deviation of 10 iterations for the carafe and 2) a mean of 6.7 iterations with a standard deviation of 2.1 iterations for the filter. Thus, the EM algorithm converges in a reasonable number of iterations. The preferential probabilities calculated from Eqs. 6 and 7 using the converged p values and q values are shown in Table 5.

4.2.1 A note on certainty in design choice over time

In a typical design selection process, it might be expected that a choice starts out with relatively low certainty and becomes more certain toward the end of a project. Information entropy (Shannon 1948) is a measurement of the amount of uncertainty of an event associated with a given probability distribution. The lower the information entropy is, the more certain the design choice is. The entropy value of the selection of the most preferred alternative in the ith time interval can be represented with E_i as in Eq. 23.

$$E_i = -\sum_{k=1}^{N} P(\pi_i = a_k)\log_2 P(\pi_i = a_k) \qquad (23)$$

The entropy values for this case derived from PPT are also shown in Table 5 and show that entropy value does decrease over time, indicating increased certainty. Note that entropy values do not change until Time Interval 2 because that is when the design team began discussion about the filter.

4.3 Analysis of survey results

4.3.1 Modified Logit model results

The modified Logit model was applied to the survey results to obtain preferential probabilities. In order to select a reasonable range for λ, the two questions given in Sect. 3.2.1 were answered. For Question 1 regarding small differences in preferential probabilities between alternatives, the answer could be a probability of at most 0.55. For Question 2 about maximum differences, the answer could be a probability of at least 0.99. With Eq. 1 and 2, a possible range for this case study is [5, 20]. This section compares the preferential probabilities extracted from the transcript with the ones converted with Eq. 2 when λ is bounded. The preferential probabilities from the survey results converted using the modified Logit model are shown in Tables 6 and 7.

4.3.2 Simulation with PPS (maximum entropy)

PPS was applied to convert survey ratings into preferential probabilities. Since the ratings in the experiment were relative preference ratings that were then normalized, the joint distributions described in Sect. 3.1.2 were used to sample both the individual and group ratings for

Table 5 Group preferential probabilities from transcripts (PPT) and information entropy

Interval	Carafe				Filter			
	Glass	Steel	Plastic	Entropy	Gold tone	Paper	Titanium	Entropy
0	0.49	0.48	0.026	1.1	0.13	0.72	0.15	1.1
1	0.86	0.14	0.0066	0.63	0.13	0.72	0.15	1.1
2	0.96	0.040	0.0020	0.26	0.16	0.75	0.087	1.0
3	0.99	0.012	3.3E-05	0.094	0.0088	0.96	0.030	0.27
4	1.0	0.0021	3.5E-05	0.023	0.0031	0.0093	0.99	0.11

Table 6 Preferential probabilities using the modified Logit model ($\lambda = 5$) on surveys

Interval	Carafe			Filter		
	Glass	Steel	Plastic	Gold tone	Paper	Titanium
0	0.45	0.45	0.10	0.15	0.69	0.15
1	0.65	0.28	0.074	0.15	0.78	0.075
2	0.67	0.25	0.090	0.098	0.85	0.050
3	0.88	0.073	0.044	0.045	0.91	0.045
4	0.88	0.073	0.044	0.095	0.080	0.83

Table 7 Preferential probabilities using the modified Logit model ($\lambda = 20$) on surveys

Interval	Carafe			Filter		
	Glass	Steel	Plastic	Gold tone	Paper	Titanium
0	0.50	0.50	0.0012	0.0025	1.0	0.0025
1	0.97	0.034	0.00017	0.0013	1.0	8.8E-05
2	0.98	0.018	0.00033	0.00017	1.0	1.2E-05
3	1.0	4.5E-05	6.1E-06	6.1E-06	1.0	6.1E-06
4	1.0	4.5E-05	6.1E-06	0.00017	8.8E-05	1.0

simulation. Take, for example, one designer's ratings for three alternatives in a one-component selection problem. If the sampled ratings of the first and second alternatives were 0.2 and 0.5, then the sampled rating of the third would be 1-0.2-0.5 = 0.3. The sampled results would be screened out if the sum of the first two ratings was greater than 1 because it conflicted with the constraint that the sum of three sampled ratings was 1.

