Naturalistic Multiattribute Choice

Sudeep Bhatia
University of Pennsylvania
Neil Stewart
University of Warwick

May 16, 2018

Address correspondence to Sudeep Bhatia, Department of Psychology, University of Pennsylvania, Philadelphia, PA. Email: bhatiasu@sas.upenn.edu. Funding for Sudeep Bhatia was received from the National Science Foundation grant SES-1626825. Neil Stewart was funded by Economic and Social Research Council grants ES/K002201/1 and ES/N018192/1 and Leverhulme Trust grant RP2012-V-022.

Bhatia, S., and Stewart, N. (in press). Naturalistic multiattribute choice. *Cognition.*

Abstract

We study how people evaluate and aggregate the attributes of naturalistic choice objects, such as movies and food items. Our approach applies theories of object representation in semantic memory research to large-scale crowd-sourced data, to recover multiattribute representations for common choice objects. We then use standard choice experiments to test the predictive power of various decision rules for weighting and aggregating these multiattribute representations. Our experiments yield three novel conclusions: 1. Existing multiattribute decision rules, applied to object representations trained on crowd-sourced data, predict participant choice behavior with a high degree of accuracy; 2. Contrary to prior work on multiattribute choice, weighted additive decision rules outperform heuristic rules in out-of-sample predictions; and 3. The best performing decision rules utilize rich object representations with a large number of underlying attributes. Our results have important implications for the study of multiattribute choice.

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

Introduction

Most choices that people make on a day-to-day basis can be seen as involving objects defined on two or more attribute dimensions. These choices involve trading off the relative values of the component attributes, so as to select the object whose attributes are, overall, the most desirable (Keeney & Raiffa, 1993). The study of these types of multiattribute choices is a key topic of inquiry across numerous fields, where scholars attempt to develop theories to predict individuals' multiattribute choices, as well as the relationship between these choices and various psychological, biological, and socio-economic variables (Hastie, 2001; Oppenheimer & Kelso, 2015; Weber & Johnson, 2009).

There is a disconnect between the way in which multiattribute choices are currently studied, and the way in which these choices are often made. Nearly all multiattribute choice experiments explicitly present choice objects and their attributes to participants in a matrix of numerical quantities (see Ettlin et al., 2015 for a summary). For example, participants may be given a choice between two hypothetical phones with each phone being described in terms of its memory, its processing speed, and its screen size. This choice would be shown in a simple 3x2 attribute-by-object matrix (e.g. Figure 1a). Although some consumer decisions do involve the evaluation of a small set of explicitly presented and quantified attributes, many other common decisions – involving, for example, movies to watch or food items to eat – do not. The objects in these common decisions may be listed using only their names (without any attribute information), but the underlying attribute structure is typically very rich and complex (e.g. Figure 1b). Decision makers do often have knowledge about these objects and their underlying attributes, but this knowledge is represented in the decision makers' minds after having been learnt through prior experience with the choice domain.

The divergence between the highly stylized stimuli used in current research and the complex multiattribute objects often involved in real-world settings is problematic. Choice processes and resulting behaviors depend greatly on the ways in which attributes and objects are presented. For example, altering attribute-by-object matrices, by displaying the objects separately rather than side-by-side, can reverse certain behavioral patterns (Bettman & Kakkar, 1977; Kleinmuntz & Schkade, 1993). Making some attributes more salient by altering the order in which they are displayed in the matrices can also have a powerful effect on behavior (Levav, Heitmann, Herrmann, & Iyengar, 2010; Russo, Medvec, & Meloy, 1996). Similarly, presenting information verbally instead of numerically can lead to different decision strategies and subsequently different choices (Stone & Schkade, 1991). There is also a well-documented difference between memory-based and stimuli-based decisions, and decision makers are known to use different choice processes when retrieving attribute information from memory vs. when using attribute information presented explicitly during the choice task (Lynch, Marmorstein & Weigold, 1988; Lynch & Srull, 1982; Rottenstreich, Sood & Brenner, 2007). This sensitivity to presentation and choice format suggests 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. Indeed, some scholars have suggested that multiattribute choice effects documented in the laboratory with artificial stimuli do not emerge in more naturalistic settings (see, e.g., Frederick, Lee, & Baskin, 2014).

The divergence between experimental research and naturalistic multiattribute choice also impedes theory development. 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, as well as the relationship between these mechanisms and the final outcomes of the decision. Of course a theory of object representation in multiattribute choice need not be completely novel: It can adopt existing insights regarding object and concept representation in semantic memory research, and combine these insights with common decision rules studied in multiattribute decision research. Such a theory would not only extend the descriptive scope of decision research, but would also help integrate two important areas of inquiry in psychology.

However, there is a significant methodological issue involved in studying multiattribute choice with naturalistic objects. Computational and mathematical theories of choice can make predictions and be tested only when underlying objects and attributes are quantified. However, unlike the attributes of object used in existing choice experiment (e.g., those in Figure 1a), the attribute of common choice objects (e.g., those in Figure 1b) are not directly observable. Although participants may know the underlying attributes of common choice objects, and use these attributes to make every-day multiattribute decisions, researchers do not currently have a way of uncovering and quantifying the precise attribute compositions of objects. Thus in addition to developing a theory of object representation in everyday multiattribute choice, it is also necessary to develop practical techniques to apply this theory to actual choice data obtained from experimental and field settings.

The goal of this paper is to address these theoretical and methodological challenges. We begin by examining how common choice objects can be represented. Here we build upon insights in semantic memory research, which suggest that people use latent attribute spaces for representing common non-choice objects and concepts (e.g., Landauer & Dumais, 1997;

Shepard, 1962). We argue that these insights can be extended to everyday multiattribute choice, with decision makers using the distribution of observable features across objects to obtain a large number of latent attributes for representing the choice objects in the environment. Furthermore, we propose that it is these latent attributes that are evaluated and aggregated during the decision process. This evaluation and aggregation can be modelled using the types of existing decision rules already used to describe choice behavior in decision making research (e.g. Gigerenzer & Gaissmaier, 2011; Keeney & Raiffa, 1993; Payne et al., 1993; Shah & Oppenheimer, 2008).

We also consider computational techniques for uncovering the latent attribute representations of common choice objects. We propose that crowd-sourced 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. To demonstrate this idea, we give experimental participants naturalistic choices between different movies (Studies 1 and 4) and between different foods (Studies 2 and 3). We attempt to predict these choices using multiattribute choice rules applied to latent attribute representations trained on crowd-sourced data from websites like www.IMDB.com and www.AllRecipes.com.

Object Representation

Imagine a choice between watching *Toy Story* and *Star Wars.* This choice does not only involve evaluative processes for comparing the two movies, but also semantic memory processes for representing the movies and knowing what the movies actually are. In order to understand how people may make these types of choices we need to study the cognitive basis of the mental representations of choice objects, as well as the ways they are integrated into evaluative choice processes during the decision.

Although the issue of representation is not often addressed in multiattribute decision research (but see Hastie, 2001 for a discussion), it has received much attention in others areas of cognitive psychology, particularly semantic memory research. The relevant object and concepts studied in this area are often described in terms of features that the objects possess (Estes, 1950; Garner, 1978; Smith & Medin, 1981; Tversky, 1972, 1977). The number of observable features possessed by a given object can be very large, making it difficult to manipulate and utilize feature-based representations. Thus individuals represent common objects and concepts using latent attributes, which they recover by performing a low-dimensional mapping on the observable feature space.

Consider, for example, a child exposed to different animals and plants (e.g., *robin*, *salmon*, *rose*), each with a different set of observable features (e.g., *wings*, *fins*, *thorns*). By examining the distributional structure of the features across objects, the child can uncover a set of latent dimensions (possibly resembling categories like *animal, fish, plant, flower*) that define this feature space. These dimensions, or attributes, can be then be used for a variety of cognitive tasks, including categorization, feature induction, object recognition, language use and comprehension, similarity judgment, as well as sophisticated reasoning and inference.

Such representations can be uncovered through techniques with varying statistical interpretations, and techniques applied to a diverse range of stimuli and training data. For example, multi-dimensional scaling (Shepard, 1962; 1980) passes pairs of similarity ratings through a matrix decomposition algorithm, resulting in the recovery of latent attributes that best describe the structure of similarity (i.e. featural proximity) for a given domain. Recently, Nosofsky and coauthors (Nosofsky, Sanders & McDaniel, 2017; Nosofsky, Sanders, Meagher & Douglas, 2017) have used multi-dimensional scaling with great success to uncover rich representations for natural categories.

Relatedly, distributional models of semantic memory learn word representations through natural language. Unlike multi-dimensional scaling these techniques do not rely on participant ratings of similarity, and can thus be applied on a very large scale to uncover representations for nearly any commonly used word. Some distributional approaches, like latent semantic analysis, perform dimensionality reduction using singular value decomposition on word-context occurrence statistics, in order to uncover multiattribute vector representations for words (Landauer & Dumais, 1997). Other approaches involve probabilistic topics (Griffiths, Steyvers, & Tenenbaum, 2007) and convolution based associative memory (Jones & Mewhort, 2007). These approaches also use latent attribute representations for words, but rely on slightly different techniques for training the representations (see Jones et al., 2015 for a recent overview).

The process of uncovering latent attributes from feature distributions can also be understood in the context of connectionist models. For example, Rumelhart and Todd's (1993) and Rogers and McClelland's (2006) model of semantic memory involves a feedforward neural network with hidden layers. These hidden layers, after training through back propagation, encode distributed representations for the various objects and concepts. When network nodes are linear, backpropagation has been shown to perform singular value decomposition (Saxe, McClelland, & Ganguli, 2013).

Latent Choice Attributes

It is likely that knowledge of choice objects, such as *Toy Story* and *Star Wars*, is learnt in a similar manner to knowledge about *robins*, *salmon*, and *roses*, and other non-choice objects commonly studied in semantic memory research. Thus we would expect the dimensionality reduction mechanisms outlined above to also play a role in choice object representation in everyday multiattribute decision making. For the purposes of this paper, we approximate these mechanisms with singular value decomposition. As discussed above, SVD is the primary assumption in approaches like latent semantic analysis, and is also a byproduct of backpropagation in linear hidden-layer neural networks.

More specifically 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 for each object i we have $\boldsymbol{x_i} = (x_{i1}, x_{i2}, \ldots x_{iM})$, with $x_{ij} = 1$ or $x_{ij} = 0$ based on whether or not feature j is present in object i. SVD involves decomposing the matrix $\mathbf{X} = [\boldsymbol{x_1}, \boldsymbol{x_2}, \ldots \boldsymbol{x_N}]$ 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 $\boldsymbol{z_i} = (z_{i1}, z_{i2}, \ldots z_{iL})$, 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 and evaluating 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 directly to latent attributes. For the purposes of this paper we consider six decision rules, which are very well studied in multiattribute choice research and are commonly used to describe behavior in multiattribute choice experiments (usually experiments involving explicit attribute-by-object matrix, as in Figure 1a) (Gigerenzer & Gaissmaier, 2011; Keeney & Raiffa, 1993; Payne et al., 1993; Shah & Oppenheimer, 2008). Many of these rules also have a linear or pseudo-linear structure giving them interpretable statistical properties and making them very easy to apply to complex object representations with a large number of attributes. Of course there are other cognitively compelling decision rules as well, and we discuss how such rules could be applied to describe behavior in our experimental task at the end of this paper.

The first rule that we examine in this paper is the weighted-additive decision rule (WAD) (Keeny & Raiffa, 1993). In the context of the latent attribute structure outlined here, this rule specifies an L dimensional vector of weights $\boldsymbol{w} = (w_1, w_2, \ldots w_L)$, and multiplies the latent attributes for an object i by these weights, so as to obtain the utility for the object $U_i = \boldsymbol{w} \cdot \boldsymbol{z_i}$. The object with the higher utility in a choice setting is chosen. As this decision rule involves assigning a separate weight to each attribute it is often seen as a rational algorithm for making multiattribute decisions.

Our second rule is the weighted pros heuristic (WP) (Huber, 1979). This is a variant of WAD that can only be applied to binary choice. WAD compares the attributes of objects against each other, so that for a pair of objects i and i' it calculates sign($\boldsymbol{z_i} - \boldsymbol{z_{i'}}$). Here we have sign($z_{ij} - z_{i'j}$) = +1 if $z_{ij} > z_{i'j}$, sign($z_{ij} - z_{i'j}$) = -1 if $z_{ij} < z_{i'j}$, and sign($z_{ij} - z_{i'j}$) = 0 if $z_{ij} = z_{i'j}$. After calculating sign($\boldsymbol{z_i} - \boldsymbol{z_{i'}}$), the WP rule aggregates attributes comparisons with a weighting vector $\boldsymbol{w}$, which allows each attribute weight to take on any positive or negative value. If $\boldsymbol{w} \cdot$ sign($\boldsymbol{z_i} - \boldsymbol{z_{i'}}$) > 0, WP chooses object i¸ whereas if $\boldsymbol{w} \cdot$ sign($\boldsymbol{z_i} - \boldsymbol{z_{i'}}$) < 0, WP chooses object i'. Intuitively, this heuristic allows weights to be flexible across attributes but considers only whether one object is better than the other on each attribute dimension (as in decision by sampling, Noguchi & Stewart, in press; Stewart, Brown, & Chater, 2006).

