Object Representation in Multiattribute Choice

Sudeep Bhatia (bhatiasu@sas.upenn.edu)
University of Pennsylvania, Philadelphia, PA.

Neil Stewart (neil.stewart@warwick.ac.uk)
University of Warwick, Coventry, UK.

Abstract

We propose a theoretical framework for understanding how everyday choice objects are represented and how decisions involving these objects are made. Our framework combines insights regarding object and concept representation in semantic memory research with multiattribute choice rules proposed by scholars of decision making. We also outline computational techniques for using our framework to quantitatively predict naturalistic multiattribute choices. We test our approach in two-object and three-object forced choice experiments involving common books, movies, and foods. Despite using complex naturalistic stimuli, we find that our approach achieves high predictive accuracy rates, and is also able to provide a good account of decision time distributions.

Keywords: Multiattribute choice, Semantic memory, Naturalistic decision making, Judgment and decision making

Introduction

Most choices that people make on a day-to-day basis, from the books they read to the foods they eat, involve trading off attributes, so as to select the object whose attributes are overall the most desirable (Keeney & Raiffa, 1993). There is, however, a disconnect between the way in which multiattribute choices are currently studied, and the way in which these day-to-day choices are typically made. Most multiattribute choice experiments explicitly present choice objects and their attributes to participants in a matrix of numerical quantities (e.g. Figure 1a). Everyday decisions, in contrast, are not usually composed of objects with a small set of explicitly presented and quantified attributes. Rather the objects in these decisions are much richer and complex (e.g. Figure 1b). Decision makers do have knowledge about these objects and their attributes, but this knowledge is represented in the decision makers' minds after having been learnt through prior experience with the choice domain.

	Phone A	Phone B
Memory	16GB	64GB
Speed	2.5Ghz	1.5GHz
Screen Size	5.1"	5.7"

Figure 1a and b. Stimuli presentation in standard multiattribute choice experiments (left) and in Study 1 (right).

The divergence between the stylized stimuli used in current research and the complex multiattribute choices made in real-world settings is problematic. Choice processes and resulting behaviors depend greatly on the ways in which attributes and objects are presented (e.g. Kleinmuntz & Schkade, 1993) suggesting that real-world decisions, which seldom involve actual attribute-by-object matrices, may be different to the types of decisions observed in current experimental work. More importantly, by using artificial designs in which the attributes of objects are directly presented to decision makers, existing theoretical work has largely ignored the role of object representation. Storing, retrieving, and processing attribute information about the objects in a given choice problem is a pivotal part of the decision process, and a complete account of choice requires an approach that is able to specify the mechanisms involved at this stage in the decision, well as the relationship between these mechanisms and the final outcomes of the decision (see Bhatia, 2013 for a discussion).

This paper provides a theoretical framework capable of addressing these issues. It relies on insights in semantic memory research which suggest that low-dimensional attribute spaces are used to represent objects and concepts. For example, multi-dimensional scaling (Shepard, 1962) passes similarity ratings through a matrix decomposition algorithm, resulting in the recovery of a small number of latent attributes that best describe the structure of similarity for a given domain. Likewise, distributional models of semantic memory typically learn low-dimensional word representations through natural language. Some approaches, like latent semantic analysis, use singular value decomposition to perform dimensionality reduction on word-context occurrence matrices (Landauer & Dumais, 1997). Others use Bayesian statistics or convolution based associative memory, but also result in low-dimensional representations for words (see Jones et al., 2015).

We suggest that these insights extend to everyday multiattribute choice, so that decision makers can be seen as using the distribution of observable features across choice objects in the environment to uncover low-dimensional latent attributes for representing the objects. Furthermore, we propose that it is these latent attributes that are evaluated and aggregated during the decision process. For simplicity we suggest that the recovery of latent attributes can be approximated using singular value decomposition on the observable feature space (as in e.g. Landauer & Dumais, 1997), and that the evaluation of the latent attributes can be approximated with a linear model with decision weights for each latent attribute (as in e.g. Keeney & Raiffa, 1993).

We also propose computational techniques for uncovering the latent attribute representations of common choice objects. Particularly, keywords, tags, and other natural language descriptors for choice objects on internet websites, can be considered suitable proxies for the observable features of these objects. For a sufficiently rich online dataset, it is possible to train semantic models and learn the latent attribute representations for the objects in a choice environment, and subsequently examine peoples' choices between these objects.

