Running head: COMPARING MODELS OF PREFERENCE

Competing theories of multi-alternative, multi-attribute preferential choice

Brandon M. Turner
Department of Psychology
The Ohio State University

Dan R. Schley
Rotterdam School of Management
Erasmus University

Carly Muller
Department of Psychology
The Ohio State University

Konstantinos Tsetsos
Department of Neurophysiology and Pathophysiology
University Medical Center Hamburg

Address correspondence to:

Brandon M. Turner

Department of Psychology

The Ohio State University

turner.826@gmail.com

Abstract

In accounting for phenomena present in preferential choice experiments, modern models assume a wide array of different mechanisms such as lateral inhibition, leakage, loss aversion, and saliency. These mechanisms create interesting predictions for the dynamics of the deliberation process as well as the aggregate behavior of preferential choice in a variety of contexts. However, the models that embody these different mechanisms are rarely subjected to rigorous quantitative tests of suitability by way of model fitting and evaluation. Recently, complex, stochastic models have been cast aside in favor of simpler approximations, which may or may not capture the data as well. In this article, we use a recently developed method to fit the four extant models of context effects to data from two experiments: one involving consumer goods stimuli, and another involving perceptual stimuli. Our third study investigates the relative merits of the mechanisms currently assumed by the extant models of context effects by testing every possible configuration of mechanism within one overarching model. Across all tasks, our results emphasize the importance of several mechanisms such as lateral inhibition, loss aversion, and pairwise attribute differences, as the mechanisms contribute positively to model performance. Together, our results highlight the notion that mathematical tractability, while certainly a convenient feature of any model, should neither be the primary impetus for model development nor the promoting or demotion of specific model mechanisms. Instead, model fit, balanced with model complexity, should be the greatest burden to bear for any theoretical account of empirical phenomena.

Keywords: Multi-attribute Linear Ballistic Accumulator model, Leaky Competing Accumulator model, Multi-alternative Decision Field Theory model, Associations and Accumulation model

Competing theories of multi-alternative, multi-attribute preferential choice

Introduction

Substantial progress has been made in our understanding of decision-making processes by studying patterns of behavioral data involving “simple” stimulus sets. We refer to a simple stimulus set as a set of usually two stimuli comprised of a single featural dimension such as direction of motion, numerical value, or color. One domain in which these simple stimuli are most popular is that of perceptual decision making. Patterns of decisions about stimuli of this type are often unsurprising: behavioral measures such as accuracy are systematically related to the levels of the featural properties of the stimuli. In other words, observers are remarkably consistent in their decision making, and this consistency makes it relatively easy to develop theories and ensuing mathematical models of the decision making process (Ratcliff, 1978; Usher & McClelland, 2001; Shadlen & Newsome, 2001; Reddi & Carpenter, 2000; Brown & Heathcote, 2005, 2008; Rouder, Morey, Gomez, & Heathcote, 2014).

Yet, simple stimuli are not often encountered in real life. Most of our daily decisions involve a choice among several options, and often these options vary along a number of dimensions. Typically, these decisions are consumer centric, where the feature dimensions (i.e., attributes) comprising the options are hedonic in nature, although perceptual studies have shown similar results (Trueblood, Brown, Heathcote, & Busemeyer, 2013). By virtue of their construction, an observer’s evaluation of hedonic stimuli cannot be evaluated in an objective sense, meaning that conventional behavioral metrics such as accuracy are not well defined. Instead, researchers must focus their efforts on understanding inconsistencies in the decisions that are made.

As with nearly all human decision making tasks, a set of reliable inconsistencies have emerged over the past few decades. These inconsistencies, referred to as “context effects”, typically occur when a third option distorts the preference share for two existing options that vary along at least two attribute dimensions. The three most prominent context effects are the similarity (Tversky, 1972), attraction (Huber, Payne, & Puto, 1982), and compromise (Simonson, 1989) effects. Context effects have been particularly intriguing for decision-making theorists because somewhere in the extension from simple stimulus sets to multi-attribute and multi-alternative stimulus sets, putatively fundamental axioms about human decision making are violated. The violation of these axioms has provided strong evidence against traditional utility models (Luce, 1959; Krantz, 1989; Tversky, 1972) that rely on the assumption of a consistent valuation of options across different contexts. In other words, context effects cannot be explained by simple decision making models (e.g., Stone, 1960) that assume that the strength of evidence for a choice is based purely on a monotonic function of the attribute values. Hence, the mere existence of context effects highlights the need for additional (or different) theoretical overhead to extend classic models of simple decision making to capture the complex preferences observed in multi-alternative, multi-choice decision making.

Assuming that simple decision making models cannot account for context effects, the obvious question to ask is “what mechanisms can be added to explain these choice inconsistencies?” Over the past half century, decision-making research has been propelled forward by integrating models of perceptual decision making into models of preference (e.g., Kahneman & Tversky, 1979; Tversky, 1977). Researchers utilize such models of perception not only because the perceptual system serves as an efficient analogy for the mechanism underlying judgment and choice, but also because the perceptual system acts as a fundamental input into the decision process (Stewart, Chater, & Brown, 2006; Summerfield & Tsetsos, 2012; Schley & Peters, 2014; Chandon & Ordabayeva, 2009; Frydman & Nave, 2016). At this point, an assortment of mechanisms have already been proposed, and four such constellations of mechanisms comprise the extant models of context effects. Most of these models rely on complex, stochastic processes to capture choice inconsistency, and as a result, these models are "simulation-based", meaning that the likelihood functions relating the models' parameters to data have yet to be derived (see I. J. Myung, 2000; Turner & Van Zandt, 2012, for tutorials on likelihood-based inference). Recently, Trueblood, Brown, and Heathcote (2014) have shown that simulation-based models may not be the only mechanistic framework for capturing context effects. Trueblood et al. present a model that makes a number of unique and simplifying assumptions, eventually arriving at a model that frees us from the shackles of stochasticity. In so doing, the predictions from the model become tractable, and analytic expressions for the likelihood function can be derived (see Brown & Heathcote, 2008). Few, if any, would oppose the notion that mathematical tractability is a feature worth striving for. Yet, mathematical tractability does not map onto the principle of model complexity as it is used to evaluate a model's ability to fit empirical data (I. J. Myung & Pitt, 1997; I. J. Myung, 2000; I. J. Myung, Forster, & Browne, 2000; Teodorescu & Usher, 2013). Instead, model complexity is defined as the range of hypothetical patterns of data a model can fit: the more complex a model, the better able it is to fit empirical data.

Many have argued that a full assessment of a model's credentials involves not only model complexity, but the model's complexity relative to the observed data (Teodorescu & Usher, 2013; Turner, Dennis, & Van Zandt, 2013). While a few studies have evaluated some models' ability to fit empirical data, to our knowledge, model fit has not been compared across all extant models simultaneously. Perhaps more alarming is that proper evaluations of model flexibility have yet to be established. The impetus for this shortcoming undoubtedly arises from the lack of suitable methodology for fitting simulation-based models to data. Recently, an ensemble of methods for fitting simulation-based models to data have been developed and used in several cognitive modeling applications (Turner & Van Zandt, 2012; Turner & Sederberg, 2012; Turner, Dennis, & Van Zandt, 2013; Turner & Sederberg, 2014; Turner & Van Zandt, 2014;

Turner, Sederberg, & McClelland, 2016; Palestro, Sederberg, Osth, Van Zandt, & Turner, 2016). These methods, which have been assigned the misnomer "likelihood-free" use simulation techniques to approximate the likelihood function and relate model predictions to empirical data. In theory, any computational model can be fit to data using these likelihood-free approaches, and especially important for this article is that they can now be fit in a Bayesian paradigm. The Bayesian paradigm affords us some opportunities that frequentist-based approaches do not, such as a direct assessment of parameter uncertainty, model complexity, and model identifiability. Hence, joining likelihood-free algorithms with Bayesian statistics provides a powerful framework for comparing theoretical aspects of simulation-based models that naturally takes into account both model fit and complexity (Turner, Dennis, & Van Zandt, 2013; Turner et al., 2016).

The goal of the present article is to provide evidence for and against the set of proposed mechanisms assumed by extant models of context effects. The outline of this article is as follows. We begin by first reviewing specific details of context effects and extant models. This section is brief out of necessity, but the reader is encouraged to consult the original publications as well as the Appendix for additional details. Second, we develop a taxonomy for comparing and contrasting the mechanisms assumed by the models. We then fit the models to data from two studies – one experiment involving consumer good stimuli, and one involving perceptual stimuli. In the first study, we fit the models to data from Berkowitsch, Scheibehenne, and Rieskamp (2014). In the second study, we describe an experiment using perceptual stimuli that also demonstrates some classic context effects (also see Trueblood et al., 2013), and fit the models to these data. Across both of these studies, we fit an assortment of hierarchical and non-hierarchical models, allowing for biases in the processing of attribute dimension information, and compare the models on the basis of the deviance information criterion (DIC; Celeux, Forbes, Robert, & Titterington, 2006). The DIC measure is a Bayesian performance metric that balances both model fit and model complexity that can be computed from

Markov chain Monte Carlo (MCMC) output. We aimed to make our model comparison agnostic by exploring multiple variations of each model, including some versions that were not initially considered by the original authors. Because presumably context effects represent a domain general phenomena impacting both perceptual and consumer decisions, Studies 1 and 2 provide consensus in our modeling analyses across these different domains.

Although Studies 1 and 2 compare and contrast the set of extant models, these analyses are limited by the set of idiosyncrasies imposed by the model structures in order to implement them. In Study 3, we attempt to integrate out the particularities of the modeling assumptions by testing all possible mechanistic configurations. Using our "switchboard" analysis, we evaluate not only the extant models, but also an assortment of hybrid models in an effort to assess the relative fidelity of each proposed model mechanism. This switchboard model serves not only as a novel methodological tool, but offers a unique ability for theory testing. We close with a discussion of theoretical considerations in model development, while emphasizing the importance of model evaluation even when models are mathematically intractable.

The Similarity, Attraction, and Compromise Effects

Before discussing the extant models, we will briefly review the similarity (Tversky, 1972), attraction (Huber et al., 1982), and compromise (Simonson, 1989) effects.

The Similarity Effect. Consider a choice between Option A and Option B that vary in terms of price and quality. For simplicity, let's assume that price and quality are indexed between 0 and 100, where expensive items have a low price index (PI) and less expensive items have a higher price index. Conversely, low quality will be represented with a low quality index (QI) and high quality with a high quality index. Suppose that Option A is inexpensive $PI_A = 90$, and of low quality $QI_A = 10$, and Option B is expensive $PI_B = 10$, and of high quality $QI_B = 90$. Now suppose that for this choice set, people have equal preferences for Options A and B (i.e., 50% choose A and 50% choose B).

A similarity effect occurs when a third option, Option C, that is similar to either Option A or B is added to the choice set (Tversky, 1972). For example, suppose that Option C is more expensive than Option A ($PI_C = 80$), and of higher quality than Option A ($QI_C = 20$). Thus, Option C is similar to Option A but dissimilar to Option B. In this case, *relative* preference for Option A compared to Option B decreases because it is seen as exchangeable with Option C (i.e., both options are relatively inexpensive and of relatively low quality). The key pattern is that when Option C similar to Option A is added to the choice set, relative to the binary choice between Options A and B, the relative preference for Option A decreases. Conversely, if Option C was more similar to Option B (e.g., $PI_C = 20$ and $QI_C = 80$), a similarity effect occurs when the preference for Option B decreases in the presence of Option C, relative to the binary choice between Options A and B.

The Attraction Effect. To illustrate the attraction effect (Huber et al., 1982), reconsider the Options A and B from above. An attraction effect occurs when a third option, Option C, is added to the choice set but is dominated by one of the Options A or B. Conventionally, Option C is often referred to as a “decoy” option, as it is similar, but inferior to, either Option A or B. For example, if Option C is more expensive than Option A ($PI_C = 80$), but is still the same quality as Option A ($QI_C = 10$), Option A dominates Option C. In this case, Option A should be more attractive because it is less expensive than Option C. When Option C is added to the choice set, an attraction effect is observed if the preference for Option A increases relative to the binary choice between Options A and B. Conversely, if Option C is dominated by Option B (e.g., $PI_C = 10$ and $QI_C = 80$), Option B should be more attractive in the presence of Option C, causing the preference share to shift toward Option B, relative to the binary choice between Options A and B.

To recap, when Option C is very similar to, but not dominated by, Option A, a similarity effect occurs if the inclusion of Option C *decreases* the preference for Option A. However, when Option C is similar to, but dominated by, Option A, an attraction effect occurs if the inclusion of Option C *increases* preference for Option A.

The Compromise Effect. To illustrate the compromise effect (Simonson, 1989), we will augment the ordering of options compared in our running example. Reconsider the Options A and B from above, and a new Option C that is not dominated by either option, and is not similar to either option (e.g., $PI_C = 50$ and $QI_C = 50$). Suppose that in a binary choice between Options A and C, people have comparable preferences (e.g., 50% prefer Option A and 50% prefer Option C). Further suppose that in a binary choice between Options B and C, people have comparable preferences (e.g., 50% prefer Option B and 50% prefer Option C). Thus, given binary choices, Option C is similarly preferred to Options A or B. Now, consider a ternary choice between Options A, B, and C. In the presence of both Options A and B, Option C is framed as a compromise between Options A and B, because Option C is less expensive than Option B but of higher quality than Option A. Although Option C is not necessarily preferred to Options A or B in binary choice trials, a compromise effect is observed if Option C commands a greater preference in a ternary choice trial (e.g., 30% prefer Option A, 30% prefer Option B, and 40% prefer Option C).

Early research on context effects demonstrated violations of normative theories of preference (Tversky, 1972; Huber et al., 1982; Simonson, 1989) using between-subject designs with one-shot choices (Bettman, Luce, & Payne, 1998; Simonson & Tversky, 1992; Huber et al., 1982). Recent interest in these context effects has shifted away from their evidential value against normative models and toward the psychological processes underlying these effects. Because context effects are theorized to be a consequence of the perceptual process (Tversky, 1972; Huber et al., 1982; Simonson, 1989), researchers have begun exploring context effects using tools developed in this domain (Usher & McClelland, 2001; Roe, Busemeyer, & Townsend, 2001; Hotaling, Busemeyer, & Li, 2010; Bhatia, 2013; Trueblood et al., 2014). These approaches adapt within-participant experimental designs and perceptual analogies to study the perceptual antecedents of context effects (Choplin & Hummel, 2005; Trueblood et al., 2013; Tsetsos, Chater, & Usher, 2012; Tsetsos, Usher, & McClelland, 2011). In addition to demonstrating analogous behavioral effects, these procedures offer rich data sets for testing models aimed at capturing the latent psychological processes underlying these choices. Model-based approaches provide unique insight because, presumably, models that best account for the data provide support for the theoretical mechanisms within these models. As the goal of the current article is to compare the relative merits of the extant models of context effects, we review them in the next section.

Mechanistic Models of Preference

At this point, there are four models (illustrated in Figure 1) that can capture all three context effects: the Multialternative Decision Field Theory (MDFT; Busemeyer & Townsend, 1993; Roe et al., 2001; Hotaling et al., 2010) model, the Multiattribute Leaky Competing Accumulator (MLCA; Usher & McClelland, 2004) model[1], the Associations and Accumulation model (AAM; Bhatia, 2013), and the Multiattribute Linear Ballistic Accumulator (MLBA; Trueblood et al., 2014) model. To account for context effects, these models vary in the theoretically-proposed mechanisms involved in the choice process. The differences in these assumptions are illustrated in Figure 1, where each diagram represents stages in the mapping from objective attribute values (i.e., left side) to the decision (i.e., right side). Arrows indicate a dependency between the nodes in the graph, and plates represent different sections in the model. For example, in the AAM, the attributes are first processed and then used to drive attention toward one feature or another. For this reason, the AAM has a double-headed arrow at the attribute selection process. Similarly, the MDFT, MLCA, and AAM models have arrows connecting each choice alternative at the decision process, indicating a lateral inhibition process among the choices.