Our third rule is the equal weights heuristic (EW), which is also seen as a simplification of WAD (Dawes, 1979; Dawes & Corrigan, 1974). EW also specifies an L dimensional vector of weights $\boldsymbol{w} = (w_1, w_2, \ldots w_L)$, and multiplies the latent attributes for an object i by these weights, so as to obtain the utility for the object $U_i = \boldsymbol{w} \cdot \boldsymbol{z_i}$. The object with the highest utility is chosen. However, unlike WAD, the weights in consideration are equal to either +1 or -1, corresponding to whether the attribute in consideration is desirable or undesirable. In essence this rule gives each attribute in the decision equal importance.

The fourth rule we consider is the tallying heuristic (TAL) (also known as the majority of confirming dimensions heuristic) which further simplifies EW by utilizing a binary representation for each attribute (Russo & Dosher, 1983). Thus prior to being multiplied by $\boldsymbol{w} = (w_1, w_2, \ldots w_L)$ (where $w_j = +1$ or $w_j = -1$ for each j), TAL first recodes the attributes to obtain sign($\boldsymbol{z_i}$). Here we have sign(z_{ij}) = +1, sign(z_{ij}) = -1, or sign(z_{ij}) = 0, depending on whether $z_{ij} > 0$, $z_{ij} < 0$, or $z_{ij} = 0$. Ultimately TAL can also be seen as computing a utility $U_i = \boldsymbol{w} \cdot \text{sign}(\boldsymbol{z_i})$ and choosing the object with the higher utility. Intuitively, this utility corresponds to the total number of good vs. bad attributes in the object.

Our fifth rule is the lexicographic heuristic (LEX) (Fishburn, 1974). This is the simplest multiattribute decision rule: It places all the decision weight on a single dimension. Thus it this rule can also be seen as calculating a utility for each object $U_i = \boldsymbol{w} \cdot \boldsymbol{z_i}$. However here we have $w_j = 1$ or -1 for only one j, and $w_j = 0$ for all other j. In essence this heuristic chooses the object with the highest (or lowest) value on a single attribute.

Our final rule is the fast and frugal tree (FFT) (Martignon, Vitouch, Takezawa, & Forster, 2003; Phillips, Neth, Woike & Gaissmaier, 2017). Unlike the other decision rules, the FFT does not involve the weighted additive utility-based evaluation of the objects. Rather FFT considers a small number of attributes sequentially, with a decision being made if the comparison on a given attribute satisfies an exit condition. This exit condition is typically of the form of a difference on the attribute, making the FFT very similar to an extension of the lexicographic heuristic known as the lexicographic semi-order heuristic (Tversky, 1969). For example, FFT may involve the following procedure for making decisions between two movies: If Movie 1 is sufficiently greater than Movie 2 on Attribute 4, choose Movie 1, otherwise if Movie 1 is sufficiently less on Attribute 7 choose Movie 2, otherwise if Movie 1 is sufficiently greater on Attribute 2 choose Movie 1, otherwise choose Movie 2.

With this framework we can now specify how choices, such as those between *Toy Story* and *Star Wars*, are made. Prior to the choice, the decision maker would have had sufficient experience with the choice domain (in our example, movies) so as to build latent attribute representations of some dimensionality *L* for the various objects in the domain. When presented with a set of available objects at the time of choice (*Toy Story* or *Star Wars*), our framework predicts that the decision maker would first map the objects onto their latent attributes. These multiattribute representations would then be aggregated into choices using one of the decision rules we have specified above. In the case of WAD, each attribute of the two movies would be assigned a separate continuous weight and the decision maker would choose the object with the higher weighted additive utility; in the case of WP, the objects would be compared against each other on every attribute dimension, and these binary comparisons would be aggregated with continuous weights to determine the object with the highest weighted pros or cons; in the case of EW the decision maker would use equal weights (+1 for desirable latent attributes or -1 for undesirable latent attributes) so as to give each attribute the same importance when computing utilities; for TAL each latent attribute would be transformed into a binary representation prior to aggregation, so as to tally the total number of good vs. bad attributes; for LEX only a single latent attribute would be used to calculate the relative desirabilities of the objects, and the object with the highest or lowest value on this attribute would be chosen; and for FFT, the decision maker would go through attributes sequentially, and choose one of the two objects if an exit condition on the attribute in consideration is satisfied.

Computational Methods

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 they consider important in describing the meaning of a given word (e.g., McRae, Cree, Seidenberg, & McNorgan, 2005; Vinson & Vigliocco, 2008). 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 be practical. Fortunately, there are now numerous crowd-sourced online datasets with detailed user-generated keywords, tags, and other descriptors for common choice objects. These descriptors are rich and comprehensive, and easy to obtain, and can be seen capturing the observable features that best describe the various objects, according to individuals who have had experience with the objects. For example, www.IMDB.com (the Internet Movie Database) allows readers to describe each movie using keywords. These keywords capture, amongst other things, the movies' key plot elements, themes, settings, aesthetics, character types, intended audience, and literary inspirations. Using all keywords used to describe all the movies on www.IMDB.com, we can build an approximate feature-based characterization of the movie universe, and represent any movie in terms of the vector of features used to describe it. We can then perform a SVD on these features to obtain latent attributes, and apply the techniques outlined above to predict choices such as those between *Toy Story* and *Star Wars*.

In this paper, we use two large online datasets: 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 160,322 unique keywords (along with actor, actress, and director names) 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 two datasets can be written as an M-dimensional feature vector $\boldsymbol{x_i} = (x_{i1}, x_{i2}, \ldots x_{iM})$, with $x_{ij} = 1$ if object i (a movie or a food dish) has observable feature j (a keyword or an ingredient), otherwise 0. An SVD on $\mathbf{X} = [\boldsymbol{x_1}, \boldsymbol{x_2}, \ldots \boldsymbol{x_N}]$ can be subsequently performed to obtain $L <<$ M latent attributes for the movies or foods in these datasets.

Now in our experiments we will be offering participants naturalistic choices between various movies (Studies 1 and 4) and between various food dishes (Studies 2 and 3). These choices will be presented in a manner similar to Figure 1b, that is, with only the names of the items of the choice items (though in Study 4 we will consider an even more naturalistic presentation format for movies). We will fit our decision rules to choices on an individual level to test the applicability of these decision rules, applied to recovered latent attributes, in predicting choice behavior.

We will allow the number of underlying latent attributes, L, to vary across participants for all heuristics except for LEX and FFT (which involve the use of only a limited number of attributes). For a given value of L, we will use the L latent attributes with the highest singular values from the SVD on the corresponding dataset. Thus for $L = 2$ we will use the two latent attributes with the largest singular values. For this *L*, each of the movies or food items in our experiment will be represented by a two-dimensional vector, and we will fit the WAD, WP, EW and TAL decision rules to participant choice-level data assuming this underlying two-dimensional representation. Likewise, for $L = 3$ we will use the first two latent attributes, as well as the latent attribute with the third largest singular value, to represent the choice items. For this *L*, we will fit our WAD, WP, EW and TAL decision rules assuming this three-dimensional representation. In order to ensure sufficient degrees of freedom for estimating decision weights, we will restrict *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 the WAD, WP, EW and TAL decision rules for each participant, corresponding to $L = 2$, $L = 3$, … $L = 100$, with a separate set of best fitting participant level parameters for each model. Note that our use of only the *L* largest singular values implies that our latent attribute representations contain the most amount of feature information possible for that level of dimensionality.

There are two interrelated issues in comparing the absolute and relative predictions of our decision rules. The first pertains to the fact that most of these decision rules can have a very large number of parameters for large values of *L*. For example, for $L = 100$, the WAD, TAL, WP, and EW decision rules have 100 parameters each. This gives our fits considerable flexibility and makes them vulnerable to overfitting. The second issue is that the number of free parameters for each decision rule does not correspond to the flexibility of the rule. For example, WAD, TAL, WP, and EW have a total of 100 flexible weights when $L = 100$, but different rules allow these weights to take on different values: WAD and WP permit any positive or negative weights for the attributes; whereas EW and TAL permit only +1 and -1 as attribute weights. As the number of parameters does not capture model flexibility, model comparisons using metrics such as the

Akaike Information Criterion and the Bayes Information Criterion are inapplicable. The same issue also holds for decision rules like FFT, which uses only a small set of attributes, but aggregates these attributes in a more sophisticated manner than the other weight-based decision rules.

In order to avoid these problems, we will use ten-fold cross-validation to test predictive accuracy and find the best performing decision rule, with the best weights and the best performing value of *L*, for describing each participant's choices. This involves randomly dividing each participant's data into ten portions, using the first nine portions (training data) to train the models, and the tenth portion (test data) to evaluate the models. Cross validation ensures that we avoid overfitting, and measuring predictive accuracy on out-of-sample predictions sidesteps the issue of measuring model flexibility with the total number of free parameters. Browne (2000) provides an introduction to cross-validation methods.

For training the WAD and WP rules (which allow the attribute weights to take on any values) we will embed the predicted utility differences into a logistic function, and use maximum likelihood estimation on the training data to uncover the best weights for each value of *L* for each participant. For training the EW and TAL heuristics (which restrict attribute weights to +1 or -1), we will individually test each of the attributes to the see whether the attribute values (continuous values in the case of EW and binary values in the case of TAL) correlate positively or negatively with choice on the training data. This indicates whether each of the attributes are desirable or undesirable, so that, when predicting the test data, we can give desirable attributes a weight of +1 and undesirable attributes a weight of -1. Again EW and TAL will be fit separately for each value of *L* for each participant.

We will restrict $L = 100$ for LEX and FFT, as these heuristics involve the selection of a small subset of attributes for use in the choice process. LEX uses only a single attribute, and this attribute will be determined based on a method similar to EW and TAL. Particularly, we will determine how correlated each of the 100 attributes are with choice on the training data, and then select the attribute with the highest absolute correlation, giving this attribute a weight of +1 if the correlation is positive and -1 if it is negative; all other attributes will be given a weight of 0.

For FFT we will use an established FFT fitting toolbox introduced by Phillips et al. (2017). This toolbox allows for FFT rules with different “depths” (i.e. different total number of attributes used in the final model), however we will only present results for an FFT model with a depth of three attributes. This three-depth model is the default model in Phillips et al.’s toolbox. We have tested the FFT model with depths of up to 10 attributes, and have found very little changes in accuracy as depth is increased. Note that our fits for FFT use all 100 latent attributes, and find the model that best predicts choices. This is done using the “ifan” algorithm, which tries many attribute orderings and thresholds for attribute differences to construct a good sequence of attributes and set of exit conditions.

For model testing we will calculate the proportion of choices in the test data predicted accurately by each decision rule for each participant. A choice is considered to be predicted accurately if the model selects the object that was chosen by the participant. In the (very few) cases where a model is unable to make a prediction (e.g. if it assigns both objects the same utility) we assume that the model makes the choice randomly. For decision rules that permit varying values of L, we will use the value of L that provides the best average accuracy in predictions on the test data.

Note that high accuracy rates obtained from the above techniques may not be due to realistic latent attribute representations for the choice objects. They may merely be a product of model fits that allow for flexible weights across a large number of dimensions. Of course cross-validation does control for this, but there is still the possibility that decision weights fit on a randomly generated attributes for each object (weakly) capture relational preferences for the objects, and thus predict behavior with an above chance accuracy.

In order to control for this, we will replicate the fits for the WAD rule with randomly generated attributes, instead of the latent attributes obtained from an SVD on our crowd-sourced data. Particularly, for each participant and each object offered to the participant, we will artificially create a 100-dimensional vector with each dimension randomly and uniformly distributed in the range [0,1]. We will then perform a 10-fold cross validation procedure that examines the fits of the weighted additive rule with flexible weights for L dimensions of the random vectors. High accuracy for this WAD-RAND rule would indicate that predictive accuracy stems not from the realism of the recovered latent attribute structure but rather from the flexibility inherent in utilizing a large number of different attribute dimensions to fit choice. In contrast, low relative accuracy for WAD-RAND would suggest that this is not the main cause of successful fits, and that latent attributes do appropriately capture underlying object representations.

To ensure that differences between WAD and WAD-RAND are not due to differences in the distribution of the attributes used to fit the two models, we will also consider a variant of WAD-RAND, called WAD-SCRM. This variant uses actual attribute vectors for each choice option, but scrambles the indices in each trial prior to model fitting. Thus the actual attribute values are meaningless, but the distribution of the attributes is the same as for the attributes used to fit WAD. Again, as with WAD-RAND, we will test WAD-SCRM with 10-fold cross validation and flexible weights for the *L* dimensions of the random vectors.