Framework

Let us consider a choice domain with N total objects**.** Each of these objects has a set of observable features, and can be written as a vector of these features. If there are M total unique features in the environment, then each for object i we have $\boldsymbol{x_i} = (x_{i1}, x_{i2}, \dots x_{iM})$, with $x_{ij} = 1$ or $x_{ij} = 0$ based on whether or not feature j is present in object i. Singular value decomposition involves processing the matrix $\mathbf{X} = [\boldsymbol{x_1}, \boldsymbol{x_2}, \dots \boldsymbol{x_N}]$ to obtain $L << M$ latent attributes, corresponding to the L largest singular values of $\mathbf{X}$. Using these singular values, we can represent an object i as $\boldsymbol{z_i} = (z_{i1}, z_{i2}, \dots z_{iL})$, with z_{ij} corresponding to the association between the object and the j^{th} latent attribute. Note that M can be very large in many naturalistic choice domains, whereas L is typically much smaller.

The use of latent attributes for representing objects implies that our approach retains the multiattribute structure assumed by theoretical decision making research. Thus we can take common multiattribute decision rules and apply them very easily to latent attributes. We use a simple linear rule, which specifies a decision weight for each attribute and aggregates weighted attributes into a measure of utility for an object (Keeney & Raiffa, 1993). The object with the higher utility is the one that is most frequently chosen. In the context of the latent attribute structure outlined here, this involves specifying an L dimensional vector of weights $\boldsymbol{w} = (w_1, w_2, \dots w_L)$, and multiplying the latent attributes for an object i by these weights, so as to obtain the utility for the object $U_i = \boldsymbol{w} \cdot \boldsymbol{z_i}$. In order to permit random noise in the choice process we embed our utilities in the logit choice rule (Luce, 1959). In a two-object choice this specifies the probability of choosing an object i over another object i' as $\Pr[i \text{ chosen}] = e^{U_i})/(e^{U_i} + e^{U_{i'}}) = e^{\boldsymbol{w} \cdot \boldsymbol{z_i}}/(e^{\boldsymbol{w} \cdot \boldsymbol{z_i}} + e^{\boldsymbol{w} \cdot \boldsymbol{z_{i'}}})$. For the general case with N' choice objects we have $\Pr[i \text{ chosen}] = e^{U_i}/(\Sigma_{n = 1 .. N'}\, e^{U_n})$.

In order to test our approach and illustrate its applicability we first need to uncover the actual attribute representations that characterize common choice objects. In related domains, such representations are usually obtained by asking experimental participants to generate features that describe the meaning of a given word (e.g. McRae et al., 2005). However common choice domains are so vast (involving thousands of features for thousands of objects) that the experimental elicitation of these feature norms may not practical. Thus we suggest that user-generated keywords, tags, and other descriptors for common choice objects on online datasets can be seen capturing the observable features that best describe the various objects.

In this paper, we use three large online datasets: www.GoodReads.com, which contains user-generated bookshelves for thousands of books; www.IMDB.com, which contains user-generated keywords for thousands of popular movies; and www.AllRecipes.com which contains user-specified ingredients for thousands of dishes. We scrapped these websites in 2014, and for each website we attempted to obtain as much information (as many objects and associated features) as was technically feasible. We obtained a total of 372,186 unique shelves for 15,737 books for the www.GoodReads.com dataset, a total of 160,322 unique keywords for 44,971 movies for the www.IMDB.com dataset, and a total of 24,688 unique ingredients for 39,979 recipes for the www.AllRecipes.com dataset. Using these user-generated descriptors as our observable features, each of the N objects in each of the three datasets can be written as an M-dimensional feature vector $\boldsymbol{x_i} = (x_{i1}, x_{i2}, \dots x_{iM})$, with $x_{ij} = 1$ if object i (a book, a movie, or a food dish) has observable feature j (a keyword, a shelf, or an ingredient). A singular value decomposition on $\mathbf{X} = [\boldsymbol{x_1}, \boldsymbol{x_2}, \dots \boldsymbol{x_N}]$ can be subsequently performed to obtain L << M latent attributes for the datasets.