We now discuss the basic details of the four models, presented in chronological order. For a more technical description of the models, we encourage the reader to consult the supplementary materials or the original publications.

Multialternative Decision Field Theory. The Multialternative Decision Field Theory (MDFT; Roe et al., 2001; Hotaling et al., 2010) model defines choice as similarity-dependent, leaky integration of information subject to competition and attentional shifts through four distinct layers, which are represented in Figure 1 with three plates. The first layer determines, according to a stochastic process, which attribute is attended at a given moment. The first plate represents the attribute selection process, where attention is directed to one of the two attributes in this illustration. The second layer represents the attribute values of the alternatives on the active attribute, and the third layer calculates valences, which correspond to advantages and disadvantages of an alternative at a particular moment in time. The valences may alternate between positive and negative as attention fluctuates between the attributes of a given choice alternative (also see Usher & McClelland, 2004), and are calculated as the difference between the value of the currently considered option and the mean value of the other options. These two layers in the model are illustrated as a single plate in Figure 1, labeled input processing. The fourth layer implements a leaky integration using as input the valences from the third layer, and a competitive inhibition process among the accumulators. This choice competition ensues until a threshold amount of evidence has been acquired, at which point a choice is made corresponding to the accumulator that reaches the threshold first. Competition among the accumulators at the fourth layer is implemented via inhibitory connections with distant-dependent (or similarity-based) strengths. Inhibitory connections in the connectivity matrix (see the Appendix) decrease as a specific distance metric between alternatives increases. This distance metric is determined based on a difference/indifference space, where options that are located on an indifference line compete more strongly (i.e. are perceived as being closer to each other) whereas those that are dominated by other options compete less strongly (Hotaling et al., 2010). The MDFT model assumes that preference states in the accumulation process may take either positive or negative values, something that is contrary to another model of choice preference that we discuss below (Tsetsos, Gao, & McClelland, 2012; Tsetsos, Chater, & Usher, 2015). The fourth layer is represented in Figure 1 as the decision plate, where the roles of lateral inhibition and leakage are illustrated by the node-to-node connections.

At its core, the MDFT model uses four free parameters to predict choice probabilities. The “feedback matrix” contains three of these parameters: ϕ_1, ϕ_2 and β. The parameters ϕ_1 and ϕ_2 are part of the MDFT model’s Gaussian mapping function which converts distances in attribute space to lateral inhibition strengths. The parameter β represents the dominance dimension weight in the the indifference/difference function, where more weight is assigned to improvements in both attributes than tradeoffs between attributes. Finally, MDFT assumes momentary fluctuations in the accumulation process which are controlled by the error term Σ. Together, the mechanisms of the MDFT model make it stochastic, and so for the experimental paradigms we consider below, it must be simulated many times if we wish to fit it to data.

The distance-dependent inhibition terms are useful in MDFT because they allow the model to account for both attraction and compromise effects. In capturing the attraction effect, the local inhibition between nearby alternatives in attribute space couples their input and results in a negative valence of the third, but inferior, option. However, because the model assumes that inhibition is calculated as a function of distance, the third option applies a negative inhibition (which is effectively excitation) toward one of the coupled options, eventually promoting the uninhibited alternative. The result is a choice advantage for the dominating option (i.e., the attraction effect). To predict the compromise effect, the MDFT model uses the inhibition from the two extreme options to increase choice preference for the mediating compromise option. Because the two extreme options are equidistant from the compromise option, the extreme options become correlated and split the choice share, so that the compromise option gains advantage and is eventually selected. Finally, in the similarity effect, two similar options are activated/deactivated together due to attentional switching instead of distance-dependent inhibition. When the two similar alternatives are less activated than a third dissimilar alternative, the dissimilar option is selected.

Multiattribute Leaky Competing Accumulator. Similar to the MDFT model, the Multiattribute Leaky Competing Accumulator (MLCA; Usher & McClelland, 2004) model conceptualizes choice as a leaky preference integration that is susceptible to choice competition and attentional shifts. The LCA model assumes two types of nonlinearity. First, the values of the accumulators that encode preference states are not permitted to be negative. To instantiate this, when any accumulator becomes negative, its value is simply reset to zero. Having the zero bound on the accumulation process allows the LCA model to eliminate inferior options and prevent noise from continuing to accrue during the decision process. Second, the MLCA model assumes an asymmetric value function where losses are weighted more heavily than gains (Kahneman & Tversky, 1979). Losses and gains here refer to disadvantages and advantages, respectively. The value function allows the model to maintain a sense of "status quo", which is important in capturing reference point effects in value comparison.

The MLCA model is arranged in four layers of a connectionist network, represented in Figure 1 as three plates (as in the MDFT model above). The first layer controls which attribute is actively attended on any given moment, with the attentional allocation alternating stochastically across attributes (i.e., first plate). The second layer represents the attribute values on the active attribute, and the third layer calculates advantages and disadvantages between all pairs of options via the asymmetric value function. These two layers are depicted in Figure 1 on the input processing plate where the nodes correspond to the attribute values for the three options (left column of nodes), and six pairwise differences (right column of nodes). In the fourth layer, the pairwise differences are integrated as preferences across time. To instantiate this, the MLCA model uses a connectivity matrix such that diagonal terms correspond to a self-connectivity coefficient (i.e., a leakage parameter), and off-diagonal elements correspond to inhibitory connections.

As in the MDFT model above, the roles of lateral inhibition and leakage are illustrated by the node-to-node connection in the decision process plate.

In total, the MLCA model has four free parameters. The parameter I_0 represents a baseline input value for each option that determines a minimum activation value. The parameter η is the noise term, akin to Σ in the MDFT model, that allows for momentary fluctuations in the accumulation process that are not governed by the inputs or other mechanisms. The parameters k and L correspond to the mechanisms of leakage and lateral inhibition, respectively, which are the primary driving force of the model's accumulation dynamics (Usher & McClelland, 2001; Bogacz, Brown, Moehlis, Holmes, & Cohen, 2006).

The value function directly allows for predicting the attraction and compromise effects due to an advantage given to similar options, and penalty to distant options. This function, in contrast to the MDFT model, does not require a distant-dependent inhibitory mechanism. The value function also contributes to the compromise effect through a pairwise comparison of each alternative. In this calculation, the two extremes options, being on average farthest from the other choice options, receive one large and one small disadvantage, whereas the compromise option only receives two small disadvantages. The value function penalizes the large disadvantages more heavily, and thus the compromise option receives greater input, and on average, is chosen more often. The MLCA model predicts a similarity effect via a correlation between activations of the two similar options, which end up splitting their choice shares (Tsetsos, Usher, & Chater, 2010; Tsetsos et al., 2015).

Associations and Accumulation Model. The Associations and Accumulation model (AAM) uses a connectionist network model of the decision process, which assumes an association between choice task and attribute accessibility within a stochastic sequential-sampling accumulation framework. Similar to MDFT and MLCA model above, Figure 1 represents the accumulation process as a three-layer computation. At the first layer, a similar attribute selection process occurs where the features are attended to on the basis of their values. The AAM conceptualizes the attribute selection process differently compared to MLCA and MDFT models, where the sum of the attributes from all options are used to determine the accessibility (i.e., probability of attending) to each attribute for a given stimulus set (i.e., a process represented as a double-headed arrow). Accessibility is proportional to the sum of the attribute values along a given dimension. Every attribute has a linear activation function and consequently, the parameter a_0 is introduced to give nonnegative constant input that is assumed to be identical across attributes. The parameter a_0 moderates the strength of associative biases, where values of a_0 are low when biases are strong and increasing values of a_0 reflect disappearance of these biases. At $a_0 = \infty$, each attribute is equally likely to be sampled. The attribute summation rule allows more salient alternatives to receive stronger inputs, and thus are assigned higher activation values. Nonsalient alternatives are given activation values equal to zero and ignored.

Once the attribute dimensions have been attended, the values corresponding to that dimension are transformed into valences using a transformation function dictated by the parameter α. This parameter is often set equal to one to keep the model simple, and we will maintain this assumption here. At the third layer, the attribute values are provided as input into the accumulation process in the same way that accumulation occurs in the MLCA and MDFT models. While not investigated in Bhatia (2013), we allow for the possibility of lateral inhibition l and decay d, which all play a role in this final layer.

Overall, the AAM model possesses five parameters: two attribute parameters, a_0 and α, two preference-state feedback parameters, d and l, and one noise parameter e that determines the degree of moment-to-moment variability in preference state. The AAM model is capable of accounting for context effects as a result of the associative attentional weights placed on the alternatives in the choice set. For example, in the attraction effect, the addition of the decoy increases the sampling probability of the dominant option's primary attribute, making it more desirable and thus more likely to be chosen.

Multiattribute Linear Ballistic Accumulator. The Multiattribute Linear Ballistic Accumulator (MLBA; Trueblood et al., 2014) model is considered an extension of the Linear Ballistic Accumulator (LBA; Brown & Heathcote, 2008) model, and was developed to circumvent perceived issues with previously proposed dynamic models. Specifically, Trueblood et al. (2014) focused on two issues, one exclusive to the MLCA model, and one that is inherent to the MDFT, MLCA, and AAM models. The issue exclusive to the MLCA model centers on the use of a loss aversion function, which was considered inapplicable to stimuli from the perceptual domain. These are based on the fact that Trueblood et al. (2013) found that context effects also occur when perceptual stimuli are used instead of consumer goods. However, it is somewhat difficult to see how a principle such as loss aversion would apply to perceptual stimuli where no ostensible feeling of loss is present. The second global issue centers on computational tractability. As discussed in the introduction, the MDFT, MLCA, and AAM models are all simulation based in that their likelihoods have not yet been derived. Because of these concerns, models such as the MLCA were not considered when assessing MLBA's ability to capture data.

To avoid the issues of computational simulations, the MLBA model relies on the linear and ballistic accumulation process assumed by the LBA model. This reliance makes a strong assumption about how context effects arise. The models discussed above all assume that context effects play out due to a stochastic attentional process across attributes and as a function of the competitive dynamics facilitated by mechanisms such as inhibition and leakage. On the other hand, the MLBA model assumes no moment-by-moment sampling across attributes and conceptualizes preference formation as an independent process by removing mechanisms such as inhibition and leakage altogether. Without these mechanisms, the MLBA model must almost exclusively rely on the models "front-end" which maps the objective attribute values of a set of alternatives to subjective values used in the "back-end" (i.e., the LBA process used to make a choice).

The front-end portion of the MLBA model is responsible for transforming stimulus inputs (i.e., attribute values) into drift rates for the back-end process. Given three options that vary along two attributes, the mean drift rate d_i for each alternative i is defined by comparing each alternative against the other two. These drift rates, once determined, do not vary during the trial. The MLBA model uses five parameters to define mean drift rates: one mapping parameter m, two decay parameters λ_1 and λ_2, one constant input parameter I_0 (i.e., akin to the MLCA model above), and one attribute bias parameter β. The mapping parameter m determines the relationship between intermediate and extreme options though introducing curvature to the attribute space. When $m > 1$, then intermediate options are preferred to extreme ones. When $0 < m < 1$, the extreme options are preferred to intermediate. Finally, when $m = 1$, the curve becomes a straight line and objective values and subjective values are equivalent. The decay parameters, λ_1 and λ_2 determine attention weights through Shepard's (1987) law of generalization, where similarity is considered an exponentially decaying function of distance (Shepard, 1987). If the difference in attribute values is positive, then $\lambda = \lambda_1$ and if the difference is negative, then $\lambda = \lambda_2$. The attention weights are intended to capture the trend that similar options receive more attention than those which are easily distinguished and are captured in a vector as a function of time. The constant parameter I_0 ensures that at least one of the mean drift rates is positive, setting a baseline rate of evidence accumulation for available options and preventing "nontermination" of the LBA back-end process. Finally, β is the attribute bias parameter. When $\beta = 1$, attribute P and attribute Q are considered equally. When $\beta > 1$, there is a bias toward attribute Q; when $0 < \beta < 1$, the bias is toward attribute P. The β parameter is intended to provide further mediation in the attention weights λ_1 and λ_2.

Once the MLBA has explicitly specified drift rates from evaluation of choice stimuli via the front-end portion, the back end component transforms these drift rates into discrete choices through the LBA process. In the LBA model, choice is represented using independent linear accumulators that race toward a threshold in a deterministic manner.

Deterministic accumulation eliminates moment-to-moment fluctuations and renders the model mathematically tractable. However, the model still possesses mechanisms that allow for trial-to-trial fluctuations in starting point and drift rate. Specifically, each accumulator starts at a randomly determined amount of evidence drawn independently from a uniform distribution with the interval $[0, A]$, and the rate of an accumulator on a given trial is determined by sampling from a normal distribution with mean determined by the front-end process and standard deviation determined by the parameter s. Given these settings, the accumulators race to a threshold amount of evidence χ until a choice is elicited that corresponds to the winning accumulator.

Figure 1 represents the MLBA model as three layers. At the first layer, the attribute values are transformed into subjective representations, controlled by the parameter m. Importantly, the MLBA model departs from the other models in that attention does not oscillate across the attribute dimensions. At the second layer, the subjective attribute values are then compared in a pairwise manner and linearly combined according to the attention weights. These process produce a set of drift rates, which are used in the back-end process represented in the third layer. Here, the MLBA model assumes that the accumulation of evidence is independent and linear, assumptions that are represented in Figure 1 by the absence of node-to-node connections.

The attraction effect is accounted for through attention weights, where options that are closer together are given greater attention weights than those far apart on the basis that closer options are more difficult to discriminate. MLBA predicts the similarity effect based on the principle that supportive information is weighted more heavily than unsupportive. This concept is reflected in the decay constants, where if $\lambda_1 < \lambda_2$, negative differences will decay more quickly than positive differences. As a result, positive differences receive more weight leading to larger mean drift rate. This relationship between decay constants is only necessary to produce the similarity effect. Lastly, the MLBA model predicts the compromise effect via the subjective value function, which produces a curvature based on the mapping parameter m. When $m > 1$, midrange options are weighed more heavily than extremes, producing a compromise effect. Through these methods, the MLBA model is able to account for all three context effects (see Trueblood et al., 2014). However, as shown in Tsetsos et al. (2015), the MLBA model accounts for the compromise effect only with alternatives that are not indifferent in binary choice sets.

A Taxonomy for Evaluating the Extant Models

The goal of the current article is to test the descriptive adequacy of the mechanisms assumed by the four extant models outlined above. Because the mathematical composition of the models differs greatly, we introduce a taxonomy to facilitate a comparison across the models throughout the article. As shown illustratively in Figure 1, our taxonomy below consists of three primary processing stages, with six additional subcategories within the various processing stages. While the three processing stages do not necessarily define the temporal structure of the decision process, such a temporal distinction among the processing stages seems psychologically plausible. Instead, the processing stages are conceived with the intent of increasing overlap among the models, allowing us to draw comparisons with higher specificity. The first processing stage in our taxonomy distinguishes models on the basis of how the objective values of the stimuli are mapped to subjective values used in the representation. The second processing stage details how attention is allocated across attribute information. The third processing stage describes how preferences dynamically evolve over time. We now discuss each of these three stages in turn.