The sampled individual ratings were used to construct the group rating distribution, and then a group rating was sampled from the distribution. The weighted average was one of the constraints for solving the parameters for the distribution. The designers were interviewed after the design task and stated that they contributed to discussion almost equally. This was verified by reviewing the transcript and video-recording of the discussion. This meant equal weightings on the individual survey analysis could be applied. The resulting values for all 5 time intervals are shown in Table 8.

The survey results in Interval 0 show that both the glass carafe and the steel carafe have a $\sim$49% chance to be selected as the "best" or most preferred choice, while the plastic carafe has only a $\sim$3% chance to be selected as the "best" or most preferred choice. It can be inferred from results from Interval 0 to Interval 4, that, as a group, the glass and the stainless steel carafes were preferred in the beginning, but that only glass was preferred in the end. For the filter design, the design team preferred the paper throughout the session until the very end when the titanium filter became the most preferred choice.

4.4 Comparisons and discussions

Preferential probabilities obtained using PPT and from surveys via the modified Logit model and PPS were compared graphically and through geometric distance and correlation. Figures 3, 4, and 5 depict the preferential probabilities translated with PPS from surveys and those converted with PPT from transcript analysis for the carafe selection problem. Overall, they suggest that the glass carafe dominates over the other two alternatives (stainless

Table 8 Group preferential probabilities from surveys (PPS)

Interval	Carafe			Filter		
	Glass	Steel	Plastic	Gold tone	Paper	Titanium
0	0.49	0.48	0.026	0.13	0.72	0.15
1	0.65	0.32	0.023	0.16	0.82	0.024
2	0.68	0.27	0.048	0.098	0.89	0.017
3	0.88	0.10	0.015	0.032	0.93	0.039
4	0.89	0.10	0.015	0.10	0.069	0.83

steel carafe and plastic carafe) throughout design discussion. Figures 6, 7, and 8 show results for filter selection. They indicate that the team's preferential probability is highest for the paper filter until the last interval, in which the preferential probability was highest for the titanium filter.

The trends are fairly similar in Figs. 3, 4, 5, 6, 7, and 8 and are consistent with the qualitative reading of the transcript. The team changed their choice for the filter because new information was given in the design process that the glass carafe and the paper filter could not function together, and so the team had to select again. They changed the filter option because they agreed that the filter was less important than the carafe.

For a more quantitative assessment of consistency, the two measures proposed in Sect. 3.3 were applied to the data shown in Tables 5, 6, 7, and 8. In this case study, there are 30 data points with a total of 6 alternatives at 5 points in time. The initial values of PPT were assigned according to the survey results before the design task started so that the preferential probabilities are equal for both PPT and PPS at Time = 12:00 (Interval 0), and so data from Time Interval 0 are excluded in the calculation of the similarity measures. Since the sum of the preferential probability values for the three alternatives for each selection problem is fixed, the data for one of the three alternatives are excluded as well.