Study 1

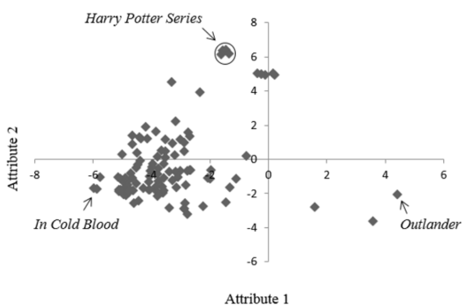
In Study 1 we tested whether our theoretical framework and the computational techniques for applying this framework, actually predict peoples' everyday multiattribute choices. This is the primary experiment in this paper: It involves incentivized choices in the laboratory with reaction time measures. In later studies we examine variants of this design using non-incentivized online samples.

Method. In this study, 73 participant made binary choices between pairs of popular books. Participants were recruited from a university subject pool, and performed the study in a behavioral laboratory on computer screens. Participants were also incentivized, and one of their chosen books was selected at random and given to them at the end of the study.

Unlike most existing multiattribute choice experiments, the choice objects were not presented alongside a set of quantifiable attributes (as in e.g. Figure 1a). Rather they were shown to participants using just the covers of the books and the accompanying titles (as in e.g. Figure 1b). Overall, each of the 73 participants made 220 choices involving 150 unique books. The books used in this study were obtained from 30 different popular genres on www.GoodReads.com.

Model Fitting. We fit participant choices using the latent attributes recovered from a singular value decomposition (SVD) on the www.GoodReads.com data. We allowed the number of underlying latent attributes, L, to vary across participants. For a given value of L, we used the L latent attributes with the highest singular values from the SVD on the www.GoodReads.com dataset. In order to ensure sufficient degrees of freedom for estimating decision weights, we restricted L to a maximum of L = 100 (and a minimum of L = 2). In essence this leads to a total of 99 unique models for each participant, corresponding to L = 2,

L = 3, … L = 100. with a separate set of best fitting participant-level attribute weights for each model. The values of the 150 books in our study on the two latent attributes with the largest singular values are shown in Figure 2. Figure 2 also shows the ten shelves with the largest absolute weights for these two latent attributes.



Latent Attribute 1 Shelf	Weight	Latent Attribute 2 Shelf	Weight
default	-0.11	*fantasy*	0.14
literature	-0.12	*teen*	0.13
2010	-0.12	*ya-fiction*	0.12
my-library	-0.10	*young-adult-fiction*	0.12
2009	-0.10	*ya*	0.11
series	0.10	*sci-fi-fantasy*	0.11
romance	0.10	*adventure*	0.11
part-of-a-series	0.09	*ya-books*	0.11
2011	-0.09	*magic*	0.11
book-club	-0.09	*fantasy-sci-fi*	0.11

Figure 2. The values of the 150 books in Study 1 on the two latent attributes with the largest singular values, alongside the ten shelves with the largest absolute weights for these two latent attributes.

In order to avoid overfitting, we used ten-fold cross-validation to test predictive accuracy and find the best performing model (i.e. best performing value of L) for describing each participant's choices. For model training, we recovered the weighting vector $\boldsymbol{w}$ that provided the best fit to the training data, with the assumption of a linear choice rule embedded in a logistic link function. This vector (whose dimensionality depended on the dimensionality of the model (value of L) in consideration), was recovered using maximum likelihood estimation. For model testing we calculated the proportion of choices in the test data predicted accurately by the recovered $\boldsymbol{w}$ for each model. A choice is considered to be predicted accurately if the utility assigned to the chosen option by the model in consideration is higher than the utility assigned to its competitor. Ultimately, the value of L and corresponding weight vector $\boldsymbol{w}$ with the highest accuracy on the test data was considered to be the overall best fitting model.

Results. The mean accuracy of our approach for predicting the test data is 83% (SD = 0.08), significantly above a baseline accuracy of 50% ($p < 0.01$). Additionally, the average best fitting value of L across our participants is 39.67 (SD = 27.95. Table 1 summarizes statistics regarding model accuracy.