1. Subjective Perceptions of the Attribute Space. Bounded rationality presumes that individuals' ability to behave rationally is governed by limitations in the amount of information decision-makers have available, the ability to process that information, and the available time to engage in the processing of said information (Simon, 1982). Theories of context effects (Tversky, 1972; Huber et al., 1982; Simonson, 1989) and broader decision models (Kahneman & Tversky, 1979; Stewart et al., 2006) suggest that judgments and decisions are made by representations of the perceived environment – representations that are assumed to be guided by the perceptual system. It is well established that the perceptual system has limited capacity and often processes perceptual stimuli in ways that are not veridical (Anderson, 1990). Given the limitations of the perceptual system, it seems reasonable that models of decision making should at least allow for the possibility of perceptual distortions. In essence, our first processing stage pins down the assumptions made by the extant models of context effects that allow the models to transform the objective values of the stimuli to subjective representations used in the decision making process.

1.a. Subjective Mapping of Attribute Values The first distinction in our taxonomy is whether or not attribute values are represented veridically in the model. As experimenters, we have access to the objective values that comprise our stimuli. However, these objective values may not necessarily map onto the subjective representations used by observers. While one could argue that using the objective values directly in the subjective representations could be an wise strategy if it maximized some behavioral metric (e.g., accuracy), assuming a perfectly objective representation of the environment could be overwhelmingly computationally costly (Jones & Love, 2011). Indeed, a number of papers have demonstrated that judgments and decisions are based on such subjective representations of the environment rather than the objective values of the stimuli (e.g., Schley & Peters, 2014; Chandon & Ordabayeva, 2009; Frydman & Nave, 2016).

Of the four extant models, only the MDFT and MLCA models assume access to the true objective attribute values. Conversely, the AAM first transforms the objective attribute values to subjective representations through a parameterized power function. Similarly, the MLBA model first applies a non-linear transformation of the attribute values through a parameterized indifference curve. Interestingly, the mapping functions work similarly, allowing the objective values to be compressed or expanded relative to one another. It is worth noting that in both models, the mapping function can produce a perfectly objective representation (i.e., no transformation is performed) under one setting of the parameter values (see Supplementary Materials for mathematical details).

1.b. Attribute Differences The models also differ in how the attribute values are represented in the decision process. The AAM assumes that the choice options are represented in an absolute sense, where the values of the attributes themselves serve as input into the accumulation process described below. Conversely, the other three models models assume that the attribute values can be represented relatively, where all possible pairwise comparisons within the choice set are calculated. For example, when considering the price attribute for Option A, decision-makers calculate the difference between the prices of Options A and B, and Options A and C. These differences are stored in a new matrix and a linear combination of the differences are used as input into the accumulation process.[2] As some models (i.e., the MDFT, MLCA, and MLBA models) operate on pairwise differences, some input terms may be negative. To avoid large negative inputs to the accumulators, some of these models assume a common additive baseline input term to force the options to accumulate positively.

1.c. Non-linear filtering of attribute differences Once the pairwise differences have been calculated, in MLCA and MLBA (but not in MDFT) the sign of these differences becomes an important factor in how they are subjectively evaluated. Essentially, a negative difference represents the amount of an attribute that is “lost” once a decision is made and an alternative is forgone. By contrast, a positive difference represents the amount that is “gained” if a choice is made corresponding to that option. In MLCA each attribute difference is transformed via a non-linear monotonic function (i.e., analogous to a utility function) that incorporates a steeper slope in the negative domain or, equivalently, loss-aversion (Kahneman & Tversky, 1979). The result of this asymmetry is that “losses” (or negative attribute differences) are more impactful than an equivalent gain (or positive attribute difference). For instance, if an attribute $QI_A = 50$ for Option A is lower than for an attribute $QI_B = 90$ for Option B, the loss of 40 when choosing A over B will be more impactful than the gain of 40 when choosing Option A over a lower-valued Option C (i.e., when $QI_C = 10$).

The MLBA model takes a very different approach based not on loss and gains, but on distances in the attribute space. Here, the attribute differences are multiplied by weights that are based on an exponentially decreasing function of the absolute difference between attribute values. As the difference between the attribute values becomes larger, the weights become smaller, such that a small difference is weighted more strongly compared to a large difference. As a result, the mapping between the attribute differences and their transformed counterparts can be non-monotonic (cf. Figure 3 in Tsetsos et al., 2015). Furthermore, the shape of exponential decay can vary depending on whether the distances are positive or negative. As a result of these additional parameters, the MLBA model can incorporate asymmetries in the relative importance of advantages and disadvantages in a way that is similar to loss aversion.

2. Attention to Attributes. Decision-making researchers have long studied the role of limited attention during the preference-construction process (Simon, 1982; Payne, 1976; Johnson & Russo, 1984; Krajbich & Rangel, 2011). Whereas other research has studied the effects of attention to a particular task, attention in this domain is meant to detail how resources are allocated in stimulus processing. In other words, the models assume that observers are always “on task”, but the manner in which they attend to specific details of the stimuli constitutes an interesting distinction among the models.

When making decisions between options consisting of multiple attributes, ascertaining the relative importance of these attributes in the decision-making process is an essential skill. For example, some attributes of the stimuli may be perceptually more diagnostic (Turner, Gao, Koenig, Palfy, & McClelland, 2017), allow one to operate in a goal-directed manner (Dai & Busemeyer, 2014), or be completely irrelevant to the task at hand (Turner & Schley, 2016). Inevitably, an observer must be able to selectively allocate attention across attributes to perform well on a task, or to express consistent preferences. In all of the extant models of context effects, the process of attention allocation is dictated by a set of attribute weights, where larger values translate to greater allocation of attention toward that particular attribute. In the MDFT, MLBA, and MLCA models, the attention weight is either assumed to be equal across attribute dimensions, or in some cases, it is freely estimated from the data. None of these models provide a theoretical basis for how attention should be allocated across the attribute dimensions. By contrast, the attribute weights in the AAM are determined not only by the intrinsic importance that each attribute has, but also by the properties of the stimulus set. Specifically, even when two attributes have equal importance a priori, the attribute dimension with the highest sum of attribute values will receive the largest weight. To illustrate, when choosing between two expensive products, because the products have large values along the price dimension, more attention will be allocated to this attribute dimension compared to a choice between two inexpensive products.

Once attribute weights have been determined, the manner in which the attribute information is integrated into a representation of the item also differs across the models. For the MDFT, MLCA, and AAM models, the allocation of attention follows a Bernoulli process where attribute dimensions are stochastically attended to with probabilities proportional to their respective attention weights. As a result of this stochastic oscillation across attribute dimensions, the valuation of a particular choice option will also oscillate, giving rise to a time-varying input signal to the accumulation process. This particular mechanism – and its interaction with preference accumulation dynamics – plays a pivotal role in capturing certain behavioral patterns, such as the similarity effect. By contrast, the MLBA model assumes that the various attribute dimensions are weighted by their relative importance (i.e., their attention weights) and then linearly combined into a single net input for each choice alternative. While the MLBA model still parameterizes the importance of each attribute dimension, it interprets classic context effects in a way that does not dependent on momentary fluctuations in attention.

3. Preference Accumulation. All of the extant models assume that preference states for each option evolves over time according to the traditions of sequential sampling theory. Namely, each choice is represented as a separate accumulator, and preference (i.e., evidence) is accumulated over time until either a common threshold amount of evidence has been achieved, or a pre-specified length of time (i.e., number of iterations) has passed. The manner in which the options accumulate preference over time is dictated by (1) the input, discussed in the first two processing stages above, (2) the type of competition among the accumulators, and (3) the type of noise in the decision process. We now discuss these last two factors in turn.

3.a. Competition Competition through lateral inhibition among different accumulators is a canonical computation, underlying both perceptual and value-based decisions (cf. Usher & McClelland, 2001) as well as an array of cognitive operations such as visual search, attention and cognitive control. Consider the choice between Options A and B. Effectively, increases in the preference state for Option A will be accompanied by decreases in the preference state for Option B. Although lateral inhibition is a biologically inspired mechanism, signatures of this mechanism can be found in higher level behavioral phenomena as well. For example, research on information distortion demonstrates that if an individual develops preference for Option A early on in their search process, they will denigrate subsequent information about Option B, resulting in increased preference for Option A (DeKay, Miller, Schley, & Erford, 2014). Essentially, accumulating preference for one option inhibits the ability to accumulate preference for another option. Inhibition can lead to winner-take-all dynamics, giving rise to extreme states with the passage of time, in which one option has a very high preference state and all other options have zero preference states. To prevent such runaway processes and to keep preference states bounded, leakage is typically used to offset the effect of inhibition. Leakage implies that the preference state of each accumulator decays over time, driving accumulators to zero preference states in the absence of input.

Leakage and lateral inhibition work in tandem to describe behavioral phenomena. For example, order effects (Hogarth & Einhorn, 1992) involve cases where information sampled earlier during a sequential process exerts proportionally more (i.e., primacy effects) or less (i.e., recency effects) influence on the eventual judgement or decision. Imagine that an individual is considering subscribing to either The New York Times (NYT) or The Wall Street Journal (WSJ). To make their decision, they buy copies of the NYT on Monday, Tuesday, and Wednesday, and buy copies of the WSJ on Thursday, Friday, and Saturday. On days that the individual buys the NYT they accumulate twice as much preference for NYT than for WSJ. On days that the individual buys the WSJ they accumulate twice as much preference for WSJ than for NYT. The difference between the amount of preference will be relatively higher when lateral inhibition is high and relatively lower when lateral inhibition is low. Over time, accumulated preference from the early days will decay (i.e., leakage). For a fixed rate of leakage, if lateral inhibition is high, preference accumulation for the NYT will likely be higher than for the WSJ, producing a primacy effect. Conversely, if lateral inhibition is low, preference accumulation for the WSJ will likely be higher than for the NYT, producing a recency effect.

The MLCA model includes both lateral inhibition and leakage parameters that operate in a manner analogous to the example above. While never investigated to our knowledge, the AAM can also include lateral inhibition and leakage in the same way as the MLCA model – a possibility that we explore in the current article. The MDFT model also includes lateral inhibition and leakage, but assumes that lateral inhibition is based on the distances between alternatives in a transformed choice space (see the Supplemental Materials). This unique assumption in the MDFT model implies that the amount of competition depends on the presented choice set, and so it may vary across conditions when fitting the model to data from an experiment. The theoretical mechanism for calculating lateral inhibition in the MDFT model has proven effective, allowing it to capture both the attraction and compromise effects. For example, reconsider the attraction effect described in the introduction, where Option A is very inexpensive $PI_A = 90$, and of low quality $QI_A = 10$, Option B is very expensive $PI_B = 10$, and of high quality $QI_B = 90$, and Option C ($PI_C = 70$, $QI_C = 10$) is similar to Option A, but worse in both attributes dimensions. According to the MDFT model, the attraction effect occurs because the accumulator for Option C inhibits the accumulator for Option A more than it inhibits the accumulator for Option B. Because Option C is inferior, it ends up with a negative preference state which, via lateral inhibition, boosts the preference state of Option A. Conversely, in the MLCA and AAM models, the amount of lateral inhibition exerted by Option C on Option A is independent of the proximity between the two options. The additional theoretical overhead of the MDFT model provides an advantage in that its lateral inhibition is governed by accessible information in the stimulus set, but has the disadvantage of being strongly tied to its theoretical commitment.

Finally, the MLBA model deviates from the other three extant models on the competition dimension as it assumes the accumulation process is completely independent, and is not prone to the passive loss of information (i.e., leakage). The additional mechanisms of leakage and lateral inhibition severely compromise the tractability of decision making models, and linearized accumulation processes have proven effective in accounting for a range of behavioral patterns across several decision making tasks (Brown & Heathcote, 2008).

3.b. Noisy Preferences The final consideration deals with the inconsistency of preferences across trials. For example, when choosing a beverage from your favorite coffee bar, you may typically have a strong preference for ordering tea early in the week. However, as days go by, your preferences may transition to a cappuccino. The idea of noisy preference accumulation mirrors similar discussions about whether unstable preferences reflect measuring fixed latent preferences with error, or whether preferences are just noisy (Regenwetter, Dana, & Davis-Stober, 2011). The MDFT, MLCA, and AAM models all assume momentary fluctuations in the preference accumulation process. Mathematically, the process of momentary fluctuations is carried out by a stochastic process called the Wiener process. Here, although the input to the accumulation process may be stable, the exact preference state of the options is volatile, fluctuating around the mean input from moment to moment.

By contrast, the MLBA model assumes a ballistic accumulation process, meaning that evidence accumulates in a linear fashion without moment-to-moment noise. As discussed above, ballistic accumulators have enjoyed widespread success by virtue of their computational efficiency compared to models that use the Wiener process (Brown & Heathcote, 2008). To account for inconsistent preferences, the MLBA model instead assumes that the input itself is corrupted by noise, such that the "drift rate" on each trial will be different despite having the exact same input. Here, we can make a distinction between models that use within-trial variability (i.e., the MDFT, MLCA, and AAM models), and those that use only between-trial variability (i.e., the MLBA model). While some models of perceptual decision making have used a combination of within- and between-trial variability, models of context effects have yet to include both sources of variability in a single model.

Summary

In this section, we provided a summary of the most robust context effects and the set of extant models that have been shown to capture these effects. We then provided a taxonomy that allowed us to compare and contrast the four extant models of context effects. We organized the assumptions of the four models on the basis of (1) the processing of objective stimulus information, (2) the allocation of attention, and (3) the dynamics of the accumulation process. Along just these three dimensions, the models vary considerably in the particular assumptions they make to effectively reproduce patterns in behavioral data. However, the models also have some key similarities. For example, all of the models build into their representations some form of context, motivated by the properties of the stimulus set (Medin & Schaffer, 1978). In the AAM, context is represented by the degree of association between items, a property than influences how attention is allocated at each moment in time. In the MDFT model, context is built up on the basis of distances between items in the stimulus set, which affects both the input to the accumulators and the strength of lateral inhibition. Similarly, the MLCA and MLBA models construct context via pairwise differences among the stimulus set, where the MLCA model includes a filtration mechanism through loss aversion, and the MLBA model uses the distances among the stimuli to establish the attention weights.

Study 1: Data from Berkowitsch et al. (2014)

Consumer choice data from Experiment 2 in Berkowitsch et al. (2014) were used to fit the four models. The study consisted of two sessions in which 48 participants saw different alternatives (sampled from 6 different product types) characterized by two attributes. In the first session, participants saw the two alternatives but one attribute value was missing for one of them. Participants had to fill in the missing value so as the two alternatives were matched in terms of subjective value. Session 2 involved choices between 3 alternatives. On each trial, two out of three alternatives corresponded to the matched alternatives from Session 1. The third alternative was a decoy. The attraction, similarity and compromise effects were tested by varying the location of the decoy. In their analyses, Berkowitsch et al. (2014) report a strong attraction effect, a strong compromise effect, and a weak similarity effect. Detailed information about the stimuli and procedure is offered in Berkowitsch et al. (2014). In order to fit the models, we used the relative choice share of the target (Berkowitsch et al., 2014) in the three effects, averaged across participants. The aggregated data are presented in Figure 2 as the black "+" signs, shown in ternary form, where the target (T), distractor (D), and competitor (C) options are shown as vertices in each triangle for the three context effects (columns).