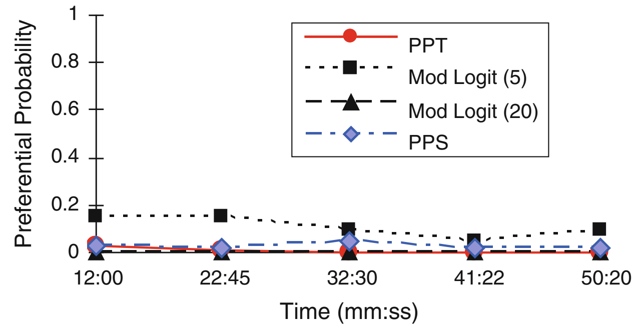

Fig. 5 Comparison of preferential probabilities for the plastic carafe

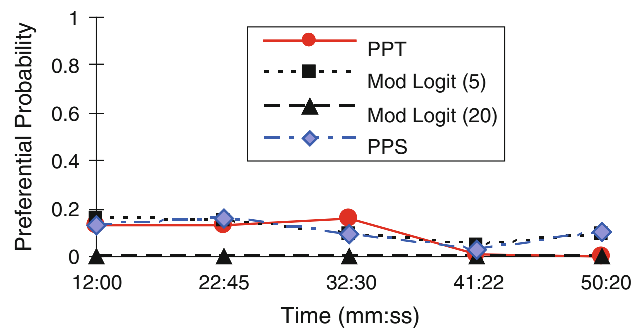

Fig. 6 Comparison of preferential probabilities for the gold tone filter

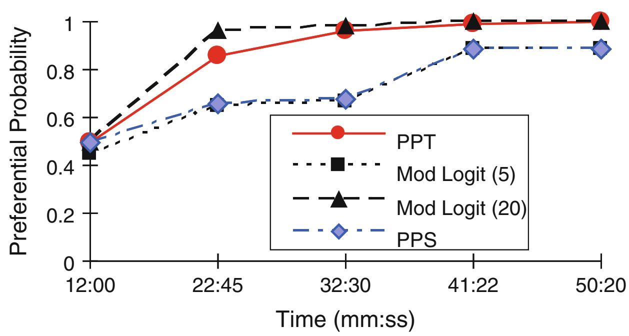

Fig. 3 Comparison of preferential probabilities for the glass carafe

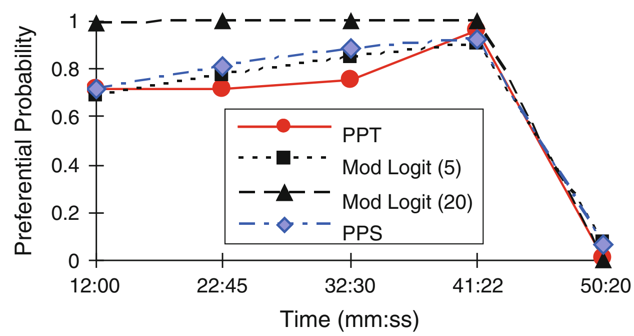

Fig. 7 Comparison of preferential probabilities for the paper filter

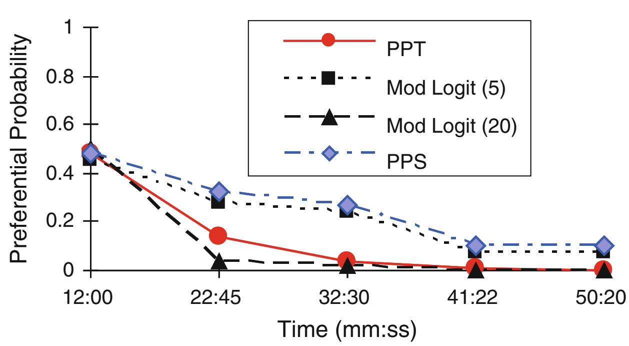

Fig. 4 Comparison of preferential probabilities for the stainless steel carafe

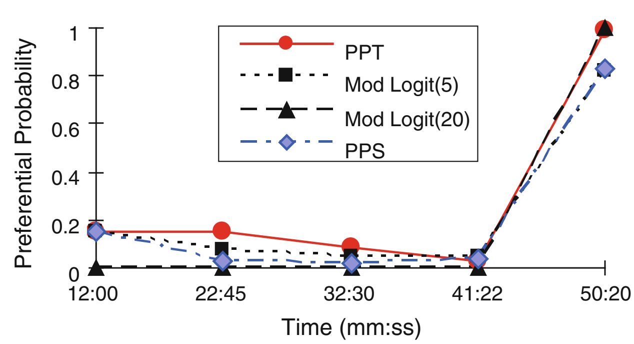

Fig. 8 Comparison of preferential probabilities for the titanium filter

Table 9 Comparisons between PPT and survey results

Measure	Possible ranges	Comparison results		
		PPT vs. Mod logit ($\lambda = 5$)	PPT vs. Mod logit ($\lambda = 20$)	PPT vs. PPS
Euclidean distance	$[0, \sqrt{14}]$	0.54	0.39	0.54
Pearson coefficient	$[-1, 1]$	0.96	0.98	0.96
		P value: 1.1E-7	*P* value: 6.1E-10	*P* value: 7.7E-9

This study uses the data from glass carafe, stainless steel carafe, gold tone filter, and paper filter for comparison. Therefore, each vector of preferential probabilities includes 16 points for comparison. The results of the Euclidean norm distance and Pearson coefficient are shown in Table 9.

Monte Carlo simulation shows that the average Euclidean distance and the standard deviation for two random feasible sequences of 16 preferential probabilities are $\bar{L}_2 = 1.3$ and $\sigma_{L_2} = 0.24$. $L_2 = 0.54$ is 3.2 multiples of standard deviation away from the average central point. Comparatively speaking, both norms of distances between PPT, the modified Logit model, and PPS are relatively small. Statistically significant correlation (Pearson coefficient) is very close to 1, implying a highly positive correlation between the vectors of the preferential probabilities from transcript and from surveys. It shows that, in this case, the evolution of the preferential probabilities from transcript is consistent with that of the surveys.

4.5 Limitations

The chief limitation of this work is its full validation. The case study presented is a small-scale design selection problem and is intended only to illustrate how the method may be applied. In order to truly demonstrate the effectiveness of the method, it must be evaluated through a full-scale set of experiments on a range of design selection problems involving longer duration discussions by teams with larger numbers of members.

It should also be noted that the PPT method is a descriptive approach to tracking variation of preference information over time. In its current form, PPT is not intended to predict a "correct" design choice, assuming that there is "right" answer to a design selection problem.