Properties and Predictions

Our use of latent attributes to predict choice implies our approach presents a number of novel benefits relative to previous applications of multiattribute decision rules. First, there is the issue of tractability. Often, there are a very large number of observable features in given choice domain. For example, as discussed above, the space of movies that we examine in this paper contains 160,322 unique keywords, or features, for 44,971 movies. Applying our multiattribute decision rules to these features, rather than latent attributes, would involve learning and applying hundreds of thousands of different tradeoffs. In contrast, our framework, by reducing the dimensionality of the underlying attribute space, facilitates a less arduous choice process.

Latent attributes are also useful for generalizing preferences. As they encode statistical regularities in the choice environment, they can be used to infer the desirability of novel objects with novel features which haven't been explicitly evaluated previously. Thus decision makers who know that they like a movie with wizards can infer that they would like a movie with goblins, if wizards and goblins both map on to the same desirable latent attribute. This would be the case even if they had never explicitly evaluated (i.e., learnt decision weights) for goblins. Of course this ability to generalize across the feature and object space is exactly what makes latent attributes particularly useful in non-preferential cognitive tasks (such as categorization, feature induction, language comprehension, etc.).

The use of latent attributes also reduces redundancy. Many observable features are highly correlated, and specifying separate decision weights for these features is unnecessary. SVD, and related techniques, map correlated features onto the same latent attributes, reducing the need for separate weights for correlated features.

Also note that the WAD decision rule applied to latent attributes uncovered through SVD can also be understood in the context of principle components regression, a technique that involves regressing the dependent variable on the main principle components of the explanatory variables. Principle components analysis and singular value decomposition are very closely related, implying that many of the properties of principle components regression also extend to our framework. These include properties pertaining to multi-collinearity and redundancy (discussed in the paragraphs above), as well as properties involving efficiency and optimality: The estimators obtained using principle components regressions have a lower mean squared errors than estimators from standard linear regressions, and are the optimal estimators (in terms of minimizing prediction error) for a large class of regularized estimators (Draper & Smith, 1981). In fact, Davis-Stober, Dana, and Budescu (2010) have recently proposed a linear model of heuristic judgment utilizing principle components, which they have shown displays similar desirable statistical properties.

Even though our use of latent attributes is relatively novel, the fact that we aggregate these attributes with existing multiattribute decision rules implies that we can use many of the findings in multiattribute choice research to make predictions in our experiments. For example, in prior work, the WAD rule has been shown to be outperformed by heuristics like EW, in making out-of-sample predictions (Dawes, 1979; Dawes & Corrigan, 1974; also see Gigerenzer & Gaissmaier, 2011 for a discussion). As WAD involves inferring separate weights for each attribute, it is sensitive to overfitting, and its predictions are not as robust as those made by simpler weighting schemes. Our model tests also use out-of-sample predictions to evaluate model accuracy. Indeed, as our underlying attribute space is far larger than that used in typical multiattribute choice experiments, fits on this attribute space should be even more vulnerable to robustness and overfitting problems. For this reason, we would expect heuristic decision rules like EW to significantly outperform WAD in terms of accuracy on our test data.

In addition to comparing the decision rules against each other we will also examine the number of latent attributes used by the best fitting variants of each of the rules. Now, prior work on multiattribute choice has suggested that decision makers do not use that many attributes while making decisions (see Shah & Oppenheimer, 2008 for an extensive discussion and review). However, this work has relied on choice tasks with explicitly presented numerical attribute values, in the types of object-by-attribute matrices shown in Figure 1a. It is possible that the number of attributes used by decision makers is significantly larger when these attributes correspond to latent representations stored in memory. Particularly, unlike exogenously provided attribute values, which are likely processed consciously and sequentially, latent representations can be retrieved and aggregated in parallel, without considerable cognitive oversight. Additionally, a large number of such latent attribute representations are necessary to fully describe complex choice objects like movies and food items. Indeed, applications of SVD to research on semantic memory, have shown that such models require hundreds of latent attributes to best mimic human judgments regarding naturalistic objects and concepts (Landauer & Dumais, 1997). Thus although prior research on multiattribute choice suggests that people only use a few attribute values in decision tasks, we predict that the types of naturalistic choices studied in this paper are best predicted by models that use a much larger number of latent attributes.

Study 1

In Study 1 we tested whether our theoretical framework and the computational techniques for applying this framework, predict peoples' everyday multiattribute choices. In this study we considered choices between movies, and we obtained latent attribute representations for these movies using user-generated keywords on www.IMDB.com.

Method

In this study, 92 participants were recruited from the Prolific Academic website. Participants made 200 binary choices between pairs of popular movies. In each choice they were asked to select the movie that they would prefer to watch. There were a total of 100 unique movies used. These were the 100 most popular movies on www.IMDB.com. All subjects were given the same set of choice pairs, and the choice pairs used in the experiment were determined by randomly combining the 100 items with each other. The choices were presented on a computer screen using just the names of the movies, as in Figure 1b. After making the choices participants were shown the list of 100 movies and, for each movie, were asked to indicate whether they had previously seen the movie, and, if not, whether they recognized the movie.

We attempted to predict participant choices using latent attributes obtained from www.IMDB.com. There are a total of 160,322 unique keywords for 44,971 movies in this dataset. The values of the 100 movies in Study 1 on the two latent attributes with the largest singular values are shown in Figure 2. Figure 2 also shows the ten movie keywords with the largest absolute weights for these two latent attributes. These keywords suggest that the first latent attribute corresponds to dark and violent movies whereas the second latent attribute corresponds to movies that are not action movies (note that all the keywords here have negative weights). Indeed, as indicated in Figure 2, the movie with the highest value on attribute 1 is *The*

Departed, the movie with the highest value on Attribute 2 is *Eternal Sunshine of the Spotless Mind*, and the movie with the lowest value on Attribute 2 is *The Dark Knight Rises.*

Results

Predictive accuracy. The accuracy rates from our analysis for the participants in Study 1 are displayed in Table 1. This table also shows the Cohen's κ measure for this statistic, calculated as $\kappa = (A–R)/(1\text{-}R)$, where A is the accuracy of the model and R is the accuracy of a random choice rule (50% in binary choice).

Contrary to our predictions we found that the best performing decision rule was WAD with a mean accuracy of 78% across participants ($\kappa = 0.56$). This was followed by EW, WP, TAL, FFT, and LEX. All decision rules outperformed the random choice rule ($p < 0.001$ for all when evaluated using a paired t-test). Although the first three of these heuristics performed almost identically, obtaining mean accuracy rates of around 70% ($\kappa = 0.40$), FFT and LEX performed much worse, obtaining an accuracy of only 61.37% ($\kappa = 0.22$) and 55% ($\kappa = 0.10$) respectively. Overall, WAD had the highest accuracy rates for 90.21% of participants, and comparisons between each pair of decision rules using a paired t-test on the participant level showed that WAD significantly outperformed all five other heuristics ($p < 0.001$ for each of the four comparisons). Scatter plots of the relative accuracies of WAD compared to each of the five heuristics, for each of our participants, is show in Figures 3a-e.

How about WAD-RAND and WAD-SCRM? Recall that these decision rules use randomly generated attribute values for each movie or scrambled attribute values for each movie, and then fit choices in the same manner as WAD to find best-performing attribute weights for these random or scrambled attributes. We found that such rules do not perform well at all. WAD-

RAND and WAD-SCRM only achieved accuracy rates of 56% (κ = 0.12) and 57% (κ = 0.13) respectively (p < 0.001 for both comparisons with WAD).

Let us also examine the dimensionality of our decision rules. We found that the average best generalizing value of L (i.e. number of dimensions) for the best fitting WAD, WP, EW and TAL decision rules across our participants was 74. Additionally, as shown in Table 1, all of these rules used a large number of latent attributes, with participant averages in excess of 60 latent attributes. In Figure 4 we can see the average accuracy of these rules as a function of L. This figure shows that the average accuracy (aggregated across participants) is increasing in L in a concave manner for all heuristics. Thus more latent attributes improve model predictions in aggregate, but with diminishing returns with each additional attribute dimension. It is likely that aggregate model performance could have been further improved if we had considered more than 100 latent attribute dimensions.

Recognition. The above tests examine predictive accuracies for our decision heuristics using all of the choice problems answered by the participants. The implicit assumption here is that participants know about each of the movies and are thus able to appropriately aggregate the latent attributes of the movies to make their decisions. This is a reasonable assumption, as our tests use the 100 most popular movies on www.IMDB.com, which are some of the best-known movies in the United States. However, a more rigorous variant of our tests would restrict the analysis to only the movies that are recognized by participants.

Overall we found that the proportion of choice trials in which both movies were recognized was 84% across participants. Thus it does seem that most of the decisions studied above are decisions in which participant have previous knowledge about the movies they are given. Moreover, when we restricted the above tests to only choice trials in which participants recognized both the movies that they were offered, we found that the above pattern of results was unchanged. Again we were able to achieve very high accuracy rates across participants, with the best performing decision rule being the WAD rule with a large number of latent attributes. These results are summarized in Table 2.

Our recognition data is also useful for testing the predictive power of the recognition heuristic. This heuristic predicts that a recognized movie is chosen over an unrecognized movie in choices in which only one movie is recognized. A similar heuristic has been shown to guide behavior in a wide range of judgment and decision tasks (Gigerenzer & Goldstein, 1996; Goldstein & Gigerenzer, 2002; Hertwig et al., 2008), and it is possible that participants also used recognition to make their decisions in our experiment. Now, on average, the recognition heuristic was only applicable in 26 out of the 200 trials for each participant (participants either recognized both movies or neither of movies in the remaining trials). For trials on which the recognition heuristic could be applied, we found that this heuristic obtained an average predictive accuracy rate of 84% ($\kappa = 0.68$) across participants (though note that for 18 participants we could not apply the recognition heuristic at all, as they recognized all movies). This accuracy rate is significantly different to a random accuracy rate of 50% ($t(74) = 22.56$, $p < 0.001$), providing strong evidence in favor of the recognition heuristic in naturalistic movie choices.

Finally note that a similar analysis is also possible when considering movies that have previously been seen by participants. In choices in which one movie had been seen and the other movie hadn't we found that participants chose the seen movie only 20% of the time on average. Thus, unsurprisingly, it seems that participants used an "unseen" heuristic to make choices when such a heuristic was applicable, indicating a preference for novelty.

Attribute weights. As a robustness test of our approach it is useful to examine the best fitting attribute weights for decision rules like WAD. Different individuals do have different tastes for movies, and we would expect these differences to show up in our fits to the latent attributes. If they don't -that is, if our fits reveal similar attribute weights for all participants- this would indicate a problem in our theoretical assumptions or methodological approach.

In order to test for heterogeneity in attribute weights we thus fit the WAD rule for each participant using only the first two latent attributes for the choice objects. Additionally, as we were not concerned with measuring out-of-sample predictions, we fit this decision rule on the entirety of each participant's data (i.e. on all 200 observations). A scatter plot of the attribute weights from this model fitting exercise is shown in Figure 5. Recall that the two attributes in this figure correspond to the two attributes in Figure 2.

As can be seen in Figure 5, there is significant variance in the weights for the two attributes across participants. The average weight on attribute 1 is -0.05 and the average weight on attribute 2 is 0.07 across participants. The correlation between the weights for these two attributes is -0.10 which is not significantly different to 0 (p = 0.35). Overall, this figure suggests that different participants do have different preferences over the latent attributes, as would be expected.

Discussion

Study 1 tested whether our proposed approach was able to describe naturalistic choices between pairs of movies. The study involved a choice experiment in which the available movies were presented using just their names. The choices of participants were predicted by applying one of six different decision rules to latent attributes recovered through a singular value decomposition on the www.IMDB.com dataset. We evaluated model predictions using ten-fold cross validation (performed on the participant level).

Overall, the most accurate decision rule was the WAD rule, which achieved an average accuracy rate of 78% and provided the best predictions for 90% of participants. This rule also outperformed WAD-RAND and WAD-SCRM which apply WAD to randomly generated or randomly scrambled latent attributes. An examination of the dimensionality of the best fitting WAD decision rules across participants, revealed that they used a large number of latent attributes.

These results were unchanged when restricting our analysis to only the movies that were recognized by participants. However, in settings in which only one of the two movies were recognized, the recognition heuristic did describe participant choices with a high accuracy. Finally, model fits using WAD did reveal substantial heterogeneity in participant attribute weights, as would be expected in such naturalistic choices.