	Latent Attributes	Random Vectors	Lex Heuristic	Tally Heuristic
Study 1				
Mean	0.83	0.69	0.50	0.72
Std. Dev.	0.08	0.08	0.03	0.09
Median	0.86	0.68	0.50	0.73
Best Fit	0.85	0.16	0.00	0.19
Significant	0.93	0.44	0.00	0.63
Study 2				
Mean	0.78	0.67	0.50	0.69
Std. Dev.	0.10	0.08	0.03	0.12
Median	0.80	0.65	0.50	0.70
Best Fit	0.74	0.18	0.00	0.21
Significant	0.88	0.46	0.00	0.59
Study 3				
Mean	0.74	0.51	0.33	0.53
Std. Dev.	0.13	0.07	0.01	0.12
Median	0.75	0.50	0.33	0.53
Best Fit	0.86	0.07	0.00	0.07
Significant	0.95	0.80	0.00	0.70
Study 4				
Mean	0.79	0.51	0.33	0.63
Std. Dev.	0.14	0.07	0.01	0.14
Median	0.80	0.50	0.33	0.65
Best Fit	0.86	0.09	0.00	0.08
Significant	0.96	0.73	0.00	0.88
Study 5				
Mean	0.80	0.51	0.33	0.62
Std. Dev.	0.12	0.07	0.01	0.13
Median	0.80	0.50	0.33	0.62
Best Fit	0.91	0.04	0.00	0.09
Significant	0.98	0.68	0.00	0.89

Table 1. Summary of model fits. Mean", "Std. Dev." and "Median" indicate the distribution of best-fitting model accuracy rates on test data across participants. "Best Fit" describes the proportion of participants for which the model has the highest accuracy (these proportions sum to greater than one as models are sometimes tied) and "Significant" indicates the proportion of participants that outperform the baseline model with $p < 0.05$.

One possibility is that our technique achieves its high accuracy rates by allowing flexible weights across a large number of dimensions. In order to control for this, we attempted the above model-fits with randomly generated attribute vectors. Particularly, for each participant and each object offered to the participant, we artificially created a 100-dimensional vector with each dimension randomly and uniformly distributed in the range [0,1]. We then performed a 10-fold cross validation procedure that examined the fits of linear models with flexible weights for L dimensions of the random vectors. With this approach we found the mean accuracy to be 69% (SD = 0.08) Additionally, 84% of participants achieved a higher accuracy rate using the recovered latent attributes from www.GoodReads.com, compared to the randomly generated vectors (and 8% of participants had equal accuracy with both approaches). A participant-level paired t-test indicates shows that this difference is significant ($p < 0.01$). Table 1 provides further statistics involving the random vectors approach.

Another alternative to our SVD-based attributes involves the use of the raw observable features for the books. Of course it is impossible to actually recover separate decision weights for each of these observable features. However, we can use well-known decision heuristics applied to these observable features. For example, using the lexicographic heuristic (Tversky, 1969) would involve considering only a single feature, and choosing the object that is the most desirable on this feature. Likewise, applying the tallying heuristic (Russo & Dosher, 1983) would involve counting up the positive and negative features of each choice object, and choosing the object with highest number of positive features relative to negative features. We applied these two heuristics to participant-level choice using 10-fold cross validation. For the lexicographic heuristic we used the training sample to determine which of the object's features has the highest absolute correlation with choice. We then used this single feature to predict the choices on our test sample. For the tallying heuristic, we used the training sample to determine whether each of the features were positively or negatively correlated with choice. If they were positively correlated with choice, they received a weight of +1, and if they were negatively correlated with choice they received a weight of -1. These weights were then applied to the observable features in the test data to predict choices according to the tallying heuristic.

We found that the lexicographic heuristic achieved a mean accuracy rate of exactly 50% ($SD = 0.03$), indicating that it is not a suitable way of making multiattribute choices with such large features spaces. In contrast, the tallying heuristic achieved a mean accuracy rate of 72% ($SD = 0.09$). When comparing these heuristics with our latent attribute approach, we found that all participants were better fit by our approach compared to the lexicographic heuristic, and that 78% of participants were better fit by our approach relative to the tallying heuristic (with another 16% tied). The differences in accuracy rates shown here are statistically significant when evaluated with a paired t-test ($p < 0.01$ for both heuristics). Table 1 provides further statistics involving the lexicographic and tallying heuristics.