5 Conclusions

The initial research question posed in this paper was: Can a design team's preference-related information be determined in a way that does not require designers to explicitly state their preferences or involve aggregation of individual opinions? The method proposed in this paper suggests that the answer to this question is "yes." This paper presents a probabilistic approach to extraction that operates only on the transcripts of team discussion and does not require explicit input from the designer about his or her preferences. The approach treats a discussion as a "bag of words" that contains all of the words used by the designers and does not take into consideration what individuals say or how individuals should be aggregated. Finally, the approach provides a time-based representation of preference-related information, which can illustrate how design selection changes over time. This work may lead to a novel way to understand the evolving nature of a team's preferences over the life of a project.

The second research question posed in this paper was: How do the preferential probabilities obtained through transcript extraction compare with those obtained explicitly through more traditional means? The work in this paper establishes quantitative consistency between the preferential probabilities extracted from a transcript and those translated from surveys and suggests that PPT obtains a reasonable reflection of what a team is likely to prefer most. Both PPS and the modified Logit model provide evidence validating this. The PPS approach proposed in this paper turns to explicit surveys as a way to explicitly obtain preferences and applies the principle of maximum entropy to translate these into preferential probabilities. The consistent results of the small-scale case study experimentally suggest that the probabilistic approach proposed in this paper and previous work (Ji et al. 2007) is worth further investigation.

6 Future work

Future work on techniques for extracting preference-related information from design team discussion may follow a number of paths:

Evaluation in other scenarios The work presented in this paper is preliminary, and a key task for future work will focus on validating these approaches with a larger, more realistic set of data that includes cases and scenarios that take into consideration a range of team sizes, backgrounds, and design task complexity. PPT can be applied to a transcript from any size discussion group, but there may be possible effects of group size on quantity and quality of discussion. PPT also requires that design alternatives be known a priori, which may be difficult in a highly complex design problem.