These results provide a number of novel insights regarding naturalistic multiattribute choice. First, they conclusively show that it is possible to predict naturalistic choices with high accuracy rates using existing decision rules applied to latent attributes. Moreover, these latent attributes do correspond to the types of representations used by decision makers, as replacing these attributes with random vectors leads to a significant decline in predictive power. Additionally, these results show that, contrary to our predictions, decision heuristics such as EW do not outperform rules like WAD, even when evaluating these decision rules using out-of-sample accuracy rates. Finally, the best performing decision rules always utilize a large number of latent attributes, and decision rules that do not use a large number of latent attributes, such as FFT and LEX, have relatively low accuracy rates. This is compatible with prior work which finds that a large number of latent dimensions are necessary for predicting behavior in semantic memory tasks (e.g. Landaur & Dumais, 1997; see also Jones et al., 2015 for a discussion).

Studies 2 and 3

In Studies 2 and 3 we wished to test for the robustness of the above results. For this purpose, we applied our approach to a second domain: food choice. We conducted two studies offering participants two-object and three-object choices between various food dishes, and we predicted these choices using latent attributes obtained from user-generated ingredients on www.AllRecipes.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. These were the most popular dishes across the different cuisines on www.AllRecipies.com. Choices in this study were presented on the screen using just the names of the dishes, as in Study 2. All participants were given the same set of choices, which were generated by randomly selecting pairs of food dishes from the list of 100 food dishes.

In Study 3, 75 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). Again all participants were given the same set of choices.

Participant data were fit using the latent attributes recovered from a singular value decomposition (SVD) on the www.AllRecipes.com data. There are 24,688 unique ingredients for 39,979 recipes in the www.AllRecipes.com data. The values of the 100 food dishes used in our study on the two latent attributes with the largest singular values are shown in Figure 6. Figure 6 also shows the ten ingredients with the largest absolute weights for these two latent attributes. These ingredients suggest that the first latent attribute corresponds to baked sweet food dishes whereas the second latent attribute corresponds to savory dishes with an onions and garlic base. Indeed, as indicated in Figure 6, the food dish with the highest value on attribute 1 is *Iced Pumpkin Cookies*, whereas the food dish with the highest value on attribute 2 is *Turkey and Quinoa Meatloaf.*

For Study 2 we considered the six decision rules used in Study 1: WAD, WP, EW, TAL, FFT, and LEX, as well as WAD-RAND and WAD-SCRM. All of these rules, except for WP and FFT (which apply only to binary choice) were also fit in Study 3. As in Study 1, model fits were evaluated using 10-fold cross validation. Accuracy rates were calculated using the average proportion of choices in the test sample predicted correctly, across 100 training-test data splits. The baseline accuracy, achieved by a random choice rule, in such an analysis is 50% for the two-object choices in Study 2, and 33% for the three-object choices in Study 3.

Results

Study 2. The accuracy rates for Study 2 are displayed in Table 3. As in Study 1, we found that the best performing decision rule was WAD with a mean accuracy of 72% across participants ($\kappa = 0.44$). This was followed by WP, EW, and TAL, which performed almost identically, with accuracy rates around 66% ($\kappa = 0.32$). FFT and LEX were again much worse, reaching an accuracy of only 59% ($\kappa = 0.16$) and 53% ($\kappa = 0.06$). All decision rules outperformed the random choice rule ($p < 0.001$ for all). Overall, WAD performed the best for 79% of participants, and comparisons between each pair of rules using a paired t-test on the participant level showed that WAD significantly outperformed all five other heuristics ($p < 0.001$

for each of the four comparisons). A scatter plot of the relative of accuracy of WAD compared to each of these heuristics, for each of our participants, is shown in Figures 7a-e.

Additionally, as in Study 1, we found that WAD-RAND and WAD-SCRM achieved a much lower accuracy rate than WAD: An average of only 57% ($\kappa = 0.14$) across participants for both ($p < 0.001$ relative to WAD). Finally, we found that the average best generalizing value of L (i.e. number of dimensions) for the best fitting WAD, WP, TAL and EW decision rules was 75. As shown in Table 3, all of these rules used a large number of latent attributes, with averages in excess of 50 latent attributes across participants. In Figure 8 we can see the average accuracy of these rules as a function of L. This figure shows that the average accuracy (aggregated across participants) is increasing in L for all heuristics.

Study 3. The accuracy rates for Study 3 are displayed in Table 4. Once again, the best performing decision rule was WAD with a mean accuracy of 67% ($\kappa = 0.51$) across participants. This was followed by TAL, EW, and LEX (note that we did not fit FFT and WP to this experiment as they apply only to binary choice). All decision rules outperformed the random choice rule ($p < 0.001$ for all). Overall, WAD performed the best for 85% of participants, and comparison between each pair of rules using a paired t-test on the participant level showed that WAD significantly outperformed all three other heuristics ($p < 0.001$ for each of the four comparisons). A scatter plot of the relative of accuracy of WAD compared to each of the three heuristics, for each of our participants, is shown in Figures 9a-c.

Once again WAD-RAND and WAD-SCRM achieved a much lower accuracy rate: An average of only 39% ($\kappa = 0.09$) and 40% ($\kappa = 0.10$) across participants. For this reason, WAD-RAND and WAD-SCRM performed significantly worse than WAD ($p < 0.001$ for both comparisons). Finally, we found that the average best generalizing value of L (i.e. number of dimensions) for the WAD, TAL and EW decision rules across our participants was 67. Again, all decision rules that permitted multiple attributes used a large number of latent attributes. In Figure 10 we can see the average accuracy of these rules as a function of *L*. This figure shows that the average accuracy (aggregated across participants) is increasing in *L* for all heuristics

Discussion

Despite using a different choice domain, Studies 2 and 3 fully replicated the results of Study 1. Again, we were able to predict out-of-sample choices with a very high degree of accuracy. The best performing decision rule was WAD, which used a large number of latent attributes. Additionally, the latent attributes played a critical role in predicting choices, as the WAD-RAND and WAD-SCRM rules performed very badly in both studies.

Study 4

Studies 1, 2, and 3, show that it is possible to predict choices presented in a naturalistic format (i.e. with only object names, and without an explicit attribute-by-object matrix of numerical quantities). However, arguably the types of multiattribute choices made by decision makers in the real world involve more than just object names. For example, movie choices on websites like www.netflix.com are often presented with movie posters and movie synopses. Do the results documented above emerge with this alternate (arguably richer and more realistic) type of presentation? More precisely, can we use existing decision rules applied to latent attributes to predict choice when choice presentation also involves additional visual and textual information? We tested this in Study 4, which involved choices between movies. As in Study 1, latent attributes for the movies were obtained from www.IMDB.com, however unlike Study 1, we used three-object choices, and presented each choice with a movie poster and a one-line synopsis.

Method

In Study 4, 88 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 the website www.IMDB.com (Internet Movie Data Base) that were also used in Study 1. Each participant was given the same set of movie choices, which were constructed randomly from our set of 100 movies. The choices were presented on the computer screen using the names of the movies, their movie posters, and a one-line synopsis (with movie posters and synopses also obtained from IMDB). All movies were presented on the screen at the same time.

Results

The accuracy rates for Study 4 are displayed in Table 5. Once again, the best performing decision rule was WAD with a mean accuracy of 64% ($\kappa = 0.46$) across participants. This was followed by TAL, EW, and LEX (note that we did not fit FFT and WP to this experiment they apply only to binary choice). All decision rules outperformed the random choice rule, which generates an accuracy rate of 33.33% ($p < 0.001$ for all). Overall, WAD performed the best for 88% of participants, and comparisons between each pair of rules using a paired t-test on the participant level showed that WAD significantly outperformed all three other heuristics ($p < 0.001$ for each of the four comparisons). A scatter plot of the relative of accuracy of WAD compared to each of the three heuristics, for each of our participants, is shown in Figures 11a-c.

Once again WAD-RAND and WAD-SCRM achieved a much lower accuracy rate: An average of only 41% ($\kappa = 0.11$) and 42% ($\kappa = 0.13$) across participants. For this reason, WAD-RAND and WAD-SCRM performed significantly worse than WAD ($p < 0.001$ for both comparisons). Additionally, all decision rules that permitted multiple attributes used a large number of latent attributes. This is shown in Table 4. In Figure 12 we can see the average accuracy of these rules as a function of *L*. This figure shows that the average accuracy (aggregated across participants) is increasing in *L* for all heuristics.

Discussion

The results of Study 4 once again replicate the results of Study 1. Despite utilizing a richer and more complex presentation format (with movie posters and synopses) we were able to predict out-of-sample choices with a very high degree of accuracy. The best performing decision rule was once again WAD with a large number of latent attributes. This rule also outperformed WAD-RAND and WAD-SRAM (showing that the latent attributes played a critical role in predicting choice).

General Discussion

Novel Insights

Accuracy. In this paper we have attempted to model naturalistic multiattribute choice. We have obtained latent attribute representations for various everyday choice objects using user-generated object descriptors in large online datasets and, in four studies, have predicted participant choices between these objects by applying existing multiattribute decision rules to the latent attribute representations. Our fits achieve high accuracy rates, which outperform accuracy rates expected by chance and accuracy rates obtained from decision rules applied to randomly generated attributes. These high accuracy rates are comparable to those in multiattribute choice studies using very tightly controlled stimuli with numerical attributes (e.g., Figure 1a). For example, Berkowitsch, Scheibehenne & Rieskamp (2014) find that logit choice models achieve accuracy rates of 73% when describing undergraduate students' choices between triplets of cameras (with five explicitly presented attributes). In Study 4, which presents analogous data involving undergraduate students' choices between triplets of movies, we are able to achieve an average accuracy rate of 67%.

Weighted additive rule. We have found that the best performing decision rule is the weighted additive rule (WAD), which makes the most accurate predictions for a large majority of participants in our four studies. In contrast, decision heuristics, such as the weighted pros heuristic (WP), equal weights heuristic (EW), tallying heuristic (TAL), fast and frugal tree heuristic (FFT), and lexicographic heuristic (LEX) all do significantly worse than WAD (though these heuristics do outperform random choice). The superior fits of WAD relative to these heuristics is contrary to prior work on multiattribute choice, which finds that simple heuristics are often better for out-of-sample predictions (Dawes, 1979; Dawes & Corrigan, 1974; also see Gigerenzer & Gaissmaier, 2011 for a discussion).

Our results suggest that our participants may be using a weighted additive multiattribute decision rule in naturalistic choice settings. These are choice settings that people frequently encounter in the real world and are thus settings in which people have sufficient experience. Decision heuristics, which simplify the evaluation and aggregation of attributes, may become less useful with practice, and thus may fail to provide an appropriate account of naturalistic choice. Relatedly, it is also possible that the difficulty of evaluating and aggregating latent attributes, which are learnt through experience and stored in memory, may be less than the difficulty of aggregating and evaluating externally provided numerical attribute values. For this reason, decision makers may not need to rely on heuristics in naturalistic choice as much as they would in the types of settings typically examined in multiattribute choice research.

Many attributes. We have also found that the best fitting decision rules all use a very large number of attributes. This too is inconsistent with prior work on multiattribute choice, which suggests that decision makers rely on only a few attributes to make their decisions (see Shah & Oppenheimer, 2008 for an extensive discussion and review). However, the use of a large number of latent attributes is fully consistent with applications of similar techniques in semantic memory tasks involving everyday objects and concepts (in fact the optimal dimensionality in such settings has been argued to be around 300 – much larger than that which we permit in this paper) (Landauer & Dumais, 1997). Everyday objects are highly complex and require rich mental representations: It would be nearly impossible to capture the space of movies and food items (and subsequently preferences over movies and food items) using only two or three underlying attribute dimensions.

The divergence we have observed between our results and prior results on multiattribute choice is not surprising. Multiattribute choice is heavily task dependent and decision makers are known to use different strategies in different choice settings. However, this task dependence does suggest that the study of naturalistic multiattribute choice may involve different slightly decision processes than those considered to be at play in the types of choice tasks examined in prior work.

Object representation. One key theoretical contribution of this paper 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 object representations that are evoked, to the use of these representations for evaluation and choice. This is in contrast to most theories of multiattribute choice, which specify the mechanisms involved in aggregating decision attributes but seldom attempt to describe how these attributes are represented in the minds of decision makers.

The success of our approach in predicting naturalistic choice behaviors suggest that latent attribute representations are not only at play in cognitive tasks involving words, concepts, and various non-choice objects but are also a critical feature of preferential decision making. There are many reasons why this would be the case. First, 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. In additional, 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, as outlined earlier in the paper, the WAD rule applied to latent attributes resembles principle components regression, which possesses a very similar set of statistical benefits.

Limitations and Extensions

Alternate interpretations. It is important to note that our high predictive accuracy rates do not allow us to make conclusive claims about underlying cognitive processes. It could be the case that participants use decision rules applied to some other mental representations. Our latent attribute-based fits may merely learn to mimic these choice processes, thereby generating accurate predictions. Unlike existing paradigms in experimental decision making research, we do not explicitly provide choice attributes to participants. Although this is a desirable feature of our research, it also implies that that we are unable to restrict or control what participants know about the available choice objects.