Modeling Analysis 1: No Individual Differences

In total, we fit 10 models to the data. For MDFT, MLCA, and MLBA, we fit two variants. The first variants assumed that there were no biases across the two attributes, variants we refer to as 1.0. The second variants assumed biases in the processing of the two attributes, which involved one additional parameter for each of these three models. We refer to these variants as 2.0. For AAM, we investigated four model variants. In total, we report four different variants of the AAM. The first variant, AAM 1.0, is the stock version of the model that assumes each attribute dimension is preferred equally and has no lateral inhibition component. Although not explored in Bhatia (2013), the second variant, AAM 2.0, allows for the additional mechanism of lateral inhibition (as used in the MLCA model). The third variant, AAM 3.0, only allows for biases in the attribute processing, but is otherwise equivalent to AAM 1.0. The fourth variant, AAM 4.0 allows for both biases in the attribute processing and lateral inhibition. In counting the number of parameters, AAM 4.0 has six total parameters, which is the most of any of the models we investigated. The specific implementation details of each model are reported in the supplementary materials.

We used the probability density approximation (PDA; Turner & Sederberg, 2014) method, to approximate the likelihood function for each model. To sample from the posterior distribution, we used differential evolution with Markov chain Monte Carlo (DE-MCMC; ter Braak, 2006; Turner, Sederberg, Brown, & Steyvers, 2013). The specific details of the sampling process are provided in the supplementary materials.

Results. We present the results in two stages. First, we compare the models on the basis of fit statistics, both across models and across variants. Second, we show predictions from the best-fitting variant of each model against the observed data. This evaluation reveals which of the three context effects are most difficult for each model to capture.

Fit Statistics We evaluated the model fits to data in two ways. First, we compared the models on the basis of the deviance information criterion (DIC; Celeux et al., 2006), which balances the number of model parameters, degree of model fit, and degree of model complexity.[3] Table 2 shows the resulting DIC values obtained for each model (rows), in the third column. Table 2 also includes the effective number of parameters p_D in the fourth column. This metric is a simple evaluation of how much the mean of the approximated likelihood deviates from the best likelihood value obtained. Large values indicate that the posterior distributions are less constrained and many values produce a reasonably close fit to the data, whereas smaller values indicate that the range of acceptable values in the posterior are fewer, which can be interpreted as better constraint. Another way of interpreting p_D is that larger values indicate model fits that are more complex. The fifth and sixth columns show the average of the log likelihood values $\bar{D}$ and the log likelihood value obtained at the best fitting parameter value $\hat{D}$.

Table 2 shows each model variant's fit statistic. For the MDFT, MLCA, and MLBA models, adding the attribute bias parameter improved the model fits to data. However, comparing the number of effective parameters, we see that only small differences between versions 1.0 and versions 2.0. For the MLBA and MLCA models, this difference is positive meaning that p_D decreases, whereas for MDFT this difference is negative.

For the AAM, adding the lateral inhibition parameter l alone actually performed worse than the base model which constrained this parameter to zero. On the other hand, adding the attribute bias parameters a_0^1 and a_0^2 greatly improved the fits to data, even after the increase in the number of parameters was taken into account. Perhaps more interesting is that with the attribute bias parameters in place, the addition of the lateral inhibition term further improved the model fit statistic, and in a way that is less complex than the other models. While this result seems paradoxical, it stems from the fact that the model is better able to capture the full range of data with these additional mechanisms, which directly affects the value of p_D.

Comparing across the four models, Table 2 shows that the best fitting model is MDFT 2.0, and a close second is AAM 4.0. The MDFT 1.0 model also performs well, even better than any of the MLBA or MLCA models we tested. Comparing MLBA to MLCA, we see that both versions of MLCA perform better than the best performing MLBA model, MLBA 2.0. It is interesting that the variants of AAM span the range from performing worst in the entire group (i.e., AAM 1.0) to performing second best (AAM 4.0). We consider this a testament to our model evaluation approach – had we not tested whether or not lateral inhibition should be present with attribute bias (i.e., AAM 4.0), we might have wrongly concluded that AAM simply performs worst than say, the MLCA 2.0 model.

Fits to Data Although the fit statistics in Table 2 are useful in describing which model best accounts for the entire data set, it is also interesting to examine how the models compare on a particular context effect. To this end, we generated predictions from the best-fitting variant of each model by randomly sampling 1,000 draws from the estimated posterior distribution, and simulating data from each model. Figure 2 shows a ternary plot of the model predictions (gray contours) against the observed data (black "+" sign). A ternary plot expresses the probabilities of choosing the target (T), distractor (D), and competitor (C) options within a single figure. For example, if the data were such that all three options were chosen equally often, the black "+" would lie on the centroid of the triangle, represented as the intersection of the dashed gray lines. However, if the data were such that the target was chosen 100% of the time, the black "+" would lie on the vertex labeled "T". The columns in Figure 2 correspond to the three context effects: attraction (left), compromise (middle), and similarity (right). The rows in Figure 2 correspond to the four models: MDFT (first), MLCA (second), AAM (third), and MLBA (fourth).

The outer border of the gray contours represents the 95% credible set of the predictions from each model. Comparing across panels, each model seems to do a reasonably good job in capturing the basic probabilities for each effect, but there are some clear discrepancies. First, for the attraction effect, each model makes predictions that are very close to the observed data, where the predictions for the MLBA model are noticeably farther from the rest of the predictions. For the compromise effect, each model makes predictions that are at least consistent with the data in that the predictions do fall into the correct shaded area. However, the MLBA model is again noticeably the worst in predicting this effect. Finally, for the similarity effect, the AAM, MDFT and MLBA models make predictions that are all consistent with the data. However, the MLCA model is noticeably worse, making predictions that are in the wrong shaded area. It is worth noting that the similarity effect is the weakest of the three effects in these data, whereas the attraction effect is the strongest. The strength of these effects may be what drives the fitting results in Table 2. For example, the MLCA model may adjust its parameters to make predictions that are consistent with the strongest effect, and this adjustment may be what causes the misfit in the similarity effect.

Summary and Conclusions. In this section, we fit the four models to the data from (Berkowitsch et al., 2014). Within each model, we tested a few variants to examine the role of the attribute bias for these data. Across all four models, attribute bias parameters improved the model fit, even beyond the penalty terms that were applied for increasing the number of parameters. Having an attribute bias parameter allowed the models to weigh the attribute dimensions unevenly, which ostensibly afforded the models enough flexibility to fit the data properly. However, it is interesting that the particular dimension that the models overweighted was inconsistent across the four models. We suspect this has to do with aggregating across consumer goods and so it isn't of any particular interest. However, this inconsistency does have some implications for interpretability when multiple models are fit to data.

Given that the MLBA model performed worst, it suggests that some element of stochasticity improves the model fits to data. However, the best-fitting versions of AAM, MDFT, and MLCA models all also have some dynamic component such as lateral inhibition or leakage. For the MDFT model, while leakage is fixed across different stimuli, lateral inhibition is dictated by properties of the stimulus set (i.e., their proximity). For the AAM and MLCA models, however, the leakage and lateral inhibition parameters, while being freely estimated, are assumed to be fixed across different stimulus inputs. Regardless, the interplay between stochasticity, lateral inhibition, and leakage did provide an advantage in the model fits.

Modeling Analysis 2: Hierarchical Models

In our second analysis of the data from Berkowitsch et al. (2014), we investigated the role that subject-to-subject variability played in the model fits. By aggregating across subjects in our Analysis 1, it is possible that important patterns in the behavioral data are obscured (Heathcote, Brown, & Mewhort, 2000; Davis-Stober, Park, Brown, & Regenwetter, 2016). In order to appreciate these subject-specific patterns, we employed hierarchical versions of the models discussed in our Analysis 1 above. Hierarchical models are powerful in the way they partition subject- and group-specific effects, allowing information from one subject to constrain what we learn in other subjects (e.g., Shiffrin, Lee, Kim, & Wagenmakers, 2008; Lee & Wagenmakers, 2013; Turner, Dennis, & Van Zandt, 2013).

In total, we fit eight hierarchical models to the data. We refer to these models as "HX", where the "X" corresponds to the acronym of the model under discussion. In parallel with our nonhierarchical analyses above, we fit two variants of each model: variants 1.0 assumed no biases in the processing of attributes, whereas variants 2.0 allowed for this possibility. To fit the models to data, we relied on a custom algorithm similar to what was used in Turner, Dennis, and Van Zandt (2013). To estimate the likelihood function for the subject-level effects, the PDA algorithm was used (Turner & Sederberg, 2014). To estimate the group-level effects, we used the Gibbs ABC algorithm (Turner & Van Zandt, 2014). Finally, to sample from the full posterior, we used the DE-MCMC algorithm (Turner, Sederberg, et al., 2013). Specific details about the specification of the models and the algorithms used to fit them to data are provided in the supplementary materials.

Results. We present the results in two stages. First, we compare the models on the basis of model fit statistics. Second, we show the posterior predictive distribution obtained by the best-fitting model for each subject as a way to assess the model's ability to capture the patterns in the behavioral data.

Table 3 shows the model fits statistics for each of the models we fit to the data. Across panels, Table 3 shows that the symmetrical models (i.e., versions 1.0) fit the data better than their asymmetrical counterparts (i.e., versions 2.0). These results are at odds with the results of our Analysis 1 in that adding parameters to allow for asymmetric processing of the attribute space seems to be an overly complex explanation of the data. Instead, simply allowing for individual differences in the decision making process provides a better account of the data for all models we investigated. Comparing across models, Table 3 shows that the HAAM 1.0 model fit best, with HMLCA 1.0 coming in second, HMLBA 1.0 coming in third, and HMDFT 1.0 coming in fourth. With the exception of the MDFT model, these results are consistent with our Analysis 1.

Next, we can evaluate the model fits to data visually by plotting the model predictions against the observed data, in an analogous manner to that of Figure 2. Figure 3 shows the posterior predictive distribution collapsed across subjects (yellow clouds) against the observed data (black dots). To generate the posterior predictive distribution, we first generated predictions from the model for each of the subjects in the data, creating a distribution of predicted probabilities for that subject's data in each of the three context effect conditions (columns of Figure 3). We continued generating predictions from the model by randomly selecting parameter values in that subject's posterior distribution. Finally, once a posterior predictive distribution had been created for each subject, we collapsed across this subject-level information so that a pattern for the observed data could be assessed. Starting with the attraction effect, Figure 3 shows that while the AAM and MLCA model capture both the shape and spread of the distribution of data, the MLBA model's predictions are slightly more variable relative to the data, and the MDFT model's predictions seem to have both the wrong mean and wrong variance. Specifically, MDFT seems to miss the more extreme subjects who exhibit the strongest attraction effects. The compromise effect best illustrates why the MDFT model does not capture the individual subject data well. While all of the other model models capture the extreme compromise effect exhibited by a few subjects, the MDFT model's predictions simply cannot produce compromise effects that are large enough to match the data. The similarity effect data best illustrates why the MLBA model does not do as well as the MLCA model and the AAM. Both the MLCA model and the AAM make conservative predictions about the magnitude of the similarity effect, whereas the MLBA makes predictions about the target probability that are too large relative to the data. This diffuse spread of the MLBA model's predictions makes the model fit less accurate for a greater number of subjects, causing the MLBA to do worse on these data.

Summary and Conclusions. After fitting two versions of each hierarchical model to data, we found that across the board, bias parameters that allow for asymmetric evaluation of the attribute space were unnecessary to capture the patterns in the individual subject data. Instead, the inclusion of these additional parameters made each variant of the models we investigated too complex once penalty terms were applied. This result is particularly at odds with the results of our Analysis 1, that did not take into account individual differences in decision making. After evaluating how well each model fit the data, we found a similar ordering of model performance as in our Analysis 1, with the exception of MDFT. After investigating the model predictions by context effect, we concluded that the MDFT was unable to capture the extreme compromise effects exhibited by some of the subjects in the data, and was penalized severely for this fault. We suspect that these differences in the results are a testament to the utility of the hierarchical models we used here. It seems that by aggregating the data, we reduce variability in ways that obscure the interpretation of the decision processes underlying individual choice behavior (Heathcote et al., 2000; Davis-Stober et al., 2016; Liew, Howe, & Little, 2016).

Study 2: A Perceptual Choice Experiment

As reviewed in the introduction, context effects in choice have been discussed as a common consequence of analogous contextual effects in perception (Tversky & Simonson, 1993; Parducci, 1965). If we assume that features of perception underlie the construction of preference (Kahneman & Tversky, 1979; Amir & J., 2008; Slovic, 1995; Schley & Peters, 2014; Chandon & Ordabayeva, 2009), then assessing context effects in the perceptual domain could provide unique insight into these shared choice mechanisms (Trueblood et al., 2013). In our Study 2, we included two manipulations to better assess and constrain the models we have been investigating here. First, the stimulus set consists of both binary and ternary trials. Second, the stimulus set consists of items that should reproduce the classic context effects discussed in the introduction, but we also included several "filler" items where one alternative in the choice set should be dominant (in terms of preference) relative to the others. The motivation for including these filler items is to avoid potentially overfitting the classic context effects. By enforcing that each model must account for basic patterns of preference (i.e., assuming that subjects are consistent in their choices) as well as the full ensemble of context effects, we can better assess the relative generalizability of the models (Pitt, Myung, & Zhang, 2002).[4]

Experiment

Participants. Fifty-six participants (61% female, mean age of 21.6 years) were recruited from a paid subject pool at Erasmus University in exchange for 20 euros. Participants completed two sessions, each on separate days, and were paid contingent upon completing both sessions to decrease attrition. Participants could not sign up for each session individually, instead they made appointments for both Session 1 and Session 2 before arriving for their first session.

Stimuli. Following Trueblood et al. (2013), the stimuli consisted of rectangles presented on a screen. The specific values of the rectangle dimensions are presented in Experiment 2 are given in Table 4 (also see the supplementary materials for a visualization in attribute space). Table 4 shows that a total of 15 stimuli were used throughout the experiment, and each subject experienced these stimuli with a consistent number of trials (i.e., the last column of Table 4). Table 4 shows that a mixture of binary and ternary trials were used. The values of the width (i.e., D1) and height (i.e., D2) attributes are provided in Table 4 on a virtual 0-10 scale, but these values were exponentiated and a random sample from a uniform distribution (i.e., from the interval [0-3] in pixels) was added to each dimension to produce the rectangles' final dimensions in pixels so as to make them an appropriate size.

Experimental Design. All features of the experimental design were within participant. The first within-participant factor involved manipulating whether the choice set was binary, which will serve as our control condition, versus ternary. Because participants completed two separate sessions, we manipulated within-participant conditions such that participants would not complete the binary and ternary choices of a particular type of stimulus in a single session. To do so, participants were randomly assigned to one of 4 within-participant-condition orders, such that number of options (binary or ternary), and session (session 1 or 2) were completely counterbalanced. On each session, the binary conditions always had 450 trials, and the ternary conditions always had 495 experimental trials. The binary trials were divided in 7 experimental conditions while the ternary trials in 8 conditions (see Table 4 for perceptual trials).

When presenting the stimuli, the position of the attributes (i.e., top or bottom row) and alternatives (i.e., left, middle, right) were counterbalanced across trials. The alternatives were placed in the left, middle or right of the screen at random, and the horizontal location of the alternatives was counterbalanced across trials. The vertical positions of the rectangles were jittered on each trial by adding random zero-centered

Gaussian jitters with pixel standard deviation of 10.