How well do our model fits predict decision time? We can perform this test by embedding our best fitting utilities into a drift diffusion model (Ratcliff & Rouder, 1978). Our utilities are a measure of the desirability of the objects and, within the drift diffusion framework, are likely to determine the drift rate. We can formalize this by allowing the mean drift rate in the drift diffusion model to be a linear function of the best fitting utility difference. Thus, for trial a for participant b, we can write this mean drift rate as $v_{ab} = \beta_0 + \beta_1 \cdot (U_{ab}^L - U_{ab}^R)$. Here U_{ab}^L is the predicted utility for the left option in the trial for the participant, based on the best fitting model for the participant. Likewise, U_{ab}^R is the predicted utility for the right option. β_1 is a multiplier mapping this utility difference on to a drift rate, and β_0 is an intercept term capturing an absolute bias in drift for the left option. In this model, hitting the upper boundary leads to the left option being selected, whereas hitting the lower boundary leads to the right option being selected.

We fit this modified drift diffusion model permitting trial-to-trial variability in starting points and trial-to-trial variability drift rates. For this purpose, we adopted a hierarchical model fitting approach, as implemented by the HDDM toolbox (Wiecki et al., 2013). This approach recovers group mean parameters for the decision threshold, non-decision time, drift rates, trial-to-trial variability in starting points, trial-to-trial variability, and trial-to-trial variability drift rates, while also permitting individual differences in these parameters. Importantly this toolbox makes it easy to fit linear functions for drift rates as we wish to do in this paper. The best fitting group mean parameters from our specification, as recovered by the diffusion analysis, are presented in Table 2. Again β_1 represents the weight on utility difference in the drift term. As can be seen, the bulk of the distribution of this parameter lies above 0, indicating that the best fitting utility difference has a strong positive relationship with mean drift in the model. Table 2 also displays the deviance information criterion (DIC) value for this fits.

	Mean	SD	Median
Full model			
Boundary	3.26	0.09	3.26
Non decision time	0.46	0.02	0.46
β_0	0.01	0.01	0.01
β_1	0.26	0.01	0.26
DIC: 60,694.76			
Restricted Model			
Boundary	3.08	0.08	3.08
Non decision time	0.48	0.03	0.48
β_0	0.01	0.01	0.01
DIC: 68,571.64			

Table 2. Summary of best fitting group mean parameters for the drift diffusion model fits in Study 1. Here β_1 represents the weight on utility difference in the drift term, in the full model. The restricted model sets this to 0. DIC indicates the deviance information criterion value for the fits.

In a related analysis, we fitted a simplified version of this model in which $\beta_1 = 0$, and drift is independent of the predicted utility difference. As shown in Table 2, the fits for this model, measured through the deviance information criterion (DIC), are much lower than those for the extended model, suggesting that the utility differences specified by our approach do improve reaction time predictions in naturalistic multiattribute choice tasks.

Studies 2-5

As a secondary demonstration we applied our approach to two other domains: food choice and movie choice. We conducted a series of online studies offering participants two-object and three-object choices between various food dishes and between various movies, and we predicted these choices using latent attributes obtained from user-generated ingredients on www.AllRecipes.com and user-generated keywords on www.IMDB.com.

Method. In Study 2, 90 participants recruited from Amazon Mechanical Turk made 200 binary choices between various food dishes. The food dishes were obtained from www.AllRecipes.com, and there were a total of 100 unique food dishes used in the study (which were the most popular dishes on www.AllRecipies.com). Choices in this study were presented on the screen using just the names of the dishes. Participants had to click on the names in order to indicate their choices. In Study 3, 88 participants recruited from Amazon Mechanical Turk made 200 three-object choices between various food dishes. The dishes used were the same as those in Study 2, and their presentation was identical to that in Study 2 (except that each screen offered three different choices, instead of two). Participants in both Studies 2 and 3 were compensated with money.

In Study 4, 75 participants recruited from an undergraduate student participant pool made 200 three-object choices between different movies. There were a total of 100 unique movies used. These were the 100 most popular movies on www.IMDB.com (Internet Movie Data Base). The choices were presented on the computer screen using just the names of the movies and their IMDB movie posters. Participants had to click on the movie name or poster in order to indicate their choices. Participants were compensated with course credit. Study 5 was identical to Study 4, except that participants were recruited from Amazon Mechanical Turk. There were 223 total participants in this study, and they were compensated with a monetary payment.