Despite this limitation, we believe that the use of radically different representations by participants is unlikely. There is extensive research on semantic memory that supports the type of analysis performed in this paper. This research is supported by various psychological measures, including participant responses in similarity judgment tasks and free association tasks (again see Jones et al. 2015 for a review). In future work, it would be useful to combine such semantic tasks with multiattribute decision making, to jointly model both the semantic and the evaluative components of the decision.

Note that it could also be the case that decision makers are using the latent attributes from our analysis, but are doing so in a manner that differs from the assumptions made in this paper. For example, it could be the case that decision maker further project the latent attributes onto a smaller representational space (perhaps using a non-linear transformation). They then could aggregate these reduced representations either with the WAD rule, or with various heuristics. Such alternative models could be easily tested using the approach outlined in this paper, and this is a useful topic for future work.

Beyond singular value decomposition. Another extension to this paper could involve more sophisticated techniques than singular value decomposition, for uncovering latent attributes. For example, there are approaches in semantic memory research that utilize probabilistic topics rather than linear matrix decompositions (e.g., Griffiths et al., 2007). Such approaches are likely to provide more intuitive latent attributes (that map on more directly to the attributes used by decision makers), and also give a rational interpretation to the latent attribute-based decision.

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 (Blei & Mcauliffe, 2008) may also facilitate the learning of such representations. Relatedly, it would be useful to test for the effect of underlying reward structure on the use of object representations in non-choice tasks such as categorization and similarity judgment.

Participant heterogeneity in representation. In many ways, reward-dependent object representations would lead to participant level heterogeneity in attribute structures. Individuals who like a certain type of choice object may also have more fine grained representations of that choice object. More generally, different individuals are likely to vary in their attribute representations due to variation in experience and personal history. This type of heterogeneity has been ignored it this paper. This has been due primarily to practical concerns: Currently it is very difficult to uncover all the objects in a given choice domain that a participant has encountered before (and it is even harder to uncover the idiosyncratic rewards the participant has associated with the objects). Thus, for tractability, we assume that all our participants have the same latent attribute representations, which are recoverable by analyzing large online datasets. Future work should attempt to move beyond this assumption, and use richer, individual-specific sources of data to uncover individual-specific choice object representations. A promising approach to this problem has been provided by Sanborn, Griffiths & Shiffrin (2010) and such techniques could be applied to the choice domains considered in this paper.

Domain-general representations. It would also be useful to uncover representations that are not restricted to a particular choice domain. The models used in this paper are specifically trained for movies or foods and cannot be applied outside of these two areas. In contrast, the ideal model would have representations for an immense number of choice objects, spanning multiple different domains, and would thus be able to characterize a large portion of the decision maker's choice universe. Scholars of decision making could use such a model to study nearly any choice offered to the decision maker. In recent work, Bhatia (2017 and in press) has outlined one such model for the study of associative judgment and associative decision making. By using core ideas in semantic memory research, as well as large online datasets, both the goals and the methods of this work resemble those outlined in this paper. Future work could attempt a synthesis of the two approaches, so as to formulate a cohesive general model of naturalistic multiattribute choice.

Cognitive choice models. Future work should also attempt to implement more sophisticated choice theories, ones that are able to predict not only choice probabilities, but also their relationship with the choice set, reference points, the feasibility of deferral, decision time, decision confidence, and other relevant variables (e.g., Bhatia, 2013; Bhatia & Mullett, 2016; Glöckner, Hilbig, & Jekel, 2014; Holyoak & Simon, 1999; Payne, Bettman & Johnson, 1993; Roe et al., 2001; Noguchi & Stewart, in press; Trueblood, Brown, & Heathcote, 2014; Usher & McClelland, 2004). There are many such models that have been proposed for multiattribute choice, and with methodological advances in the computational modelling of decision making, we may be able to apply these more sophisticated models to the types of naturalistic decisions studied in this paper.

Model similarity. It would also be valuable to not just fit various decision models and evaluate out-of-sample predictions, but also to analyze the similarity of model performance on choice data. One approach to this type of analysis has been suggested by Broomell, Budescu and Por (2011). This approach involves making pairwise comparisons between models based on the proportion of choice data for which they make identical predictions. In the context of this paper, such an approach could illustrate when and where the WAD decision rule outperforms the other heuristics, which could in turn yield novel predictions regarding the types of choice environments where heuristics make the most accurate predictions. When applied to the fifteen different pairs of models considered in this paper, such an approach could also be used to uncover the latent structure of model similarity. Variants of such an analysis could also involve global model analysis (Pitt, Kim, Navarro & Myung, 2006) and model landscaping (Navarro, Pitt & Myung, 2004). Such techniques also provide more sophisticated and intuitive controls for model flexibility, than the cross validation procedure adopted in the current paper.

Multiattribute choice effects. Thus far, multiattribute decision making is almost always studied using highly stylized stimuli, presented to decision makers in explicit object-attribute matrices (e.g., Figure 1a). The reason for this design choice is that formal theories of multiattribute choice need quantifiable information about the attributes in order to be tested and compared. Many everyday decisions, however, do not involve explicitly presented attributes. Rather the attribute information used by decision makers is stored in their minds after having been learnt through experience (and additionally is far richer and more complex than information that could be presented in an explicit object-attribute matrix). Our approach provides one way of uncovering this attribute information. With the use of our recovered attributes, it is possible to test the robustness of documented multiattribute choice effects in more realistic choice settings. As discussed above, multiattribute choice is heavily task dependent, and it is possible that behaviors in these choice settings diverge from the types of behaviors observed using current designs. Indeed, some scholars have suggested that prominent decoy effects (which involve a change in choice probabilities as irrelevant choice objects are added or removed from the choice set) do not emerge in naturalistic settings where information is not presented in object-attribute matrices (Frederick et al., 2014). This is a topic of considerable debate in the field (see Huber et al., 2014, Simonson, 2014; and Yang & Lynn, 2014), and in future work we hope to use our approach to help address this disagreement.

Memory vs. stimuli-based decisions. One way in which this can be done involves insights from research on memory-based vs. stimuli-based decisions (Lynch et al., 1988; Lynch & Srull, 1982; Nedungadi, 1990; Rottenstreich et al., 2007). The most common experimental paradigm in this area makes decision makers memorize attribute information and then retrieve this information from memory while making multiattribute choices. These decisions are compared against decisions made when attribute information is presented explicitly during the choice. As an inversion of the approach in the current paper, it may be possible to adopt the memory-based vs. stimuli-based paradigm, and present some the object-feature information (e.g. food ingredients or movie keywords) explicitly on the screen. Choices with this explicit information presentation would then be compared to the types of choices elicited by our current experimental approach. Of course, given the richness of naturalistic objects like movies and food items (which are composed of thousands of features), it is not immediately clear how to present feature information in a manner that can be easily read and evaluated by participants. Future work should attempt to refine this paradigm so as to better understand the difference between naturalistic memory-based and stimulus-based multiattribute choices.

A variant of the design outlined above can also be used to study multiattribute decisions in which explicit attribute information accompanies object names with rich learnt representations. For example, in many food choice scenarios, linguistic descriptions of the food items are often presented alongside quantitative nutritional data (e.g. number of calories). Likewise, in many consumer decisions, verbal object descriptions (e.g. brand names) are often accompanied by quantitative attribute information (e.g. memory and processing speed). How are these two types of information aggregated? Although we have argued that the experiments in this paper capture naturalistic multiattribute choices, it is clear that most everyday choices involve both quantitative and linguistic information. Understanding how these choices are made is necessary for fully characterizing multiattribute decision making.

Prior information. Another extension to the paradigm presented in this paper involves the analysis of the information that people have prior to experiencing the choice objects. In this paper we have intentionally limited our experiments to settings involving popular movies and food items – objects which decision makers have previously experienced and for which they have rich learnt representations. This allows us to conveniently assume that our uncovered latent attributes match those of participants. Yet most multiattribute choices do not involve prior experience with the objects in consideration; instead, decision makers have to evaluate novel choice objects based on limited information. For example, when choosing a movie to watch, people typically consider new movies that they have never seen before, and evaluate these new movies based on reviews or trailers. It should not be difficult to extend the approach presented in this paper to predict such choices. Instead of learning latent attributes from keywords listed on websites like www.IMDB.com, it may instead be possible to learn these attributes from the content of movie reviews, or the words used in a movie trailer. More generally, the approach presented in this paper is not limited to large online datasets like www.IMDB.com or www.AllRecipes.com. This approach can be applied to nearly any information source used by participants to form mental representations of choice objects.

Free responses. Our approach also permits the analysis of a new type of decision problem: One involving completely free responses without the presentation of actual choice stimuli. Decision research is currently limited to settings in which participants are provided with explicit choice sets to choose between or to evaluate (e.g., “Do you want to watch Movie A or Movie B?”). However, everyday choices often involve individuals having to retrieve feasible choice objects from their memory, from a very large list of different objects they could possibly choose (e.g., “What movie do you want to watch?”). Our methods, which provide techniques for uncovering attribute representations for nearly all possible movies and food items, can be combined with existing memory models, to study choices in these free response settings.

Real World Choice

A notable benefit of our approach is the ability to extend the psychological analysis of multiattribute choice beyond the laboratory and predict real world choice data. Besides being necessary for evaluating the external validity of psychological theories, such an analysis has numerous commercial applications. For example, recommendation systems have, over the past decade, transformed online commerce and entertainment. The primary goal of recommendation systems is to suggest, for consumers, the types of goods, movies, books, restaurants, and other objects they would most like to consume. Although this is a natural area to apply existing psychological insights regarding decision making, most current approaches to building recommendation systems do not use these insights. More generally, the growth of the internet has led to increased choices for individuals and increased data regarding these choices, which in turn has stimulated the need for companies to quantitatively predict peoples' choices. The psychological study of decision making, which for the past seventy years has developed a detailed understanding of how these choices are made, has much to contribute to these applications, and our techniques provide one approach to translating decision making research for these purposes (see Golstone & Lupyan, 2016; Griffiths, 2015; and Jones, 2016 for an extensive discussion).

As existing multiattribute theories cannot be applied to real world decisions, they also cannot be used to improve these decisions. Our approach can however shed light on the ways in which people represent and aggregate the attributes of real objects. This opens up an avenue for the application of behavioral interventions to decisions involving these objects. Consider, for example, food choice, which is one of the two choice domains studied in this paper. An analysis of food choice has implications for public health, and our approach makes it possible to examine how individuals represent food items and how these representations influence choice. This can, in turn, allow for the study of socio-economic and situational determinants of food choice, and for the development of appropriate incentives, decision aids, and nudges to improve food decision making.

Ultimately, by combining existing theories of 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 decision making research.

References

Bettman, J. R., & Kakkar, P. (1977). Effects of information presentation format on consumer information acquisition strategies. *Journal of Consumer Research*, 3(4), 233-240.

Bhatia, S. (2013). Associations and the accumulation of preference. *Psychological Review*, 120(3), 522-543.

Bhatia, S. (2017). Associative judgment and vector space semantics. *Psychological Review*, 124(1), 1-20.

Bhatia, S. (in press). Semantic processes in preferential choice. *Journal of Experimental Psychology: Learning, Memory, and Cognition*.

Bhatia, S. & Mullett, T. (2016). The dynamics of deferred decision. *Cognitive Psychology*, 86(7), 112-151.

Blei, D. M., & McAuliffe, J. D. (2008). Supervised topic models. In *Advances in Neural Information Processing Systems* (pp. 121-128).

Brenner, L., Rottenstreich, Y., Sood, S., & Bilgin, B. (2007). On the psychology of loss aversion: Possession, valence, and reversals of the endowment effect. *Journal of Consumer Research*, 34(3), 369-376.

Browne, M. W. (2000). Cross-validation methods. Journal of Mathematical Psychology, 44, 108–132.

Davis-Stober, C. P., Dana, J., & Budescu, D. V. (2010). A constrained linear estimator for multiple regression. *Psychometrika*, 75(3), 521-541.

Dawes, R. M. (1979). The robust beauty of improper linear models in decision making. *American Psychologist*, 34, 571–582.

Dawes, R. M., & Corrigan, B. (1974). Linear models in decision making. *Psychological Bulletin*, 81(2), 95.

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

Estes, W. K. (1950). Towards a statistical theory of learning. Psychological Review, 57, 94–107.

Ettlin, F., Bröder, A., & Henninger, M. (2015). A new task format for investigating information search and organization in multiattribute decisions. *Behavior Research Methods*, 47(2), 506-518.

Fishburn, P. C. (1974). Lexicographic orders, utilities and decision rules: A survey. *Management Science*, 20(11), 1442-1471.

Frederick, S., Lee, L., & Baskin, E. (2014). The limits of attraction. *Journal of Marketing Research*, 51(4), 487-507.