Team dynamics The interplay among characteristics of team members and group leaders is complex, and there is a rich body of research on the topic. If additional detailed information on the design team, its members and the nature of the interaction between them is known, appropriate weightings may be formulated that can better represent overall team preferences. Such information might include knowledge of a team member's interpersonal skills or technical background, or their past decision-making making patterns in both groups and as individuals.

Other forms of preference information In future research, more information such as individual weightings and group dynamics patterns will be used to build even more realistic distribution models for PPS. Future research may also consider approaches to convert preferential probabilities extracted with PPT into generic preference ratings which are more widely used in formal design.

Time variation of preferences In design selection, it is expected that the design preferences may vary throughout the design process, and this was found to be true for both the surveys and the transcripts in the experiment. The probabilistic approaches described in this paper demonstrate the likelihood that a design team will prefer one alternative over the others and may lead to a novel way to understand the nature of a team's preferences over time.

Linguistic information This study directly associates design choices with the utterances of design alternatives in a discussion. However, the unambiguous identification of a design concept or alternative in text is an open area of research in linguistics. Lexical analysis of the related words, such as synonyms, antonyms, hypernyms, hyponyms, meronyms, holonyms, and troponyms (Miller et al. 1990; Dong et al. 2005) could improve the accuracy of design alternative identification in a transcript. Finally, appraisal analysis (Dong 2006a, b) in a discussion may also reflect designers' preferences. Preliminary work in integrating appraisal analysis and lexical relationships has already been conducted to improve the accuracy of the models for PPT in the future (Honda et al. 2010).

Acknowledgments The work described in this paper was supported in part by the National Science Foundation under Awards CMMI-0830134 and CMMI-0900255. The opinions, findings, conclusions, and recommendations expressed are those of the authors and do not necessarily reflect the views of the sponsors.

References

Arnold K (2001) Making team decision. In: Biech E (ed) The Pfeiffer book of successful team building tools. Jossey-Bass/Pfeiffer, San Francisco

Arrow KJ (1970) Social choice and individual values. Yale University Press, New Haven

Arrow KJ, Raynaud H (1986) Social choice and multicriterion decision-making. MIT, Cambridge, MA

Ben-Akiva M, Lerman SR (1985) Discrete choice analysis. MIT, Cambridge, MA

Bertrand M, Mullainathan S (2001) Do people mean what they say? Implications for subjective survey data. Am Econ Rev 91(2):67–72

Bilmes JA (1998) A gentle tutorial of the EM algorithm and its application to parameter estimation for Gaussian mixture and hidden Markov models. Technical report. International Computer Science Institute, Berkeley, CA, USA

Bockenholt U (2002) A Thurstonian analysis of preference change. J Math Psychol 46(3):300–314

Brans JP, Vincke P (1985) A preference ranking organisation method: (The PROMETHEE method for multiple criteria decision-making). Manag Sci 31(6):647–656

Brockman JB (1996) Evaluation of student design processes. The 26th annual frontiers in education conference, Salt Lake City, UT

Busemeyer JR, Diederich A (2002) Survey of decision field theory. Math Soc Sci 43(3):345–370

Cross N, Christiaans H, Dorst K (1996) Analysing design activity. Wiley, Chichester

Dempster A, Laird N, Rubin D (1977) Maximum likelihood from incomplete data via the EM algorithm. J R Stat Soc Ser B 39(1):1–38

Dong A (2005) The latent semantic approach to studying design team communication. Des Stud 26(5):445–461

Dong A (2006a) Concept formation as knowledge accumulation: a computational linguistics study. Artif Intell Eng Des Anal Manuf 20(1):35–53

Dong A (2006b) How am I doing? The language of appraisal in design. Design computing and cognition '06 (DCC06). J S Gero. Kluwer, Dordrecht, The Netherlands, pp 385–404