Model Fitting. The model fitting in Study 2 was identical to Study 1, except that the latent attributes were recovered from a singular value decomposition on the www.AllRecipes.com data. Study 3 used a very similar model fitting technique, except that instead of a binary logit choice rule, there was a three-object (multinomial) logit choice rule. Studies 4 and 5 also used this choice rule, applied using latent attributes recovered from a singular value decomposition on the www.IMDB.com data.

Results. The accuracy rates from our analysis for the Studies 2-5 are displayed in Table 1. The mean accuracy for Study 2 is 78% (SD = 0.10), the mean accuracy for Study 3 is 74% (SD = 0.13), the mean accuracy for Study 4 is 79% (SD = 0.14) and the mean accuracy for Study 5 is 80% (SD = 0.12). All of these are significantly ($p < 0.01$) higher than the baseline accuracy of 50% (for Study 2) and 33% (for Studies 3-5).

We also found that the best fitting latent attribute models have a relatively low dimensionality, for most participants. Overall, the average best fitting value of L (i.e. number of dimensions) across our participants is 31.95 (SD = 28.55) for Study 2, 56.02 (SD = 27.18) for Study 3, 50.05 (SD = 28.12) for Study 4, and 52.64 (SD = 25.93) for Study 5.

Table 1 also displays the results of a random vector model for these studies. Again it shows that the majority of participants are better described by our approach relative to the random vector approach. Finally, Table 1 shows the fits of the lexicographic and tallying heuristics. For Study 2, these fits are performed similarly to Study 1. However, Studies 3-5 involve three object choice. Thus the weights for the individual features necessary for fitting these heuristics cannot be obtained through a simple correlation analysis between the relative presence or absence of a feature and the choice in a trial. Instead we calculated, for each feature in each trial, *Relative Presence* = $C - 0.5[UC_1 + UC_2]$. Here $C = 1$ if the feature is present in the chosen option and 0 otherwise. Likewise, $UC_1 = 1$ if the feature is present in the first unchosen option and 0 otherwise, and $UC_2 = 1$ if the feature is present in the second unchosen option and 0 otherwise. For each feature, we summed *Relative Presence* over all the observations in the training data for the participant in consideration. This gave us a measure of the *Total Relative Presence* of the feature in the chosen options for the participant. For the lexicographic heuristic, we then selected the single feature with the highest absolute *Total Relative Presence* for the participant in the training data, and used this feature to predict the participant's choices in the test data. For the tallying heuristic we recoded the *Total Relative Presence* for a feature to generate a weight of +1 if *Total Relative Presence* was positive and -1 if it was negative. These binary weights were then used to predict the participant's choices according to the tallying heuristic. Using this approach, we again found that the lexicographic and tallying heuristics were out performed by the latent attribute approach, as shown in Table 1.

Discussion

In this paper we have proposed that decision makers use low-dimensional latent attributes in order to make decisions in naturalistic multiattribute choice settings. We have obtained latent attribute representations for various everyday choice objects using user-generated object descriptors in large online datasets, and in five experiments, have predicted participant choices between these objects by fitting linear models with our latent attributes. Our fits reveal that our approach provides high accuracy rates, which significantly outperform accuracy rates obtained through other sophisticated methods (such as linear models with random attribute vectors, and lexicographic and tallying heuristics). The best fitting models in our analysis often have small or moderate number of dimensions. Additionally, these models are able to quantitatively predict decision times, when their estimated utilities are embedded within a drift diffusion process.

Our primary theoretical contribution involves the formal characterization of the processes involved in choosing between everyday choice objects. In doing so we extend insights from semantic memory research to the field of multiattribute decision making. The resulting framework attempts to describe all key aspects of the decision process, from the learning of object representations for common choice objects, to the use of these representations for evaluation and decision making. This is in contrast to most theories of multiattribute choice, which specify the mechanisms involved in aggregating decision attributes but seldom attempt to describe what these attributes actually are (see Bhatia, 2013 for a discussion).