Garner, W. R. (1978). Aspects of a stimulus: Features, dimensions, and configurations. In Eleanor Rosch & Barbara B. Lloyd (Eds.), *Cognition and Categorization* (pp. 99–133). Hillsdale, NJ: Lawrence Erlbaum

Gigerenzer, G., & Gaissmaier, W. (2011). Heuristic decision making. *Annual Review of Psychology*, 62, 451-482.

Gigerenzer, G., & Goldstein, D. G. (1996). Reasoning the fast and frugal way: models of bounded rationality. *Psychological Review*, 103(4), 650.

Glöckner, A., Hilbig, B. E., & Jekel, M. (2014). What is adaptive about adaptive decision making? A parallel constraint satisfaction account. *Cognition*, 133(3), 641-666.

Goldstein, D. G., & Gigerenzer, G. (2002). Models of ecological rationality: the recognition heuristic. *Psychological Review*, 109(1), 75.

Navarro, D. J., Pitt, M. A., & Myung, I. J. (2004). Assessing the distinguishability of models and the informativeness of data. *Cognitive Psychology*, 49(1), 47-84.

Griffiths, T. L. (2015). Manifesto for a new (computational) cognitive revolution. *Cognition*, 135, 21-23.

Griffiths, T. L., Steyvers, M., & Tenenbaum, J. B. (2007). Topics in semantic representation. *Psychological Review*, 114(2), 211-244.

Hastie, R. (2001). Problems for judgment and decision making. *Annual Review of Psychology*, 52(1), 653-683.

Hastie, T., Tibshirani, R., & Friedman, J. (2009). *The elements of statistical learning: Data mining, inference, and prediction.* New York: Springer. doi:10.1007/978-0-387-84858-7

Hertwig, R., Herzog, S. M., Schooler, L. J., & Reimer, T. (2008). Fluency heuristic: a model of how the mind exploits a by-product of information retrieval. *Journal of Experimental Psychology: Learning, memory, and cognition*, 34(5), 1191.

Hinton, G. E., & Salakhutdinov, R. R. (2006, July 28). Reducing the dimensionality of data with neural networks. *Science*, 313(5786), 504-507.

Holyoak, K. J., & Simon, D. (1999). Bidirectional reasoning in decision making by constraint satisfaction. *Journal of Experimental Psychology: General*, 128(1), 3-31.

Huber, O. (1979). Nontransitive multidimensional preferences: Theoretical analysis of a model. *Theory and Decision*, 10, 147–165.

Huber, J., Payne, J. W., & Puto, C. P. (2014). Let's be honest about the attraction effect. *Journal of Marketing Research*, 51(4), 520-525.

Jones, M. N., & Mewhort, D. J. (2007). Representing word meaning and order information in a composite holographic lexicon. *Psychological Review*, 114(1), 1-37.

Jones, M. N. (2016). Developing cognitive theory by mining large-scale naturalistic data. Big data in cognitive science, 1-12.

Jones, M. N., Willits, J. A., Dennis, S., & Jones, M. (2015). Models of semantic memory. In J. R. Busemeyer, Z. Wang, J. T. Townsend, & A. Eidels, *Oxford Handbook of Mathematical and Computational Psychology*, (pp. 232-254). Oxford, England: Oxford University Press.

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), 211-240.

Levav, J., Heitmann, M., Herrmann, A., & Iyengar, S. S. (2010). Order in product customization decisions: Evidence from field experiments. *Journal of Political Economy*, 118(2), 274-299.

Lynch Jr, J. G., Marmorstein, H., & Weigold, M. F. (1988). Choices from sets including remembered brands: Use of recalled attributes and prior overall evaluations. *Journal of Consumer Research*, 15(2), 169-184.

Lynch Jr, J. G., & Srull, T. K. (1982). Memory and attentional factors in consumer choice: Concepts and research methods. *Journal of Consumer Research*, 9(1), 18-37.

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.

Navarro, D. J., Pitt, M. A., & Myung, I. J. (2004). Assessing the distinguishability of models and the informativeness of data. *Cognitive psychology*, 49(1), 47-84.

Noguchi, T., & Stewart, N. (in press). Multialternative decision by sampling. *Psychological Review.*

Oppenheimer, D. M., & Kelso, E. (2015). Information processing as a paradigm for decision making. *Annual Review of Psychology*, 66, 277-294.

Payne, J. W., Bettman, J. R., & Johnson, E. J. (1993). *The Adaptive Decision Maker*. New York: Cambridge University Press.

Phillips, N. D., Neth, H., Woike, J. K., & Gaissmaier, W. (2017). FFTrees: A toolbox to create, visualize, and evaluate fast-and-frugal decision trees. *Judgment and Decision Making, 12,* 344-368.

Pitt, M. A., Kim, W., Navarro, D. J., & Myung, J. I. (2006). Global model analysis by parameter space partitioning. *Psychological Review*, 113(1), 57.

Roe, R. M., Busemeyer, J. R., & Townsend, J. T. (2001). Multialternative decision field theory: A dynamic connectionist model of decision making. *Psychological Review*, 108(2), 370-392.

Rogers, T. T., & McClelland, J. L. (2008). Précis of semantic cognition: A parallel distributed processing approach. *Behavioral and Brain Sciences*, 31(6), 689-714.

Rumelhart, D. E. & Todd, P. M. (1993) Learning and connectionist representations. In: Attention and performance XIV: Synergies in experimental psychology, artificial intelligence, and cognitive neuroscience, ed. D. E. Meyer & S. Kornblum, pp. 3–30. MIT Press.

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

Russo, J. E., Medvec, V. H., & Meloy, M. G. (1996). The distortion of information during decisions. *Organizational Behavior and Human Decision Processes*, 66(1), 102-110.

Saxe, A. M., McClelland, J. L., & Ganguli, S. (2013). Learning hierarchical category structure in deep neural networks. In M. Knauff, M. Pauen, N. Sebanz, & I. Wachsmuth (Eds.), Proceedings of the thirty-fifth annual conference of the Cognitive Science Society (pp. 1271-1276). Austin, TX: Cognitive Science Society.

Sanborn, A. N., Griffiths, T. L., & Shiffrin, R. M. (2010). Uncovering mental representations with Markov chain Monte Carlo. *Cognitive Psychology*, 60(2), 63-106.

Shah, A. K., & Oppenheimer, D. M. (2008). Heuristics made easy: an effort-reduction framework. *Psychological Bulletin*, 134(2), 207.

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

Shepard, R. N. (1980). Multidimensional scaling, tree-fitting, and clustering. *Science*, 210(4468), 390-398.

Simonson, I. (2014). Vices and virtues of misguided replications: the case of asymmetric dominance. *Journal of Marketing Research*, 51(4), 514-519.

Smith, E. E., & Medin, D. L. (1981). *Categories and Concepts*. Cambridge, MA: Harvard University Press.

Stewart, N., Chater, N., & Brown, G. D. A. (2006). Decision by sampling. Cognitive Psychology, 53, 1–26.

Stone, D. N., & Schkade, D. A. (1991). Numeric and linguistic information representation in multiattribute choice. *Organizational Behavior and Human Decision Processes*, 49(1), 42-59.

Tenenbaum, J. B., De Silva, V., & Langford, J. C. (2000). A global geometric framework for nonlinear dimensionality reduction. *Science*, 290(5500), 2319-2323.

Trueblood, J. S., Brown, S. D., & Heathcote, A. (2014). The multiattribute linear ballistic accumulator model of context effects in multialternative choice. *Psychological Review*, 121(2), 179.

Tversky, A. (1972). Elimination by aspects: A theory of choice. Psychological Review, 79, 281–229.

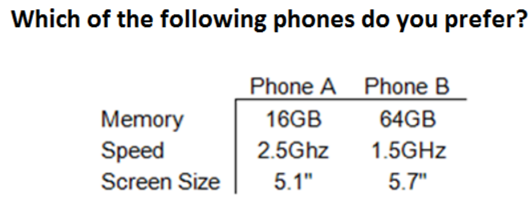
Tversky, A. (1977). Features of similarity. Psychological Review, 84, 327–352.

Usher, M., & McClelland, J. L. (2004). Loss aversion and inhibition in dynamical models of multialternative choice. *Psychological Review*, 111(3), 757.

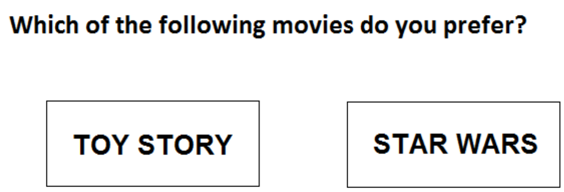
Vinson, D. P., & Vigliocco, G. (2008). Semantic feature production norms for a large set of objects and events. *Behavior Research Methods*, 40(1), 183-190.

Weber, E. U., & Johnson, E. J. (2009). Mindful judgment and decision making. *Annual Review of Psychology*, 60, 53-85.

Yang, S., & Lynn, M. (2014). More evidence challenging the robustness and usefulness of the attraction effect. *Journal of Marketing Research*, 51(4), 508-513.

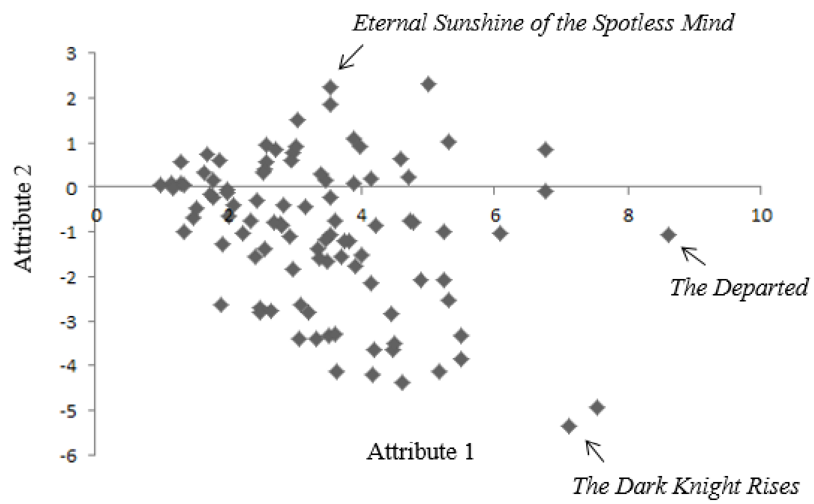


(a)



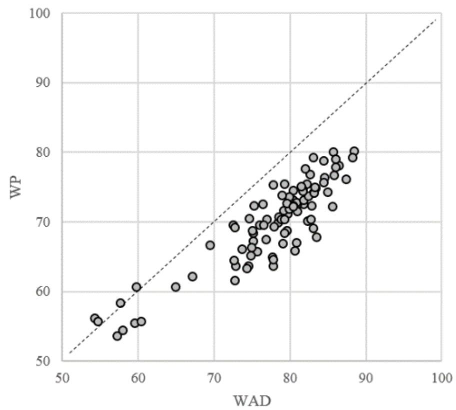
(b)

Figure 1. (a) A typical attribute-by-object matrix presentation for a choice between two phones. (b) The type of naturalistic decision studied in this paper.

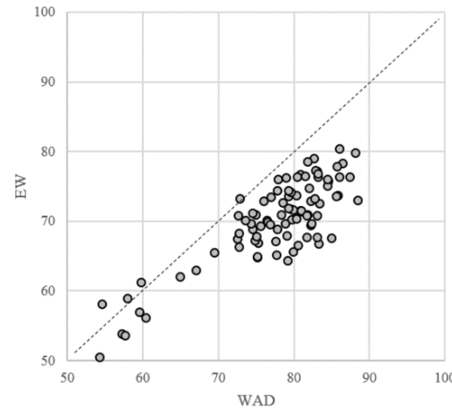


Latent Attribute 1		Latent Attribute 2	
Keyword	Weight	Keyword	Weight
murder	0.12	*martial-arts*	-0.17
female-nudity	0.11	*fistfight*	-0.15
violence	0.11	*tough-guy*	-0.15
independent-film	0.10	*showdown*	-0.15
blood	0.10	*hero*	-0.15
sex	0.10	*brawl*	-0.15
death	0.10	*shootout*	-0.15
nudity	0.09	*gunfight*	-0.14
husband-wife-relationship	0.09	*hand-to-hand-combat*	-0.14
based-on-novel	0.09	*action-hero*	-0.14

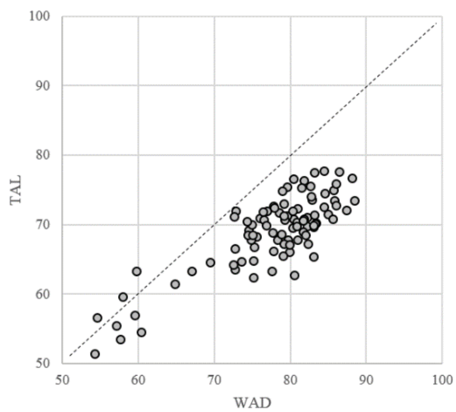
Figure 2. The values of the 100 movies in Studies 1 and 4 on the two latent attributes with the largest singular values, alongside the ten keywords with the largest absolute weights for these two latent attributes.