Dym CL, Wood WH, Scott MJ (2002) Rank ordering engineering designs: pairwise comparison charts and Borda counts. Res Eng Des 13(4):236–242

Fishburn PC (1978) Choice probabilities and choice functions. J Math Psychol 18:205–219

Fisher RA (1922) On the mathematical foundations of theoretical statistics. Philos Trans R Soc 222:309–368

Geslin MM (2006) An argumentation-based approach to negotiation in collaborative engineering design. Department of Aerospace And Mechanical Engineering, University of Southern California, Los Angeles

Giffin M, De Weck O, Bounova G, Keller R, Eckert C, Clarkson PJ (2009) Change propagation analysis in complex technical systems. J Mech Des 131(8):081001

Gigone D, Hastie R (1997) The impact of information on small group choice. J Pers Soc Psychol 72(1):132–140

Green PE, Srinivasan V (1990) Conjoint analysis in marketing: new developments with implications for research and practice. J Market 54(4):3–19

Grefenstette G (1993) Automatic thesaurus generation from raw text using knowledge-poor techniques. The 9th annual conference of the UW centre for the new OED and text research, Oxford, England

Hanley N, Mourato S, Wright RE (2001) Choice modelling approaches: A superior alternative for environmental valuation? J Econ Surv 15(3):435–462
Hauser JR, Clausing D (1988) The house of quality. Harv Bus Rev 66(3):63–73
Hazelrigg GA (1998) A framework for decision-based engineering design. J Mech Des 120(4):653–658
Hensher DA, Johnson LW (1981) Applied discrete choice modeling. Halsted Press, New York
Hey JD (1998) Do rational people make mistakes? Game theory, experience, rationality. In: Leinfellner W, Kohler E (eds) Kluwer, The Netherlands, pp 55–66
Honda T, Yang MC, Dong A, Ji H (2010) A comparison of formal methods for evaluating the language of preference in engineering design. ASME design engineering technical conferences. Montreal, Canada
Jabeur K, Martel J-M, Khelifa SB (2004) A distance-based collective preorder integrating the relative importance of the group's members. Group Decis Negot 13(4):327–349
Jain VK, Sobek DK II (2006) Linking design process to customer satisfaction through virtual design of experiments. Res Eng Des 17(2):59–71
Jaynes ET (1957) Information theory and statistical mechanics. Phys Rev 106(4):620–630
Jaynes ET (1968) Prior probabilities. IEEE Trans Syst Sci Cybern 4(3):227–241
Ji H, Yang MC, Honda T (2007) A probabilistic approach for extracting design preferences from design team discussion. In: Proceedings of ASME 2007 international design engineering technical conferences and computers and information in engineering conference
Keeney RL, Raiffa H (1976) Decisions with multiple objectives: preferences and value tradeoffs. Wiley, New York
Kelley CT (2003) Solving nonlinear equations with Newton's method. Society for Industrial and Applied Mathematics, Philadelphia
Kohrs A, Merialdo B (2000) Using category-based collaborative filtering in the ActiveWebMuseum. The 2000 IEEE international conference on multimedia and expo, vol 1, pp 351–354
Krantz DH, Luce RD, Suppes P, Tversky A (1971) Foundations of measurement volume 1. Academic Press, New York, NY
Kulok M, Lewis K (2005) Preference consistency in multiattribute decision making. ASME conference proceedings 2005 (4742Xa), pp 291–300
Li W, Jin Y (2006) Fuzzy preference evaluation for hierarchical co-evolutionary design concept generation. ASME conference proceedings (4255X), pp 31–41
Luce RD (1959) Individual choice behavior. Wiley, New York
Mabogunje A, Leifer LJ (1996) 210-NP: measuring the mechanical engineering design process. Frontiers in Education Conference. FIE '96. In: Proceedings of 26th annual conference, vol 3, pp 1322–1328