Procedure. Upon arriving, participants were kept in a waiting room until their session began. Participants were brought back to the lab space one at a time. The lab space consisted of 8 completely sealed rooms, each 100 centimeters wide and 168 centimeters deep. The rooms were fully enclosed with reasonable sound insulation. Participants were seated between 50 and 70 centimeters from the screen, which was 55.5 centimeters diagonal with a 16:9 screen ratio and 1920×1080 screen resolution. After sitting the participant in the room and starting the program, the research assistant returned to the waiting room to start the next participant. When all of the participants in a given block were initialized, the research assistant waited in a separate control room where they could monitor participants in the 8 experimental rooms.

The experimental session began with a welcome message stating that they would be completing two tasks in this session, that the session would last for around 45 minutes on average, and that they would be required to complete the two sessions to receive payment. Participants then read instructions about their first task, the specific task was determined by randomly assigning the participant to one of the 4 condition orders discussed above. Subjects were instructed that their task was to identify which of the presented rectangles had the largest area.

On the following screens, participants were asked to place their dominant hand on the arrow keys. In the binary-choice conditions, the left and right arrows corresponded to the options on the left and right sides of the screen. In the ternary-choice conditions, the up arrow indicated choice of the option in the middle of the screen. Although participants were instructed to only use the arrow keys, participants could press any other keys on the keyboard to advance the page and that key would be recorded in the data. After reading the instructions, participants completed 25 practice trials. These 25 trials were randomly sampled without replacement from the experimental stimuli. After completing the practice trials, participants then completed the 450, from the binary-choice conditions, or the 495, from the ternary-choice conditions, experimental trials. For both practice and experimental trials, each trial had a 500 millisecond inter-trial interval as well as a 10 second maximum duration. We did not use a fixation stimulus between trials to avoid potential confound between attention and preference (Krajbich & Rangel, 2011). Instead, during the inter-trial interval the background of the screen changed to a lighter gray. This 500 millisecond lightening signaled to participants the end of the previous trial and the beginning of the subsequent trial.

Upon completing the experimental trials, participants were thanked and were instructed to take a brief break before continuing to the second task. Participants were then presented the instructions corresponding to the second condition from their random assignment. Participants again completed 25 practice trials followed by the 450 or 495 experimental trials in the second task. The experimental session concluded by thanking participants, asking simple demographic questions, and reminding participants that they would have to complete the second session to receive payment. Participants returned between 1 day and 8 days after the first session to complete the second session. The procedure during the second session was nearly identical to the first session. In total, participants completed 945 experimental choice trials over the two days. Participants were paid electronically within 7 days of completing their second session.[5]

Analysis of Raw Data

The raw data, aggregated across subjects by partitioned into the 15 stimulus items, are shown in the supplementary materials. In the sections that follow, we analyzed the raw data in a consistent manner with previous studies, and report the strength of each context effect.

Quantification of context effects. Our design involved both critical trials (corresponding to the attraction, similarity and compromise effects) as well as filler trials aiming to maximally constrain the estimation of the different model parameters.

Condition B1 in Table 4 offered baseline choice probabilities for the attraction and similarity effects. In particular, we calculated the attraction effect as the average of two measures: $\frac{P(A|C1)}{(P(A|C1)+P(B|C1))} - P(A|B1)$ (i.e. A is the target and B is the competitor) and $\frac{P(B|C2)}{(P(A|C2)+P(B|C2))} - P(B|B1)$ (i.e. B is the target and A is the competitor). The similarity effect was the average of $\frac{P(B|C3)}{(P(A|C3)+P(B|C3))} - P(B|B1)$ (i.e. B is the dissimilar target and A is the competitor) and $\frac{P(A|C4)}{(P(A|C4)+P(B|C4))} - P(A|B1)$ (i.e. A is the dissimilar target and B is the competitor). Finally the compromise effect was the average of $\frac{P(C|C5)}{(P(A|C5)+P(C|C5))} - P(B|B2)$ and $\frac{P(C|C5)}{(P(B|C5)+P(C|C5))} - P(B|B3)$.

Attraction effect. The magnitude of the attraction effect was positive ($M = 0.054, SE = 0.010, 95\%$ CI $[0.033, 0.075]$) and deviated significantly from 0 ($t(41) = 5.246, p < 0.001, d = 0.809$). This finding replicates previously reported attraction effects using the same perceptual stimulus (Trueblood et al., 2013; Trueblood, Brown, & Heathcote, 2015; Farmer, Warren, El-Deredy, & Howes, 2016).

Similarity effect. The similarity effect was significantly lower than 0 ($M = -0.036, SE = 0.011, 95\%$ CI $[-0.058, -0.014], t(41) = -3.340, p = 0.002, d = -0.515$). Significant negative similarity effects have not been previously reported using rectangles or any other stimuli. Nevertheless, the configuration of the alternatives in our study differs from other studies using perceptual stimuli (Trueblood et al., 2013, 2015). In particular, the added alternative in conditions C3, C4 (Table 4) was inlying to the target and the competitor. In (Trueblood et al., 2013, 2015) the similarity decoy was outlying relative to the target and competitor alternatives, extending thus the range of values in both dimensions. Our design compares to another study employing consumer products, in which inlying decoys did not yield a significant similarity effect (Noguchi & Stewart, 2014). It is thus conceivable that our failure to obtain the similarity effect is driven by the placement of the decoys in-between the target and the competitor.

Compromise effect. The compromise effect was not significantly different from 0 ($M = 0.017, SE = 0.026, 95\%$ CI $[-0.036, 0.070], t(41) = 0.514, p < 0.001, d = 0.102$). In two previous studies employing rectangles, the compromise effect was not significant either, although, in contrast with our findings here, the corresponding p-values were close to 0.05 (Trueblood et al., 2013, 2015). In one case this weak effect was characterized by the authors as "fairly consistent" and significant if a one-tailed t-test were applied (Trueblood et al., 2013). However, in these studies the compromise effect was reported as the average of two ternary conditions, and this way of quantifying the compromise effect does not rule out the possibility that the reported trend was an artifact.

Consider four alternatives A, B, C, D, ordered on an indifference line in a 2 dimensional space (with A being on the left end and D on the right end of the indifference line). Trueblood et al. quantify the compromise effect as the average of

$$P(B|A, B, C) - P(B|B, C, D) \quad \text{(compromise-1), and}$$

$$P(C|B, C, D) - P(C|A, B, C) \qquad \text{(compromise-2).}$$

A compromise effect that is positive in both conditions unequivocally constitutes a context effect. However, if compromise-1 is positive and compromise-2 is negative (or vice versa), but the absolute magnitude of the former is larger than the absolute magnitude of the latter, then an artifactual compromise effect would ensue on average. To further illustrate, assume a rational agent whose subjective values of the 4 alternatives are stable across choice-sets: $v_A = 0, v_B = 40, v_C = 20, v_D = 15$. We further assume a simplified Luce choice rule to map subjective values onto choice probabilities: $P(X|X, Y, Z) = \frac{v_X}{v_X + v_Y + v_Z}$. Using the above subjective values compromise-1 has a magnitude of 0.133 and compromise-2 a magnitude of -0.067. Averaging thus across the two condition results in a positive effect. In this example the positive context effect is spurious because the subjective values of the observer do not change as a function of the choice set. The way we quantified the compromise effect, obtaining baseline preferences in binary trials, is not subject to this limitation. We thus hold that the lack of compromise effect in our study is not surprising given that this effect was not unequivocally demonstrated in other studies using the same stimuli.

Correlations between effects. The attraction effect was negatively correlated with the similarity effect ($r = -0.442, p = 0.003$) and negatively (but weakly) correlated with the compromise effect ($r = -0.294, p = 0.058$). There was no correlation between the similarity and compromise effects ($r = 0.034, p = 0.831$). Previous studies have also found a negative attraction-similarity correlation but also a positive correlation between attraction and compromise effects and a negative correlation between similarity and compromise effects (Berkowitsch et al., 2014; Trueblood et al., 2015), which we did not obtain.

Modeling Analysis 1: No Individual Differences

We fit eight models that were used in our Study 1 with no individual differences. For MDFT, MLCA, and MLBA, we fit the 1.0 and 2.0 variants, whereas for the AAM, we fit the 2.0 and 4.0 versions. Recall that for MDFT, MLCA, and MLBA, variants 1.0 all assume that there are no biases in the processing of attributes, whereas variants 2.0 allow for this possibility. For AAM, variant 2.0 assumes no biases in the processing of attributes, whereas variant 4.0 allows for this possibility; however, both AAM variants include the lateral inhibition term. Although we fit each model to the full set of 15 stimuli, no additional parameters were needed beyond the versions of the models that were reported above. To fit the models to data, we used the same algorithm as in the no individual differences analyses from Study 1. Additional details of the model specification and the sampling algorithm are provided in the supplementary materials.

Results. We evaluated the models in two ways. First, we compared the relative fits of the models via the DIC measure. Second, we qualitatively compared the model fits to the observed data by generating predictions from the best-fitting model parameters. The second comparison allows us to evaluate the fits of the models on each of the perceptual stimuli separately, whereas the first comparison collapses across the stimulus set.

Table 5 shows the fit statistics for each of the eight models fit to our data. The second column shows the DIC statistic, the third column shows the effective number of parameters p_D, the fourth column shows the average deviance $\bar{D}$, and the fifth column shows the best deviance value obtained. Across the models, Table 5 shows that the AAM variants performed best, then the MLCA variants, then the MLBA variants, and then the MDFT variants. This same ordering happens whether one considers the DIC statistic, which takes into account model flexibility or just the best log likelihood obtained $\hat{D}$, which would be the critical measure in frequentist-type model comparisons. Comparing within model variants, there are mixed results for whether or not the asymmetric version should be preferred to the symmetric version. For example, for the MDFT and MLBA models, the symmetric versions provided better fits, whereas for the AAM and MLCA models, the asymmetric models provided better fits. So, for the aggregate level at least, there is no clear guide as to whether parameters that allow for attribute biases should be used in perceptual experiments.

Once the models had been fit to the data, we estimated the MAP for each model parameter of the best fitting model variant. We generated 1,000 predictions from each model for the entire set of data (i.e., a total of 15 experimental cells) by simulating the model with the MAP estimate. We then averaged the model predictions to obtain an average prediction from each model for each stimulus item. The Supplementary Materials provide plots showing these average predictions for each item against the data so that one can better assess the models' performance on each item. In general, these plots revealed that each of the best-fitting model variants could produce predictions that were at least somewhat consistent with the data. As these plots were not particularly diagnostic, we chose to collapse across items and compare the set of predictions against the observed data. Figure 4 shows the model predictions across all items (y-axis) against the observed data (x-axis) for each of the four models: the MDFT (top left), MLCA (top right), AAM (bottom right), and the MLBA (bottom right). Each point in Figure 4 corresponds to a response probability for a given choice set, aggregated across subjects in the experiment.

Comparing across the panels, there does not appear to be any systematic differences in predictions vs. observed data across the models, although the predictions from MDFT do seem to be more variable. When calculating the correlations between the model predictions and the observed data, the AAM model obtained the highest correlation ($r = 0.975$), the MLCA model the second-highest correlation ($r = 0.97$), the MLBA model the third-highest correlation ($r = 0.958$), and the MDFT the fourth-highest correlation ($r = 0.909$). These correlation analyses produce the same ordering of model performance as was observed in the DIC calculations reported in Table 5.

Summary and Conclusions. In this section, we fit a total of eight models to the perceptual data from a new experiment. For each model, we investigated models that either included or did not include attribute dimension biases. Our results were mixed in that some models performed better when attribute biases were allowed (e.g., MLCA and AAM), whereas other did not (e.g., MDFT and MLBA). Considering the model fits to data, the AAM with lateral inhibition fit best, followed by the MLCA model, followed by the MLBA model, followed by the MDFT model. When generating predictions from the model and correlating them with the observed data, we observed an identical ordering of results, further substantiating this particular ordering of model performances for our perceptual data.

Modeling Analysis 2: Hierarchical Models

To be consistent as possible with Study 1, we investigated identical versions of each model as was described above. We fit two variants of each model: one variant that assumed asymmetric weighting of the attribute dimensions (i.e., the 2.0 variants), and one variant that assumed symmetric weighting (i.e., the 1.0 variants). Because these models are identical to the ones formulated in Study 1, we do not reproduce the model specifications here. Details about the sampling algorithm used are provided in the supplementary materials.

Results. In parallel with the hierarchical analyses in Study 1, we present the results in two stages. First, we compare the models on the basis of model fit statistics. Second, we show the posterior predictive distributions obtained by the best-fitting model for each subject, collapsed across stimuli and effect. The supplementary materials show predictions from the best fitting model for each of the 15 stimuli used in the experiment.

Table 6 shows the model fits statistics for each of the models we fit to the data. Across models, Table 3 shows that the asymmetric models (i.e., versions 2.0) fit the data better than their symmetric counterparts (i.e., versions 2.0). In contrast to our results from Study 1, Table 6 suggests that some form of biased attention weighting is essential to capture the patterns in the behavioral data. Across models, the results suggest that HAAM 2.0 performed best, HMLCA 2.0 performed second best, HMLBA 2.0 performed third best, and HMDFT 2.0 performed worst. Once again, these conclusions are on the basis of the DIC measures, which take into account model complexity, yet similar conclusions would be drawn by just considering the log likelihood values at the best-fitting parameter values.

Although the HAAM 2.0 model provided the best fit to the data at the aggregate level, we can also examine how well the models fit each individual subject. In the supplementary materials, Table 2 provides the DIC statistic for each subject separately for each of the four best-fitting models. Going across the subjects, we also tabulated how many times each model provided the best fit. By this analysis, the HAAM 2.0, HMLCA 2.0, and HMLBA 2.0 models provided the best fit for 12 subjects each, and the HMDFT 2.0 model provided the best fit for 6 subjects.

Next, we can evaluate the model fits to data by plotting model predictions against data from the experiment. For this analysis, we obtained the best-fitting model parameters (i.e., the MAP estimates) for each subject from the best-fitting model. Using the MAP estimates, we simulated 1,000 predictions from each model for each of the 15 stimulus values used in the experiment. We then averaged these 1,000 predictions to obtain model predictions for the response probabilities for each item. The supplementary materials show the model predictions against the data from the experiment for each stimulus. These plots suggest that each of the models accounted reasonably well for the data, and so we collapsed across stimuli and plotted the model predictions (y-axis) against the data (x-axis) in Figure 5 for each of the four models: the HMDFT 2.0 (top left), HMLCA 2.0 (top right), HAAM 2.0 (bottom right), and the HMLBA 2.0 (bottom right). Each point in Figure 5 corresponds to a response probability for a given choice set, for each subject in the experiment. Comparing across the panels, there does not appear to be any systematic differences in predictions vs. observed data across the models, although the predictions from HAAM 2.0 seem to miss the observed data on occasion. The correlations between the model predictions and observed data revealed that HMLCA 2.0 obtained the highest correlation ($r = 0.869$), HMDFT 2.0 obtained the second-highest correlation ($r = 0.86$), HMLBA 2.0 obtained the third-highest correlation ($r = 0.856$), and HAAM 2.0 obtained the fourth-highest correlation ($r = 0.829$). These correlation analyses produce an ordering that is different from the DIC statistics reported in Table 6.