Our results suggest that dimensionality reduction is not only at play in representing words, concepts, and various non-choice objects (as in e.g. Landauer & Dumais, 1997; Shepard, 1962) but is also a critical feature of multiattribute choice object representation in preferential decision making. There are many reasons why this would be the case. Firstly, common multiattribute choice objects involve a large number of observable features, as well as systematic relationships between the features. Good decision making involves understanding these feature relationships, and using these relationships to make inferences about the objects. Even though the inferences in preferential choice are primarily evaluative, knowledge is used in a very similar manner as in categorization, language comprehension, object recognition, and other related tasks. Additionally, the use of latent attributes also offers a number of distinct advantages relative to the use of raw observable features. There are fewer latent attributes than there are observable features, and for this reason, latent attributes simplify the decision process. These attributes also reduce redundancy in object representation, and do so in the most efficient manner possible. In fact, our approach is not unlike principle components regression, which possesses a very similar set of statistical benefits (see Draper & Smith, 1981).

That said, the approach presented in this paper is fairly simplistic: It involves a linear technique for dimensionality reduction combined with a linear multiattribute utility model. Both of these assumptions should be tested, it wouldn't be surprising if more sophisticated and more realistic approaches to building semantic representations (Jones et al., 2015) and making choices (Oppenheimer & Kelso, 2015) outperform the current approach. It may also be the case that the representations of choice objects depend not only on feature co-occurrence, but also on the reward structure of the domain in consideration. Individuals may, for example, learn object representations that best predict rewards, rather those that best predict feature occurrence. If this is the case then it would be necessary to train models of object representation alongside models of evaluation and choice (rather than training the former separately, as is done in this paper). This could be accomplished using neural networks with backpropagation from a preference (reward) layer to an object representation layer. Supervised topic models may also facilitate the learning of such representations.

Despite the need to test more sophisticated representation and choice models, the success of our current approach nonetheless opens up a new avenue for studying naturalistic multiattribute choice. It can be applied to examine whether existing multiattribute choice effects also emerge in more realistic choice settings, where attribute information is not presented numerically (as in Figure 1a). It can also be used to extend the psychological analysis of multiattribute choice beyond the laboratory and predict real world choice data. Ultimately, by combining existing theories of semantic representation and multiattribute choice with rigorous analysis of large-scale data, this paper has proposed tools to capture the large number of important decisions made in the real-world, that are not currently within the scope of decision making research. This has the potential to significantly expand the theoretical, descriptive, and practical scope of this area of study.

References

Bhatia, S. (2013). Associations and the accumulation of preference. *Psychological Review*, 120(3), 522.

Draper, N., and Smith, H. (1981), *Applied Regression Analysis*, New York: Wiley.

Jones, M. N., Willits, J. A., Dennis, S., & Jones, M. (2015). Models of semantic memory. *Oxford Handbook of Mathematical and Computational Psychology*, 232-254.

Keeney, R. L., & Raiffa, H. (1993). *Decisions with Multiple Objectives: Preferences and Value Trade-offs*. Cambridge University Press.

Kleinmuntz, D. N., & Schkade, D. A. (1993). Information displays and decision processes. *Psychological Science*, 4(4), 221-227.

Landauer, T. K., & Dumais, S. T. (1997). A solution to Plato's problem: The latent semantic analysis theory of acquisition, induction, and representation of knowledge. *Psychological Review*, 104(2), 211.

McRae, K., Cree, G. S., Seidenberg, M. S., & McNorgan, C. (2005). Semantic feature production norms for a large set of living and nonliving things. *Behavior Research Methods*, 37(4), 547-559.

Oppenheimer, D. M., & Kelso, E. (2015). Information processing as a paradigm for decision making. *Annual Review of Psychology*, 66, 2774.

Russo, J. E., & Dosher, B. A. (1983). Strategies for multiattribute binary choice. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 9(4), 676.

Shepard, R. N. (1962). The analysis of proximities: Multidimensional scaling with an unknown distance function. *Psychometrika*, 27(2), 125-140.

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

Wiecki, T. V., Sofer, I., & Frank, M. J. (2013). HDDM: hierarchical bayesian estimation of the drift-diffusion model in python. *Frontiers in Neuroinformatics*, 7, 14.