Manski CF (1977) The structure of random utility models. Theory Decis 8:229–254
Miller GA, Beckwith R, Fellbaum C, Gross D, Miller K (1990) WordNet: an on-line lexical database. Int J Lexicogr 3(4):235–244
Otto KN, Antonsson EK (1991) Trade-off strategies in engineering design. Res Eng Des 3(2):87–104
Otto KN, Antonsson EK (1993) The method of imprecision compared to utility theory for design selection problems. ASME 1993 design theory and methodology conference
Packard DJ (1979) Preference relations. J Math Psychol 19(3):295–306
Press WH, Teukolsky SA, Vetterling WT, Flannery BP (2007) Numerical recipes: the art of scientific computing. Cambridge University Press, New York
Pugh S (1991) Total design: integrated methods for successful product engineering. Addison-Wesley, Wokingham
Reich Y (2010) My method is better!. Res Eng Des 21(3):137–142
Ross SM (2006) Simulation. Academic Press, Burlington, MA
Saaty TL (2000) Fundamentals of decision making and priority theory with the analytic hierarchy process. RWS Publications, Pittsburgh
Scott MJ, Antonsson EK (1998) Aggregation functions for engineering design trade-offs. Fuzzy Sets Syst 99(3):253–264
Scott MJ, Antonsson EK (1999) Arrow's theorem and engineering design decision making. Res Eng Des 11(4):218–228
Scott MJ, Antonsson EK (2005) Compensation and weights for trade-offs in engineering design: beyond the weighted sum. J Mech Des 127(6):1045–1055
See T-K, Lewis K (2006) A formal approach to handling conflicts in multiattribute group decision making. J Mech Des 128(4): 678–688
Shah JJ, Vargas-Hernandez N, Summers JD, Kulkarni S (2001) Collaborative sketching (C-Sketch)—an idea generation technique for engineering design. J Creat Behav 35(3):168–198
Shannon CE (1948) A mathematical theory of communication. Bell Syst Tech J 27(3):379–423, 623–656
Song S, Dong A, Agogino AM (2003) Time variation of design "story telling" in engineering design teams. In: Proceedings of the 14th international conference on engineering design (ICED 03), Stockholm, Sweden
Thompson LL (2003) Making the team: a guide for managers. Prentice Hall, Upper Saddle River
Thurston D (1991) A formal method for subjective design evaluation with multiple attributes. Res Eng Des 3(2):105–122
Tribus M (1969) Rational descriptions, decisions, and designs. Pergamon Press, New York
Ueda N, Nakano R (1998) Deterministic annealing EM algorithm. Neural Netw 11(2):271–282
Ueda N, Nakano R, Ghahramani Z, Hinton GE (2000) SMEM algorithm for mixture models. Neural Comput 12(9):2109–2128
von Neumann J, Morgenstern O (1947) Theory of games and economic behaviour. Princeton University Press, Princeton
Wang J (1997) A fuzzy outranking method for conceptual design evaluation. Int J Prod Res 35(4):995–1010
Wassenaar HJ, Chen W (2003) An approach to decision based design with discrete choice analysis for demand modeling. J Mech Des 125(3):490–497
Wassenaar HJ, Chen W, Cheng J, Sudjianto A (2005) Enhancing discrete choice demand modeling for decision-based design. J Mech Des 127(4):514–523
Wood KL, Antonsson EK (1989) Computations with imprecise parameters in engineering design: background and theory. ASME J Mech Transm Autom Des 111(4):616–625
Yang MC (2003) Concept generation and sketching: correlations with design outcome. ASME Conf Proc 37017b:829–834
Yang MC (2005) A study of prototypes, design activity, and design outcome. Des Stud 26(6):649–669
Yang MC, Ji H (2007) A text-based analysis approach to representing the design selection process. In: Proceedings of the 16th international conference on engineering design (ICED 07)
Yang MC, Wood WH, Cutkosky MR (2005) Design information retrieval: a thesauri-based approach for reuse of informal design information. Eng Comput 21(2):177–192