Finally, we can compare the models' predictions for the strength of each context effect relative to the data. For example, Liew et al. (2016) have shown that different clusters of decision making patterns can appear across the three classic context effects. For example, Liew et al. (2016) have shown that few people consistently exhibit all three context effects simultaneously (also see Trueblood et al., 2015), and so individual differences – and the ability of models to capture these differences – can potentially be an elucidating feature in the model evaluation process. Using the posterior predictive distributions from Figure 5 (also see the supplementary materials), we can calculate the strength of each context effect by calculating the response probabilities predicted by the models for some subsets of the stimulus set in the same way that we did for the data (see description above). Figure 6 shows the joint distribution of each effect for each model (yellow clouds) against the data (black dots). The rows correspond to the predictions from the best fitting hierarchical versions of the models: HMDFT 2.0 (first row), HMLCA 2.0 (second row), HAAM 2.0 (third row), and HMLBA 2.0 (fourth row). The columns correspond to the various joint distributions of context effects: similarity vs. compromise (first column), similarity vs. attraction (second column), and attraction vs. compromise (third column). In each panel, dashed lines represent points of zero effect as a guide. Figure 6 shows that all four models predict, to some extent, all three effects. The important test is whether these predictions have similar statistical properties to those observed in the data. For example, the MDFT model appears to make strong predictions for the attraction effect, strong predictions for a negative similarity effect, and only weak predictions for the compromise effect, relative to the data. Other models like the AAM predict strong similarity effects; effects that are not as apparent in the observed data. On the other hand, the MLBA model predicts relatively weak context effects, with one possible exception being the similarity effect. The strength of the model predictions relative to the data, in conjunction with the filler trials, are inevitably what give one model a better fit than another in Table 6.

Summary and Conclusions. In this section, we reanalyzed our perceptual data by fitting eight hierarchical models. In parallel with our previous analyses, we investigated versions of the models that allowed for biases in the attribute dimensions, and versions that assumed attributes were weighted equally. Here, our results suggest that all models performed better when attribute dimension biases were allowed for each subject, a result that is at odds with our results from the consumer goods task. One reason for this might be that the perceptual system is biased to perceive the vertical dimension as being larger than the horizontal dimension, a prediction that is confirmed across models. By contrast, averaging across the consumer goods in Study 1 could have rendered the dimension information less diagnostic in the model fits. Considering the ordering of the fits of the hierarchical models, the AAM model fit best, followed by the MLCA model, followed by the MLBA model, followed by the MDFT model. This ordering of hierarchical models parallels the ordering of the non-hierarchical models from our Analysis 1. When correlating the model predictions against the observed data, a different order of results were obtained where the MLCA model performed best, followed by the MDFT model, followed by the MLBA model, followed by the AAM. However, as the differences in the correlations across the models are quite small (i.e., a 0.04 difference in correlation between the best and worst performing models), we do not believe these results to be inconsistent with our model fitting results.

Study 3: A Switchboard Analysis of Model Mechanisms

Echoed from the introduction, the extant models of context effects have explored a variety of mechanisms in their attempt to capture patterns present in experimental data. Each model takes a stimulus set comprised of attribute values as an input, and generates a choice among the set options as an output. Where the models differ is in the specific set of assumptions mediating this input-to-output process. Yet, the models still share a common space if one views each assumption used by the models as a decision along a model-making path. We refer to these decision points as “nodes”, and the type of decisions one can make at the nodes as “switches”.

To use an analogy, the nodes in the path might correspond to a type of relay in an electrical circuit. At each node in the relay, a switch can control the path that the circuit takes, much like the ensemble of assumptions that produce a particular model. Following this logic, we developed a “switchboard” where all possible combinations of processing assumptions could be realized through different configurations of switches at each node.

Given some global assumptions about how to simulate the models, we can define the set of nodes as well as the set of switches present at each node. This analysis should allow us to better isolate the unique contributions of particular mechanisms when fitting preference models to data (see Van den Berg, Awh, & Ma, 2014; Donkin, Brown, &

Heathcote, 2011; Heathcote, Loft, & Remington, 2015; Rae, Heathcote, Donkin, Averell, & Brown, 2014, for a similar analyses). We first discuss the global assumptions used by all of the models. Next, we discuss the set of nodes, as well as the values the switches can take at each node. We then discuss the model fitting results.

Global Specification of the Models

Following traditional sequential sampling theory, we assume the attribute values serve as input into an accumulator model. Each alternative is represented as its own accumulator, and the preference states evolve according to a recursive equation. Letting $P[t]$ denote the vector of preference states for the m alternatives at time t, preferences evolved according to the following equation:

$$\begin{aligned} P[t] &= SP[t-1] + VA_t + I_0 + \epsilon_t, \text{ and} \\ P[t] &\leftarrow \max\left(0, P[t]\right), \end{aligned} \tag{1}$$

where S denotes a "feedback" matrix, V is the input matrix, A_t specifies the dimension to attend at time t, I_0 is a baseline input term common to all accumulators, and

$$\epsilon_t \sim \mathcal{N}(0, \sigma).$$

The second statement about the accumulation dynamics in Equation 1 specifies that a floor on activation is used, such that if an accumulator's value goes below zero, it is reset to zero. The lower bound constraint is commonly used in the MLCA (and LCA) model, and we retain this assumption for the switchboard models so that we can appreciate the roles of lateral inhibition and leakage (cf. van Ravenzwaaij, van der Maas, & Wagenmakers, 2012; Bogacz et al., 2006). For all of the models investigated here, the baseline input term was always freely estimated. To evolve the preferences, we ran the accumulators for $T = 300$ iterations, and assumed that no threshold existed to avoid any early terminations.

List of Model Mechanisms

In developing the switchboard, our goal was to create a set of nodes that, depending on the values of the switches at each node, could subsume all of the extant models. Ultimately, our derivations produced a set of seven nodes along the model-making path, with varying levels that the switches could take at each node. In this section, we describe the seven nodes, and the set of different values that each switch at these nodes can take.

1. Processing. At the first node, one must decide whether or not the stimuli are represented veridically, or whether the values of the attributes representing the stimuli are internally transformed in some way. In the case where three alternatives $\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z}$ are presented, each possessing values for two attributes P and Q, we can store the stimulus set into the matrix M such that

$$M = \begin{bmatrix} x_P & x_Q \\ y_P & y_Q \\ z_P & z_Q \end{bmatrix}, \tag{2}$$

where the columns correspond to the levels of the attributes and the rows correspond to the items. The veridical representation in M is used directly in the MLCA and MDFT models, but the AAM and MLBA models assume that some transformation is made to M prior to the deliberation process. To subsume the models, we assumed the transformation function $f(\cdot)$ such that $N = f(M)$, where N contains the subjective representation of the attributes, arranged in the same manner as M. At this node, the transformation function $f(\cdot)$ serves as the switch, and can take on one of three values. First, $f(\cdot)$ can take on the unity function such that $N = M$. Second, $f(\cdot)$ can implement the transformation function assumed by the MLBA model, in which case one new parameter is introduced. Third, $f(\cdot)$ could be a simple power function, as assumed by the AAM. In this case, a different new parameter would be introduced to make the transformation flexible. Because the addition of new parameters makes the model more flexible, we took the addition of new parameters into account when assessing model performance.

2. Attribute Comparison. Following the formation of the subjective representation matrix N, we have three options when deciding how to compare the attribute values in N with one another. If we let the matrix D contain the set of attribute comparisons, we can define the function $g(\cdot)$ such that $D = g(N)$. As in the processing stage, the function $g(\cdot)$ can take on three values. First, the function $g(\cdot)$ can be a simple unity function where $D = N$, as in the AAM. Here, D will have the same dimension as N. Second, $g(\cdot)$ can simply evaluate every pairwise distance between the attributes. In this case, we can assume that $g(\cdot)$ expands the matrix N to create a partition of attribute P differences and attribute Q differences when creating D. For example, when three options are presented,

$$D = g(N) = \left[\begin{array}{ccc|ccc} x_P - x_P & x_P - y_P & x_P - z_P & x_Q - x_Q & x_Q - y_Q & x_Q - z_Q \\ y_P - x_P & y_P - y_P & y_P - z_P & y_Q - x_Q & y_Q - y_Q & y_Q - z_Q \\ z_P - x_P & z_P - y_P & z_P - z_P & z_Q - x_Q & z_Q - y_Q & z_Q - z_Q \end{array}\right]. \tag{3}$$

This particular difference matrix is used in the MLBA and MLCA models, where a linear combination of the elements of D are fed into the accumulators as part of the input. For example, in the MLCA model (assuming no loss aversion), when attention is focused on attribute P, the input to the first accumulator is the first row of D in the left partition:

$$x_P - x_P + x_P - y_P + x_P - z_P = 2x_P - (y_P + z_P). \tag{4}$$

The third possibility for $g(\cdot)$ is an average distance between the attribute values, as the one used in the MDFT model. This form of $g(\cdot)$ creates a matrix D with the same size as N, but D contains different information. In the three alternative case,

$$D = \left[\begin{array}{c|c} x_P - \dfrac{(y_P + z_P)}{2} & x_Q - \dfrac{(y_Q + z_Q)}{2} \\ y_P - \dfrac{(x_P + z_P)}{2} & y_Q - \dfrac{(x_Q + z_Q)}{2} \\ z_P - \dfrac{(x_P + y_P)1}{2} & z_Q - \dfrac{(x_Q + y_Q)}{2} \end{array}\right]. \tag{5}$$

This particular difference matrix is used as input in the MDFT model in a similar way to that of the MLBA and MLCA models. For example, when attribute P is attended to, the first accumulator receives the input $x_P - .5(y_P + z_P)$. By comparing this input to the input term in Equation 4, we see that only a scalar separates the two input terms. Because our switchboard analysis always allows the momentary integration noise to be a free parameter in the model, the relative contribution of the inputs relative to the noise can always scale accordingly. As a consequence of this tradeoff, only the first two $g(\cdot)$ functions are needed to subsume the set of mechanisms used by the extant models, and neither function requires any additional parameters. Hence, the switch can only take on two values at this node.

3. Filtration. The final attribute preprocessing stage the models may assume is filtration. Essentially, filtration works by applying some subjective valuation to the particular values within the matrix D. For example, in the MLCA model, Usher and McClelland (2004) have argued for the use of loss aversion as a way to asymmetrically weigh losses and gains within D. In the MLBA model, Trueblood et al. (2014) have argued for a somewhat similar asymmetric function for losses and gains, where different weight parameters are used depending on the values within D.

Letting V represent the final input matrix following perceptual processing, we can use the function $l(\cdot)$ to apply a transformation to D such that $V = l(D)$. We allow the function $l(\cdot)$ to take on three forms. First, $l(\cdot)$ could simply be the unity function such that $V = D$, which is used by both the MDFT model and the AAM. Second, $l(\cdot)$ could be the loss aversion function, where we use the same function applied in our previous analyses (i.e., the same form used by Usher & McClelland, 2004). This version of the loss aversion function is convenient as it does not require any additional parameters. Third, $l(\cdot)$ could be the weighted combination function used in Trueblood et al. (2014). Here, a weight matrix W is used to apply decreasing weight to values of attributes that are farther apart, according to a parameterized exponential function (Shepard, 1987). As W is computed for every element in D, W is the same size as D. Hence, once the weight matrix is computed, a Hadamard product is taken between W and D to create V: $V = D \circ W$. When the third $l(\cdot)$ function is used, three additional parameters are needed to allow for asymmetric processing across the attribute space (Trueblood et al., 2014).

4. Attention. The next node in the path considers how attention should be divided across the attribute space. In most context effects experiments, each stimulus consists of two attributes, and the models must make some assumption about the relative importance of each attribute dimension. At this node, we assume that attention can be divided in two different ways. First, attention could be divided according to an attribute weight parameter $\omega \in [0, 1]$ that is freely estimated when the model is fit to data. This particular method of attention allocation is what is assumed by both the MLCA and MDFT models. Second, attention could be divided according to the associations made between the presented stimuli and their corresponding attributes. The AAM is the only model to specify how attention should be divided, where it assumes that ω is calculated based on the sum of the attribute values for each of the presented stimuli in the set. In addition, the AAM assumes that biases in the allocation of attention manifest through a set of parameters which are added to the summation of attribute values (see supplementary materials). Hence, when the second mechanism for calculating ω is used, two additional parameters must be freely estimated.

A note about how the attentional node in the path relates to the way attention is allocated in the MLBA model is in order. The MLBA model allows for attribute bias through a parameter that affects the attentional weights used in the model. However, as this calculation appears in the same equation as the asymmetric weighting of losses and gains, we consider it to be a filtration mechanism rather than an assumption about how attention should be divided. In the end, as the ω weight can be freely estimated at this node in the path, the specific details of how the MLBA model combines information from each attribute dimension can be subsumed with changes in the filtration and attention nodes in the path.

5. Attribute Integration. The next node in the path determines how the attribute dimensions are integrated over time. While the attention node details how attention should be allocated across the two dimensions through the parameter ω, we have not yet described how moment-to-moment fluctuations in attention might play a role in the evolution of preferences over time. At this node in the path, we assume two methods for attribute integration. The first method of attribute integration involves a stochastic process where attention is randomly allocated to one dimension or another at every moment in time. This process is known as the Bernoulli process and is assumed by three of the extant models (i.e., AAM, MLCA, MDFT).[6] Consider the variable a_t that specifies which dimension attention should be allocated to at time t in the deliberation process. Assume that if $a_t = 0$, attention is allocated to attribute Q, and if $a_t = 1$, attention is allocated to attribute P. Models like AAM, MLCA, and MDFT assume a Bernoulli process on a_t such that

$$a_t \sim \text{Bernoulli}(\omega).$$

It is both convenient and conventional to represent the allocation of attention in matrix form. To do this, we define a matrix A_t such that

$$A_t = \begin{bmatrix} a_t & 1 - a_t \end{bmatrix}^\intercal .$$

Equation 1 specifies a matrix multiplication of the input matrix V and the attention matrix A_t. As the dimensionality of V changes depending on the function $g(\cdot)$ (i.e., depends on how pairwise differences are calculated), we must also adjust the size of A_t depending on the size of V. Specifically, when the second $g(\cdot)$ function is used (i.e., all pairwise differences are obtained), we set

$$A_t = \begin{bmatrix} a_t & a_t & a_t & 1 - a_t & 1 - a_t & 1 - a_t \end{bmatrix}^\intercal .$$

The second method of attribute integration is a simple weighted average, which is assumed by the MLBA model. Here, the integration of attribute information is not stochastic, but deterministic, meaning that the same amount of attribute P and attribute Q information is used at every moment in time. In this case, we can express the weighted average by adjusting the matrix A_t, such that

$$A_t = \begin{bmatrix} \omega & 1 - \omega \end{bmatrix}^{\mathsf{T}},$$

with an analogous adjustment for the dimension differences in V discussed above. Note that while we have defined A for every moment in time t, this is purely for convenience: A does not change over time in this second method.