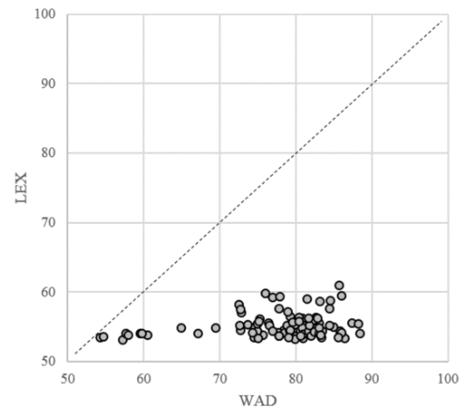
(a)



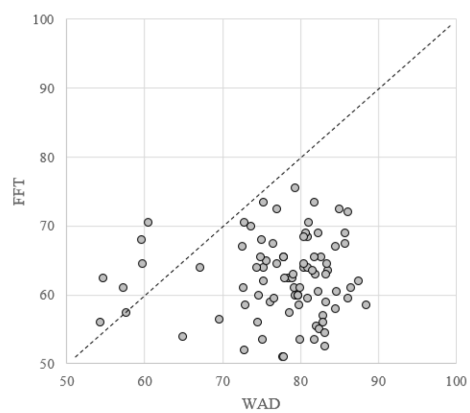
(b)



(c)



(d)



(e)

Figures 3a-e. Scatter plots of average participant accuracy rates of WAD compared to the five heuristic rules for Study 1. Here accuracy rates correspond to the proportion (%) of out-of-sample choices predicted accurately, by the best performing version of the corresponding decision rule. Each point corresponds to a single participant. Note that a random choice rule would obtain an accuracy rate of 50%.

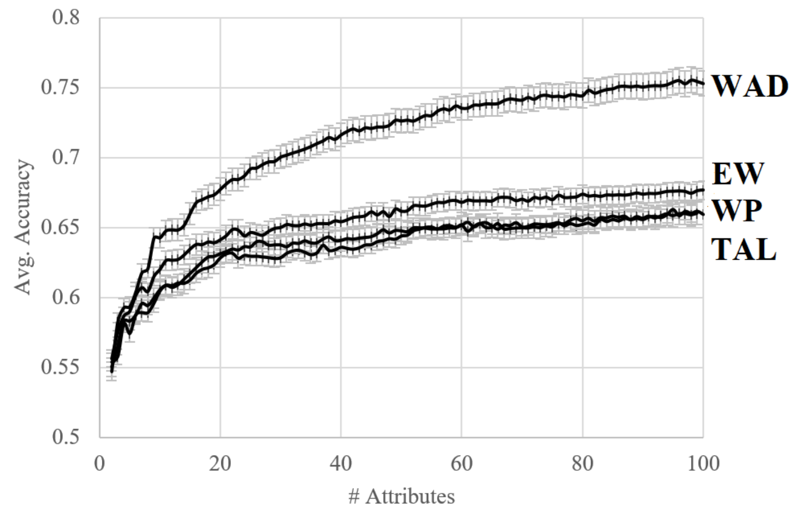


Figure 4. Average accuracy for the WAD, EW, WP, and TAL decision rules across participants in Study 1 as a function of number of latent attributes used (*L*). Error bars display +/- 1 standard error.

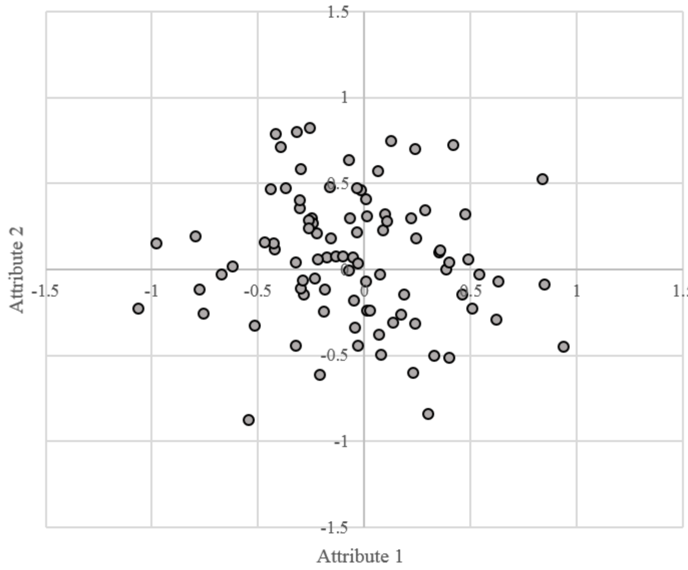
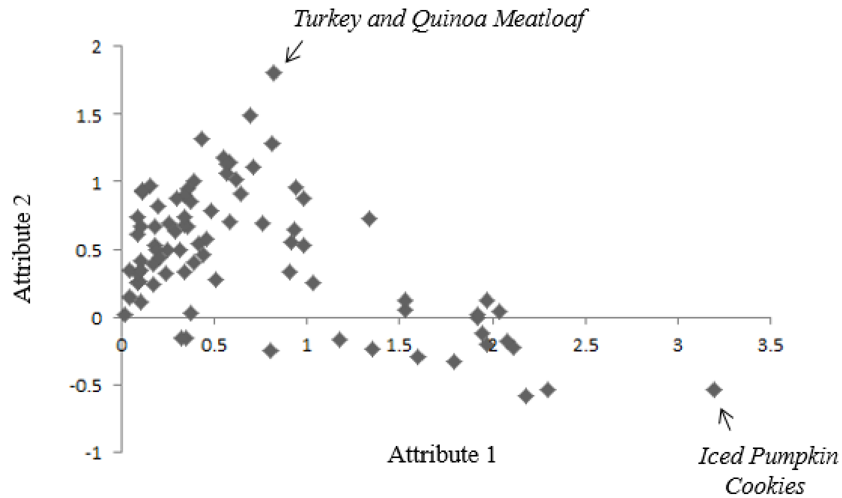
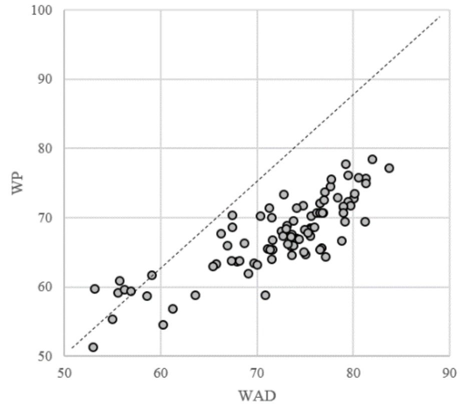


Figure 5. Best fitting attribute weights across participants for a 2-dimensional WAD model for Study 1. Here attributes 1 and 2 correspond to the two attributes in Figure 2.

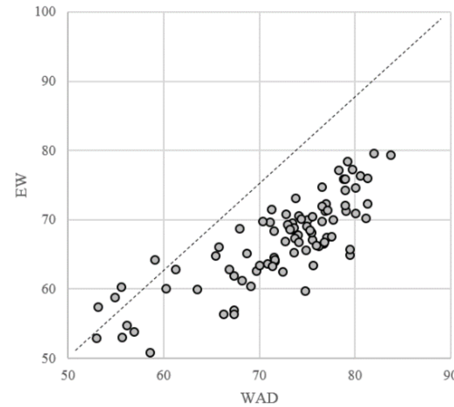


Latent Attribute 1		Latent Attribute 2	
Ingredient	Weight	Ingredient	Weight
all-purpose flour	0.35	*olive oil*	0.25
white sugar	0.34	*ground black pepper*	0.25
vanilla extract	0.33	*garlic, minced*	0.23
eggs	0.26	*water*	0.22
baking powder	0.26	*onion, chopped*	0.20
baking soda	0.25	*salt and pepper to taste*	0.19
salt	0.25	*vanilla extract*	-0.18
milk	0.21	*garlic powder*	0.18
butter	0.19	*baking soda*	-0.15
ground cinnamon	0.18	*dried oregano*	0.14

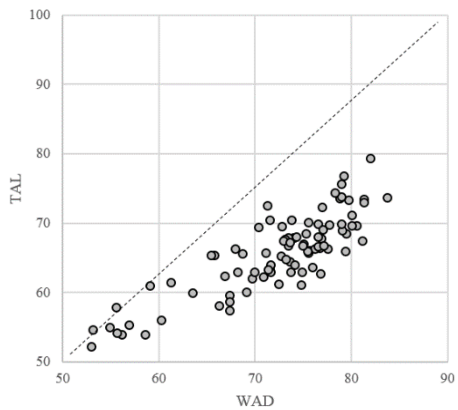
Figure 6. The values of the 100 food dishes in Studies 2 and 3 on the two latent attributes with the largest singular values, alongside the ten food ingredients with the largest absolute weights for these two latent attributes.



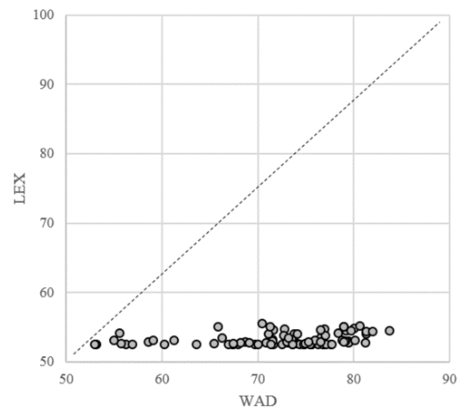
(a)



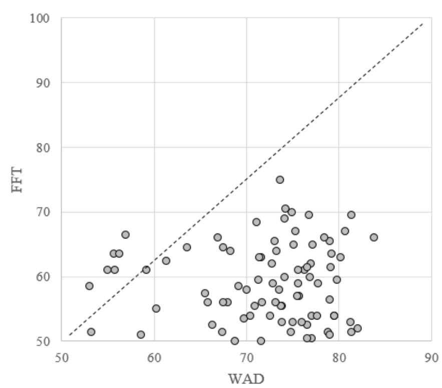
(b)



(c)



(d)



(e)

Figures 7a-e. Scatter plots of average participant accuracy rates of WAD compared to the five heuristic rules for Study 2. Note that a random choice rule would obtain an accuracy rate of 50%.

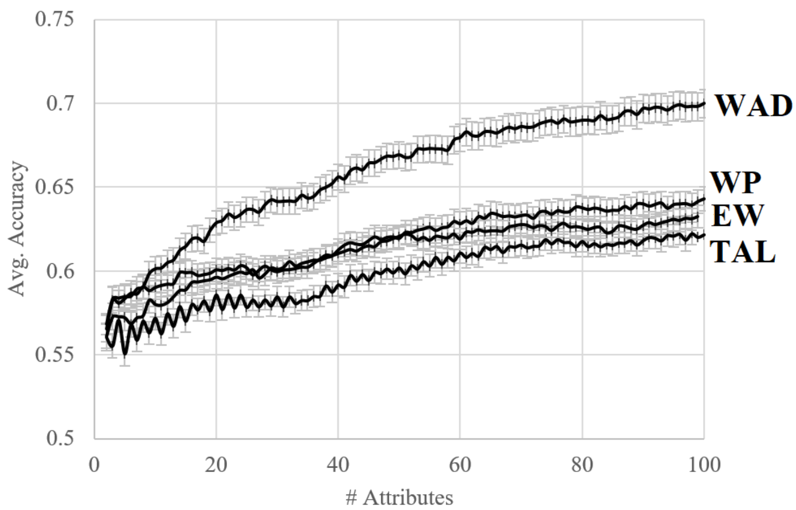
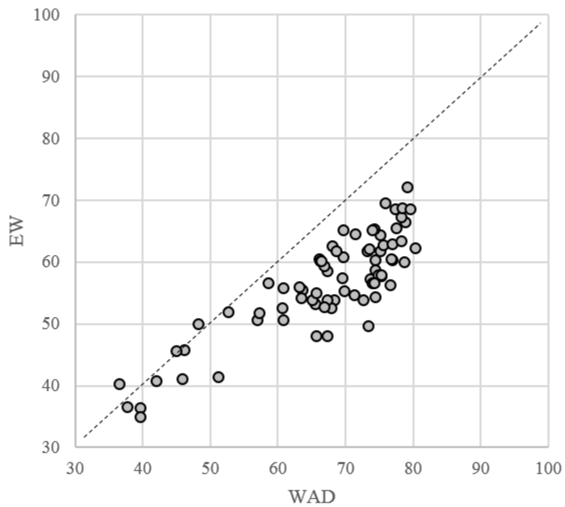
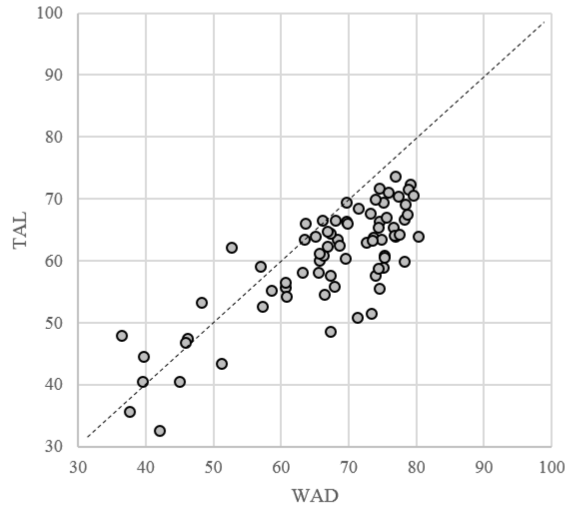