6. Competition. The next node in the path corresponds to the role of competition within the deliberation process. During the accumulation process in the switchboard model, one option may have some influence on the other options in the stimulus set, and the type of influence can be determined on the basis of the attributes themselves, or they can be simply freely estimated. In total, we assume that three types of competition may exist during the deliberation process. First, the options may not interact with one another at all, and so the accumulators would evolve over time completely independently, as they do in the MLBA model. Mathematically, this switch can be implemented through the specification of the feedback matrix S in Equation 1. Specifically,

$$S = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}.$$

Second, the options may interact with one another through leakage and lateral inhibition. The parameters that control these mechanisms can be freely estimated, as they are in the MLCA model and the version of AAM we investigated above. In this method of competition, two free parameters are added to the model to implement lateral inhibition and leakage. Importantly, the parameters for lateral inhibition and leakage are constrained to be equal across different stimulus sets. Mathematically, the feedback matrix can be adjusted to incorporate these parameters by setting

$$S = \begin{bmatrix} k & -L & -L \\ -L & k & -L \\ -L & -L & k \end{bmatrix}.$$

Third, the mechanisms of lateral inhibition and leakage may be determined by the relative distances between the stimuli in the attribute space, as they are in the MDFT model. While the lateral inhibition and leakage parameters are not estimated directly, there are three parameters that are freely estimated to determine their values. One major difference between this third version of competition is that the lateral inhibition and leakage terms vary for different stimulus sets, if the distances between the stimuli are different. The adjustment of S to incorporate this third switch requires a few calculations and extra definitions. We have relegated these equations to the Supplementary Materials, as they are exclusive to the MDFT model.

7. Valuation Noise. The final consideration in the switchboard model is the role that variability plays in establishing preferences. Mirroring typical decision making models, the modeling choice we have to make is whether or not to include within-trial or between-trial variability. Most of the extant models assume a stochastic process in the deliberation period where moment-to-moment integration is subject to noise. The MLBA model is an exception here, as it assume that the integration of stimulus information is ballistic. Instead, the MLBA model captures variability in decision making data by assuming that the rate of accumulation varies from one trial to the next. While some classic models of perception (e.g., the DDM; Ratcliff, 1978) assume the presence of both within- and between-trial variability, because we are only fitting choice probabilities, we ignored the possibility of having both sources of variability as this configuration did not appear to be identifiable.

Hence, at this node in the path we have two possibilities for the type of variability present in the model predictions. For the first source of variability, we assume the presence of within-trial noise where momentary fluctuations in preference are observed through time. When the switch is set to this value, the parameter σ is freely estimated. For the second source of variability, we assume the presence of between-trial noise where the strength of evidence for the accumulators varies from one trial to the next. To implement this, we assumed a free parameter s such that

$$V \sim \mathcal{N}(0, s),$$

for all of the elements of the input matrix V. Hence, both switches involve one additional parameter, meaning that every model that the switchboard can create has at least two free parameters: a variability term (i.e., s or σ), and a baseline input term I_0.

Fitting the Models

With the seven nodes in the model-making path and the various number of switches at each node, we arrived at $3 \times 2 \times 3 \times 2 \times 2 \times 3 \times 2 = 432$ total models to investigate. We chose to fit each of the models to the full set of perceptual data from our Study 2, as these data were more constrained that those of Study 1. The average preferences were taken for each response option, and individual differences were not taken into account (i.e., no hierarchical models were investigated).

We again used the probability density approximation (PDA; Turner & Sederberg, 2014) method to fit each model to data. For each parameter proposal, we simulated the model 1,000 times to obtain a stable approximation of the option preferences. We again used a purification step for all chains every 10 iterations (Holmes, 2015). This time, to reduce our computational burden, we focused on obtaining the MAP estimate rather than full posterior estimates. Here, we used the "burnin" mode of the ABCDE algorithm (Turner & Sederberg, 2012). We used 24 chains and ran the algorithm for 400 iterations. For the first 100 iterations, a migration step was used (see Turner & Sederberg, 2012, 2014, for details) with probability 0.1, but after this period, the migration probability was set to zero.

Results

The switchboard analysis allows us to not only investigate which models perform particularly well for our data, but also which mechanisms perform well by examining the average performance of a class of models. The results section is organized into two parts. First, we discuss which models performed best out of the 432 models we explored. Second, we investigate the relative merits of the levels of each switch by aggregating the model fitting results over the nodes in the model-making path.

Evaluating Model Performance. The first comparison involves evaluating all 432 models and their particular constellation of mechanisms. To do this, we first fit the models to the data from our perceptual experiment, and calculated the Bayesian information criterion (BIC; Schwarz, 1978). We chose the BIC for the switchboard analysis because our focus was on obtaining the best-fitting parameter estimates, rather than the full posterior distribution used in the previous studies. In addition to assessing model fit, the BIC also contains a model complexity term based on the number of parameters and number of data points, and in many cases functions similarly to the DIC.

Figure 7 shows the BIC value for every model investigated, color coded according to the legend on the far right-hand side. As with the DIC values, lower BIC values indicate better performance. Figure 7 is organized according to the nodes discussed in the previous section, where each layer represents a particular node. Along the rows, the outermost factor is competition, followed by the type of valuation noise, followed by the type of processing (i.e., subjective mapping of attribute values). Along the columns, the outermost factor is attribute integration, followed by the type of attention, followed by distances computed between the attributes, followed by filtration. Figure 7 is organized to reveal patterns across the various factors that can be manipulated to arrive at a better performing model. For example, by comparing columns 3, 9, 15, and 21 to columns 6,12,18, and 24, we can infer that using the weighted average filtration method assumed by the MLBA model only helps to improve the fit if it is used in conjunction with pairwise differences. A similar result appears for the loss aversion switch.

In general, many of the models perform quite well. Figure 7 shows that one switch – the weighted average filtration method (as used in the MLBA model) – creates particularly bad models, and this poor performance distorts the scale of the figure such that it is difficult to appreciate the relative performance of the better-performing models. Another approach is to look at the top performing models. The first ten rows of Table 8 show the ten best-performing models in our set. The leftmost column corresponds to a model label, the rightmost column shows the BIC value, and the columns in between show the values of the switch at each node in the model-making path. Looking down the columns, we see that the best performing models (1) all possess stochastic integration of attributes, (2) have either freely estimated lateral inhibition and leakage or accumulate independently, (3) mostly possess directed attention via the associations of attributes (i.e., assumed by AAM), and (4) mostly use loss aversion. The results on attribute processing, distance calculations, and valuation noise are mixed.

We can also compare the best-performing models to the switchboard versions of the extant models. We stress that the switchboard versions do not map on identically to the set of four extant models. There are several processing assumptions, such as the number of time steps in the simulation, step size in the Euler approximation, floor on activation, and baseline input terms that may affect the BIC statistics either directly (e.g., by adding additional parameters) or indirectly (e.g., the model simulation may not be representative of the long term dynamics). Instead, these models are the closest switchboard version to the extant models as they incorporate the same set of mechanisms in generating predictions for behavioral data. Within these models, the MLCA variant performs best, the MDFT model variant performs second best, the AAM variant performs third best, and the MLBA model variant performs fourth best. Relative to the best-performing models, the MLCA model is comparable (see the legend of Figure 7 to gain a sense of the variability across models), and the MDFT and AAM models also performed well.

In light of the possibility that the switchboard versions were not equivalent to the extant models, we refit the four extant models with attribute biases (i.e., MDFT 2.0, MLCA 2.0, AAM 4.0, and MLBA 2.0) and obtained BIC values rather than DIC values obtained in the nonhierarchical fits from Study 2. We obtained a BIC of 2449.17 for the MDFT model, 987.71 for the MLCA model, 1140.68 for the AAM, and 3999.05 for the MLBA model. While the exact numbers are different from the switchboard versions, the ordering is roughly the same except with AAM outperforming MDFT in this analysis. Regardless, even the best performing model MLCA 2.0 did not achieve a small enough BIC statistic to rank among the top 10 performers suggesting that the models presented in Table 1 might present new versions worth consideration in future studies.

Evaluating Model Mechanisms. The switchboard analysis also allows us to investigate whether certain switches in the model-making path lead to particularly good or bad model performance. To do this, we can aggregate the BIC statistics across values of specific switches. Figure 8 shows the median BIC statistic (across models) for every switch by node combination. The dashed vertical line corresponds to the overall median BIC statistic as a guideline. The nodes are arranged in the order that they were discussed from the section above. By comparing the variability in the median BIC statistics across switches, Figure 8 shows that the three factors that contribute most strongly to the BIC statistics are filtration (green), attribute comparison (blue), competition (red), and valuation noise (pink). For the filtration node, Figure 8 shows that no filtration and loss aversion contribute positively to model performance (i.e., they lower the BIC statistic relative to the global median), whereas the weighted average indifference curve assumed by the MLBA model contributes negatively to performance. For the attribute comparison node, Figure 8 shows that having a pairwise comparison among the options improves the model performance. For the competition node, Figure 8 shows that having no competition contributes positively to model performance, having freely estimated competition does not markedly improve the model performance, and having competition determined by the similarity among the stimuli as in the MDFT model contributes negatively. In the valuation noise node (pink), having between-trial variability performs better than having within-trial variability.

Table 1

*Absolute results of the switchboard analysis. The leftmost column designates the model number, the rightmost column designates the Bayesian information criterion (BIC), and other columns designate the value of the "switches" at each node in the model-making path. The first ten rows show the ten best performing models, whereas the last four rows show the switchboard versions of the extant models. *Note that the processing assumptions for these switchboard versions are not identical to the true extant models.*

Model	Processing	Distance	Filtration	Attention	Integration	Competition	Noise	BIC
318	curve	pairwise	loss aversion	associations	stochastic	free	between	555.65
299	power	pairwise	loss aversion	free	stochastic	free	between	740.58
97	none	none	loss aversion	associations	stochastic	free	within	763.56
313	none	none	loss aversion	associations	stochastic	free	between	767.17
240	curve	pairwise	none	associations	stochastic	none	between	771.29
243	curve	none	loss aversion	associations	stochastic	none	between	774.65
21	curve	none	none	associations	stochastic	none	within	775.28
241	none	none	loss aversion	associations	stochastic	none	between	782.41
27	curve	none	loss aversion	associations	stochastic	none	within	784.60
26	power	none	loss aversion	associations	stochastic	none	within	785.38
MDFT*	none	pairwise	none	free	stochastic	proximity	within	1329.37
MLCA*	none	pairwise	loss aversion	free	stochastic	free	within	921.65
AAM*	power	none	none	associations	stochastic	free	within	1617.81
MLBA*	curve	pairwise	indifference	free	deterministic	none	between	6553.57

Other factors are less impactful on the BIC statistic. In the processing node (orange), having no preprocessing is the best value of the switch. In the attention node (yellow), freely estimating ω performs better than the associations mechanism assumed by the AAM in the aggregated analyses. Although we caution against arriving at this conclusion because the associations rule was used by all of the top 10 models. Finally, in the attribute integration node (gray), stochastic and deterministic integration perform similarly. We believe this to be completely reasonable as we are not assuming the presence of an accumulation threshold and instead running the models for $T = 300$ iterations. In this setup, oscillations in attention will closely mimic a simple weighted average.

General Discussion

In this article, we evaluated the four extant models of preferential choice that have been shown to account for the three most robust context effects. In Study 1, we evaluated a few variants of each model on the basis of their relative fits to the data from Berkowitsch et al. (2014). We determined that across all models, some element of attribute bias was essential to best capture the patterns in the data when collapsing across subjects. However, when modeling subject-to-subject variability hierarchically, we found that adding the attribute bias term was unnecessary. For the AAM and MLCA models, we also determined that mechanisms like lateral inhibition and leakage provided better fits to data that went over and above the penalty incurred for increases in complexity by way of number of parameters. The MDFT model already assumes a fixed amount of competition, and we found that this model fit the data better than any of the other model variants when attribute bias was allowed. The MLBA model, which eliminates mechanisms like leakage and lateral inhibition, performed the worst of all the model variants from Study 1 when attribute bias was allowed. However, when moving to hierarchical models, the MDFT model performed worst, and the MLBA model performed substantially better. Our conclusion is that MDFT lacks the flexibility necessary to explain the amount of variability across subjects that is typically observed in decision making tasks. Other models such as AAM and MLCA performed consistently well regardless of whether the hierarchical structure was in place.

In Study 2, we subjected the four models to a new set of data consisting of two- and three-choice alternatives in the perceptual domain. We included a number of filler trials, which are not commonly modeled, further constraining the model fits (see Table 4). Once again we fit versions of each model that allowed for attribute dimension biases in both the hierarchical and non-hierarchical versions. In the non-hierarchical models, the results were mixed: some models performed better with attribute bias parameters (i.e., MLCA and AAM) whereas others did not (i.e., MDFT and MLBA). However, when fitting the data hierarchically, we found strong evidence across all models that some form of attribute bias is essential in capturing preference in the perceptual domain. Across both hierarchical and non-hierarchical analyses, the AAM and MLCA models performed consistently well, whereas models like MDFT and MLBA performed relatively worse. The MLCA model's ability to capture data from a perceptual task seems to be inconsistent with Trueblood et al.'s (2014) criticism of MLCA on the basis that loss aversion is not a realistic mechanism for perceptual stimuli. The fact that the MLCA model is able to account for both perceptual (i.e., Study 2) and consumer goods (i.e., Study 1) data suggests that the loss aversion function, in conjunction with the other mechanisms in the model such as leakage and lateral inhibition, is at least a plausible function relating the objective values used in the experimental setting to the subjective values used in the evaluation process.

In Study 3, we attempted to provide some consensus about the efficacy of the numerous mechanisms at play across the four models. To do this, we developed a "switchboard" analysis that conceives of each proposed model mechanism as a choice along a model-making path. Essentially, every possible choice was combinatorially (i.e., a total of 432 models) fit to the data from our perceptual choice experiment. Following this procedure, the BIC was calculated and each model was compared. Considering interaction effects, we found that the top ten performing models had consistent features such as the inclusion of stochastic integration, competition, loss aversion, and association-based attention (i.e., as assumed by AAM). We also examined the utility of model mechanisms by aggregating the performance metrics over values of the switches in the switchboard. This analysis revealed four dominant trends in the model making process that have large effects on the BIC statistics (see Figure 8): filtration, attribute comparison, competition, and valuation noise. Regarding filtration, it seemed that having either no filtration or loss aversion worked substantially better than the weighted average method (i.e., as assumed by the MLBA model). Regarding attribute comparison, pairwise comparisons performed better than no comparisons at all. Regarding competition, having either no competition or competitive mechanisms that were free to vary performed better than competitive mechanisms based on associations (i.e., as assumed by the MDFT model). Finally, having between-trial variability performed better than having within-trial variability.

We focus our General Discussion on three important topics. First, we draw some conclusions on the basis of our three studies about the relative utility of the model mechanisms assumed by the extant models of preference. We then discuss a few limitations of the simulation-based approach, as well as the hierarchical models we developed throughout the article. Finally, we close with a discussion about the importance of mathematical tractability in the development of cognitive models.

Summary of Conclusions About the Extant Models and Their Mechanisms

In this article, we have provided many detailed fits of the four extant models to two datasets. Table 7 provides a summary of our model-fitting results, where each of the best-fitting model variants has been ranked relative to the other models for each of the five analyses we performed in this article. In general, Table 7 shows that while the AAM and MLCA models performed consistently well across analyses and data types, models like MDFT and MLBA performed relatively worse. But what have we learned about the suitability of the assumptions made by these extant models? What assumptions are particularly problematic, and what features of the models could benefit from revision?

Starting with the MDFT model, perhaps the most striking performance difference was that of the similarity-based competition rules that MDFT uses to specify the lateral inhibition term. Across all of our model fits, we found that MDFT did not perform relatively well, and given the similarities between the MDFT and MLCA models, we suspect that the specification of the lateral inhibition term is to blame. From our switchboard analysis, we found strong evidence to suggest that this mechanism did not perform well relative to a simple, freely-estimated lateral inhibition term as used by the MLCA and AAM models. While the switchboard version of the MDFT model performed well overall (i.e., BIC= 1329.37), we found that by simply switching the competition node to be freely estimated instead, a better model variant could be produced (i.e., BIC= 1127.10). Other mechanisms such as the pairwise differences among attributes seem to be strong features of this model, providing apparently enough information to allow good fits to data without extra processing assumptions (e.g., AAM or MLBA models).