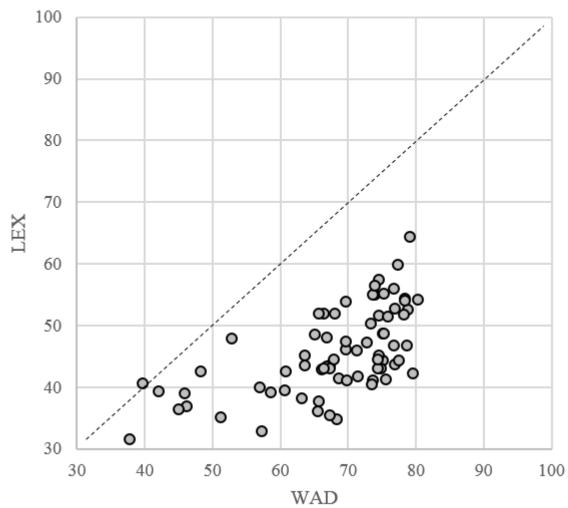
Figure 8. Average accuracy for the WAD, EW, WP, and TAL decision rules across participants in Study 2 as a function of number of latent attributes used (*L*). Error bars display +/- 1 standard error.



(a)



(b)



(c)

Figures 9a-c. Scatter plots of average participant accuracy rates of WAD compared to the three heuristic rules for Study 3. Note that a random choice rule would obtain an accuracy rate of 33.33%.

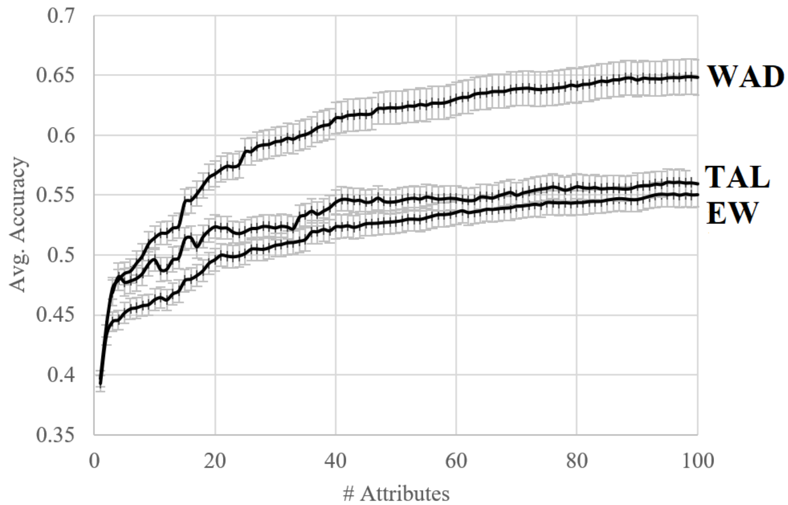
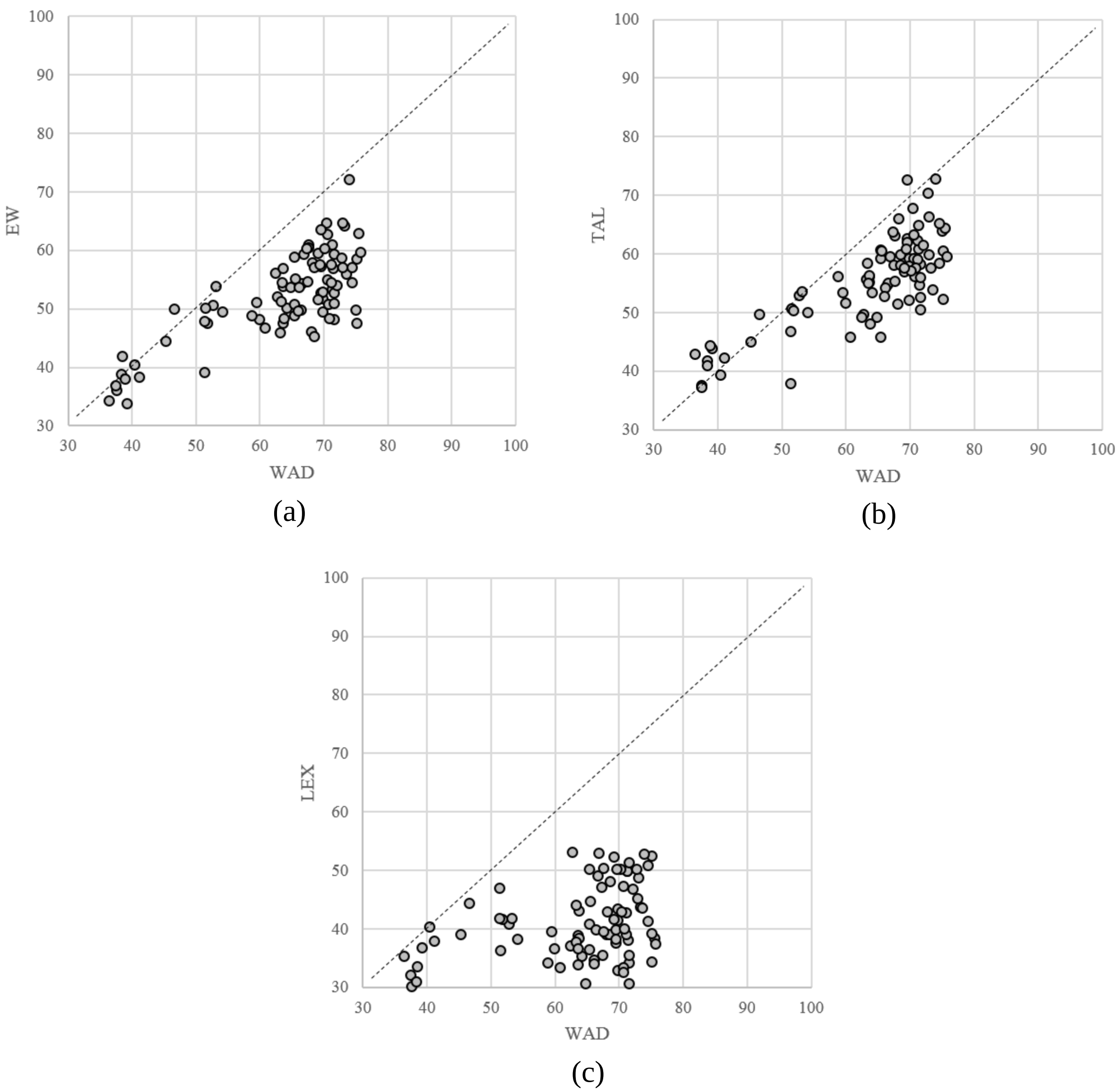


Figure 10. Average accuracy for the WAD, EW, and TAL decision rules across participants in Study 3 as a function of number of latent attributes used (*L*). Error bars display +/- 1 standard error.



Figures 11a-c. Scatter plots of average participant accuracy rates of WAD compared to the three heuristic rules for Study 4. Note that a random choice rule would obtain an accuracy rate of 33.33%.

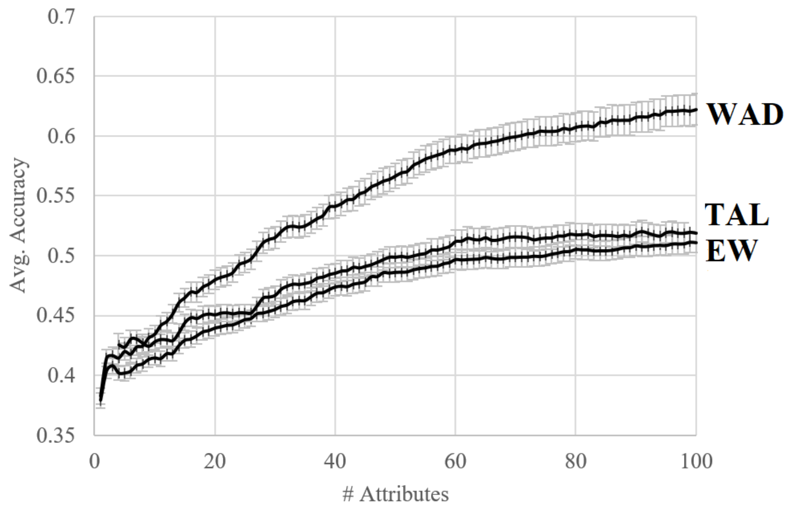


Figure 12. Average accuracy for the WAD, EW, and TAL decision rules across participants in Study 4 as a function of number of latent attributes used (*L*). Error bars display +/- 1 standard error.

Table 1. Summary of model fits for Study 1. “Accuracy” corresponds to average out-of-sample predictive accuracy for the best-fitting decision rule. κ indicates Cohen’s κ, which captures accuracy relative to chance. “Best” captures the percentage of participants for which the decision rule in consideration had the best accuracy. “# Attributes” captures the number of attributes (L) in the best-fitting decision rule. Here averages (“Avg.”) and standard deviations (“SD”) are taken across participants.

	Accuracy (Avg.)	Accuracy (SD)	Accuracy (Avg. κ)	Best (%)	# Attributes (Avg.)
WAD	77.65	7.59	0.55	90.21	75.70
WP	69.94	6.16	0.40	2.17	60.62
EW	70.21	5.95	0.40	1.10	68.84
TAL	68.95	5.41	0.38	2.17	63.54
FFT	61.37	6.96	0.22	4.34	--
LEX	55.18	1.72	0.10	0.00	--
WAD-RAND	56.08	3.05	0.12	--	36.09
WAD-SCRM	56.71	3.41	0.13	--	23.48

Table 2. Summary of model fits for Study 1 using only choice trials in which participants recognized both movies.

	Accuracy (Avg.)	Accuracy (SD)	Accuracy (Avg. κ)	Best (%)	# Attributes (Avg.)
WAD	78.03	7.26	0.56	84.70	73.45
WP	70.53	5.82	0.41	2.37	60.46
EW	70.68	5.73	0.41	2.35	65.27
TAL	69.66	4.99	0.39	2.35	60.44
FFT	61.01	8.40	0.22	8.23	--
LEX	55.70	1.85	0.11	0.00	--
WAD-RAND	57.06	3.82	0.14	--	31.42
WAD-SCRM	57.22	3.75	0.14	--	30.10

Table 3. Summary of model fits for Study 2.

	Accuracy (Avg.)	Accuracy (SD)	Accuracy (Avg. κ)	Best (%)	# Attributes (Avg.)
WAD	72.26	7.15	0.45	78.89	75.51
WP	67.43	5.39	0.35	5.55	56.51
EW	66.94	6.20	0.34	4.44	62.40
TAL	65.60	5.56	0.31	2.22	60.19
FFT	58.56	6.40	0.16	8.88	--
LEX	53.21	0.86	0.06	0.00	--
WAD-RAND	57.02	3.83	0.14	--	33.03
WAD-SCRM	57.83	3.36	0.16	--	40.10

Table 4. Summary of model fits for Study 3.

	Accuracy (Avg.)	Accuracy (SD)	Accuracy (Avg. κ)	Best (%)	# Attributes (Avg.)
WAD	67.05	11.18	0.51	85.33	69.29
EW	56.47	8.07	0.35	1.33	75.29
TAL	60.20	8.75	0.40	13.33	49.27
LEX	44.96	7.59	0.17	0.00	--
WAD-RAND	38.68	2.86	0.08	--	29.83
WAD-SCRM	39.71	2.77	0.10	--	35.86

Table 5. Summary of model fits for Study 4.

	Accuracy (Avg.)	Accuracy (SD)	Accuracy (Avg. κ)	Best (%)	# Attributes (Avg.)
WAD	63.75	10.72	0.46	88.00	78.22
EW	52.37	7.38	0.29	4.00	75.82
TAL	55.30	7.68	0.33	8.00	57.74
LEX	40.84	6.28	0.11	0.00	--
WAD-RAND	41.23	4.11	0.12	--	14.02
WAD-SCRM	42.12	5.50	0.13	--	24.96