The MLCA model was a strong performer overall, as both the hierarchical and non-hierarchical models consistently fit data well across all studies presented here. Furthermore, the switchboard variant of the MLCA model performed strongest of the extant models, suggesting that the particular combination of these model mechanisms produced a consistent architecture for explaining preference. In the switchboard analysis, we found that only one dimension could be improved on for the MLCA model; when the valuation noise was allowed to vary between trials, the model performed better (i.e., BIC= 826.26) than its within-trial counterpart (i.e., BIC= 921.65).

The AAM model performed arguably the best across all of our data sets, although we emphasize that the best variant of AAM was slightly different from its original form (Bhatia, 2013). For example, we found that adding mechanisms like lateral inhibition and leakage greatly improved the model fits to data, even after applying penalties for complexity. We also found that when biases in the attribute space were essential, such as for the perceptual stimuli, adding two attribute bias parameters – one for each dimension

– greatly improved the model's fit to data. Perhaps the most compelling feature of AAM is in its specification of how to allocate attention. Essentially, AAM assumes that when the values of a set of items are high on one attribute dimension, more attention should be directed toward said dimension. While the association rule certainly didn't prohibit AAM from fitting the data investigated here, after penalizing for complexity, on average, this mechanism did not perform substantially better than simply adding a free parameter (see Figure 8). Regardless, we consider the attention mechanism in AAM to be a power feature of the model, as nine of the top ten performing models possessed this model feature (see Table 8).

For the MLBA model, the curvature assumption mapping objective values to subjective ones was comparable to other preprocessing assumptions. Across all studies, we did find some evidence that competition among the options via lateral inhibition and/or leakage is important for models of preference, an assumption that is at odds with that of the MLBA model. First, we observed that models like AAM performed substantially better when these competitive mechanisms were introduced. Second, we observed that models like AAM and MLCA performed, on the whole, better than models like MDFT and MLBA, for reasons that may be due to the freely estimated competition mechanisms. Third, the four best-performing models in our switchboard analysis all assumed competition (see Table 8). That is not to say that models that assume no competition do poorly – indeed the next six best performing models did not assume competition – but having competition across the board seemed to improve model fits. Exclusive to the MLBA model, changing the value of the switch at the competition node produced markedly better fits, regardless of the switch value: when competition was freely estimated, the BIC was 4192.39, and when competition was determined through associations as in the MDFT model, the BIC was 4735.93 (i.e., compare these values to a BIC of 6553.57 when competition was off). Finally, the method used to filter distances between stimuli was found to be problematic. In the switchboard analysis, this weighted averaging technique was the worst model feature of any mechanism we investigated (see Figures 7 and 8).

Limitations of Hierarchical Modeling

Hierarchical models are advantageous because they enforce constraints on the lower-level (e.g., subject-specific) parameters on the basis of the information learned at the hyper-level (e.g., group-specific) parameters. The information at the hyper level is an extrapolation of the variability from one subject to another, and so one can view the type of constraint that hierarchical models offer as being data driven. In psychology, hierarchical modeling has become a recent trend in describing human performance, at least partially due to the wide variability often observed in decision making.

Despite the many advantages of hierarchical models, they come with a cost. Hierarchical models balance the information at the subject level (i.e., the likelihood) with the information at the group level (i.e., the prior). In so doing, each posterior estimate at the subject level is a combination of two sources of information, and the relative weight of these two sources depends on the number and quality of the observations going into each source. In some cases, when the information at the subject level is sparse, as in the data presented in this article, the hierarchical model will resort to using the information contained in the group level. When this happens, the estimates for the subject-level parameters may systematically differ from estimates that would have been obtained had a hierarchical model not been used. This general feature of hierarchical models is know as shrinkage.

Hierarchical models are not the only way to ascertain individual differences. Recently Liew et al. (2016) present results from a cluster-based analysis of context effects. The goal of their analysis was to identify commonalities among subjects in their decision making behavior. To do this, Liew et al. (2016) used a nonparametric Bayesian classifier model (Navarro, Griffiths, Steyvers, & Lee, 2006) to separate subjects into groups based on their pattern of choice probabilities across the three classic context effects. Their results show that almost no one displayed all three context effects simultaneously. Instead, clusters emerged such that some subjects show either the similarity or compromise effect, but never both. Importantly, they found that attribute dimension biases played a much larger role in the patterns of decision making that did any of the context effects.

Given the clusters found in Liew et al. (2016), one may wonder whether or not the hierarchical structures proposed in this article somehow damage the cluster structure that may be present in context effect data. There are at least two responses to this concern. First, the analyses presented in Liew et al. (2016) were not model based. Instead, they relied on the pattern of response probabilities to form their clusters of decision patterns. By contrast, the model-based approach advocated in this article essentially uses a computational model to guide the formation of structure across the various types of decision making patterns. If the computational model is an accurate description of context effects, then there should be parameter sets that can only produce compromise effects, and other sets that can only produce similarity effects when the data show these patterns. The degree to which these two parameter sets are different will affect the group-level parameters in the hierarchical model. However, because the hierarchical structure is flexible, it will still allow the model to capture subject-to-subject differences in decision making (e.g., see Figure 3). Assuming these clusters of parameter sets exist in the model, then the ability of the hierarchical model to capture these clusters will depend on the size of the underlying clusters (i.e., how many subjects comprise the cluster). If a cluster is small in size, the cluster will be pulled toward the mean of all the clusters due to shrinkage.

Second, the Dirichlet process used in Liew et al. (2016) is not hierarchical. In the hierarchical analyses presented here, we could have assumed a nonparametric prior over the subject-level parameters, and allowed clusters of model parameters to emerge in a similar way as that of Liew et al. (2016). If one were concerned about the role of shrinkage in biasing the estimates of the subject-level parameters, a nonparametric Bayesian prior over these parameters would be the ideal approach. Unfortunately, these priors are computationally much more difficult to implement compared to the generic priors we used here. The computational difficulty, combined with the likelihood-free approach used here, seemed outside the bounds of feasibility for the myriad number of model fits we performed. We save these nonparametric extensions for future research.

The Role of Mathematical Tractability in Model Development

Considering model structure, the mostly recently proposed MLBA model deviates from the other three models in its assumption about independent, ballistic accumulation of evidence. While it is not clear whether this assumption was made for the purposes of mathematical tractability per se, the advantages gained by this particular departure from what was a nearly conventional modeling choice motivates a stimulating discussion on the role that mathematical tractability should play in model development. In developing mathematical models, our goal is to put forth a model that can not only fit data well, but also makes a strong yet accurate commitment to the distribution of data we should see in our experiments (Roberts & Pashler, 2000; Teodorescu & Usher, 2013). Echoed from the introduction, the assessment of a model's full credentials involves two important considerations: model fit and model complexity (I. J. Myung & Pitt, 1997; I. J. Myung, 2000; I. J. Myung et al., 2000).

In the domain of model development, the word complexity can sometimes refer to either the flexibility of a model, and can sometimes refer to the ease of implementation (Turner, Forstmann, Love, Palmeri, & van Maanen, 2017). However, these are two different concepts. The ease of implementation is related to the mathematical tractability, but it is not related to complexity (J. I. Myung, Montenegro, & Pitt, 2007; Montenegro, Myung, & Pitt, 2011). The MLBA model stands alone in that it is extremely easy to implement and fit to data because there are analytic expressions relating the model parameters to the data. These expressions make the model very easy to fit to data via maximum likelihood or Bayesian approaches. As stressed in Trueblood et al. (2014), for internally-driven experimental paradigms, where the subject is allowed to deliberate as long as they desire, computer intensive simulations are required to fit the AAM, MDFT, and MLCA models. Unfortunately, tractability does not necessarily map onto fewer parameters, or the degree of model flexibility. As such, tractability is also unrelated to complexity when used as a measure of model performance.

To illustrate, consider as analogy the bind cue decide model of episodic memory (BCDMEM; Dennis & Humphreys, 2001) model. The BCDMEM model was proposed as a pure context model of episodic memory, an assumption that was at odds with the dominant models at the time. The model was proposed as a simulation model, meaning that the likelihood function relating model parameters to predictions about the hit and false alarm rates was intractable. For years, anytime a researcher wanted to fit BCDMEM to their data, they were forced to rely on simulation methods, such as approximate least squares. Eventually, J. I. Myung et al. (2007) produced analytic expressions for the model. While these expressions are computationally difficult to evaluate, they can be used to assess the model's flexibility, complexity, and identifiability (Turner, Dennis, & Van Zandt, 2013). Ultimately, the expressions derived by J. I. Myung et al. (2007) unlocked one key facilitator in the endeavor of rigorous model evaluation: mathematical tractability.

What can we make of the research conducted in the time between the development of the original model in 2001, and the derivation of analytic expressions in 2007? As the assumptions of BCDMEM were never changed during this time period, the complexity of BCDMEM also never changed. Hence, the ability of BCDMEM to fit data also never changed. In a similar vein, if a researcher published a paper deriving analytic expressions for say, the MDFT model tomorrow, nothing about the previous *fits* of the MDFT model over the past decade will have changed. Nothing about the model's *complexity* will have changed either. Instead, the MDFT model would simply be given a compelling *pragmatic* advantage in choosing among the various models for application purposes because the model would now be (potentially) easier to fit to data.

While mathematical tractability is highly advantageous, the analyses in this manuscript highlight the importance of new methods for performing inference on simulation-based models (see Palestro et al., 2016, for a review). In theory, any computational model can now be fit to data using the likelihood-free approach, allowing researchers to regain access to tried-and-true methods for model evaluation. Our view is that by using these methods, researchers are free to experiment with as many complex, stochastic model variants as they can imagine, while still assessing model flexibility relative to the data. Of course tractable models offer compelling advantages, but if compromising assumptions are required to produce tractability, these assumptions may now be rejected on the basis of a theoretical position, prior research, curiosity, or even something as simple as preference.

Conclusions

The model evaluations and comparisons presented in this article speak to the possibilities that likelihood-free algorithms provide. Our approach allowed us to evaluate the models on the basis of model fit, while still controlling for model complexity. Our results show that some of the stochastic models of context effects provided excellent accounts of empirical data from two studies in both the consumer and perceptual domains. Our switchboard analyses show that some decisions in the model-making process are more consequential than others. Namely, factors such as the way attributes are compared, the way objective distance in the stimulus space affects our representation of the stimuli, and the manner in which the deliberation among alternatives is carried out play an essential role in determining whether or not a model can account for context effects. By comparing and contrasting the various model mechanisms, our model-based approach underscores the notion that mathematical tractability, while certainly a convenient feature of any model, should not be the primary impetus for model development. Instead, the degree to which a model fits data – relative to its complexity – should be the ultimate test in the evaluation of model mechanisms.

References

Amir, O., & J., L. (2008). Choice construction versus preference construction: The instability of preferences learned in context. *Journal of Marketing Research*, *45*, 145-158.

Anderson, J. A. (1990). *Cognitive psychology and its implications.* San Francisco, CA: Freeman.

Ando, T. (2007). Bayesian predictive information criterion for the evaluation of hierarchical bayesian and empirical bayes models. *Biometrika*, *94*, 443-458.

Berkowitsch, N. A. J., Scheibehenne, B., & Rieskamp, J. (2014). Rigorously testing multialternative decision field theory against random utility models. *Journal of Experimental Psychology*, *143*, 1331-1348.

Bettman, J. R., Luce, M. F., & Payne, J. W. (1998). Constructive consumer choice processes. *Journal of Consumer Research*, *25*, 187-217.

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

Bogacz, R., Brown, E., Moehlis, J., Holmes, P., & Cohen, J. D. (2006). The physics of optimal decision making: A formal analysis of models of performance in two-alternative forced choice tasks. *Philosophical Transactions of the Royal Society, Series B: Biological Sciences*, *362*, 1655-1670.

Brown, S., & Heathcote, A. (2005). A ballistic model of choice response time. *Psychological Review*, *112*, 117-128.

Brown, S., & Heathcote, A. (2008). The simplest complete model of choice reaction time: Linear ballistic accumulation. *Cognitive Psychology*, *57*, 153-178.

Busemeyer, J., & Townsend, J. (1993). Decision Field Theory: A dynamic-cognitive approach to decision making in an uncertain environment. *Psychological Review*, *100*, 432-459.

Celeux, G., Forbes, F., Robert, C. P., & Titterington, D. M. (2006). Deviance information criteria for missing data models. *Bayesian Analysis*, *1*, 651-673.

Chandon, P., & Ordabayeva, N. (2009). Supersize in one dimension, downsize in three dimensions: Effects of spatial dimensionality on size perceptions and preferences. *Journal of Marketing Research*, *46*, 739-753.

Choplin, J. M., & Hummel, J. E. (2005). Comparison-induced decoy effects. *Memory & Cognition*, *33*, 332-343.

Dai, J., & Busemeyer, J. R. (2014). A probabilistic, dynamic, and attribute-wise model of intertemporal choice. *Journal of Experimental Psychology: General*, *143*, 1489-1514.

Davis-Stober, C. P., Park, S., Brown, N., & Regenwetter, M. (2016). Reported violations of rationality may be aggregation artifacts. *Proceedings of the National Academy of Sciences of the United States*, *113*, E4761-E4763.

DeKay, M. L., Miller, S. A., Schley, D. R., & Erford, B. M. (2014). Proleader and antitrailer information distortion and their effects on choice and postchoice memory. *Organizational Behavior and Human Decision Processes*, *125*, 134-150.

Dennis, S., & Humphreys, M. S. (2001). A context noise model of episodic word recognition. *Psychological Review*, *108*, 452-478.

Donkin, C., Brown, S., & Heathcote, A. (2011). Drawing conclusions from choice response time models: a tutorial. *Journal of Mathematical Psychology*, *55*, 140-151.

Farmer, G. D., Warren, P. A., El-Deredy, W., & Howes, A. (2016). The effect of expected value on attraction effect preference reversals. *Journal of Behavioral Decision Making*.

Frydman, C., & Nave, G. (2016). Extrapolative beliefs in perceptual and economic decisions: Evidence of a common mechanism. *Management Science, Forthcoming*.

Heathcote, A., Brown, S. D., & Mewhort, D. J. K. (2000). The power law repealed: The case for an exponential law of practice. *Psychonomic Bulletin and Review*, *7*, 185-207.

Heathcote, A., Loft, S., & Remington, R. W. (2015). Slow down and remember to remember! A delay theory of prospective memory costs. *Psychological Review*, *122*, 376-410.

Hogarth, R. M., & Einhorn, H. J. (1992). Order effects in belief updating: The belief-adjustment model. *Cognitive Psychology*, *24*, 1-55.

Holmes, W. R. (2015). A practical guide to the probability density approximation (pda) with improved implementation and error characterization. *Journal of Mathematical Psychology*, *68*, 13-24.

Hotaling, J. M., Busemeyer, J. R., & Li, J. (2010). Theoretical developments in decision field theory: Comment on tsetsos, usher, and chater (2010). *Psychological Review*, *117*, 1294-1298.

Huber, J., Payne, J. W., & Puto, C. (1982). Adding asymmetrically dominated alternatives: Violations of regularity and the similarity hypothesis. *Journal of Consumer Research*, *9*, 90-98.

Johnson, E. J., & Russo, J. E. (1984). Product familiarity and learning new information. *Journal of Consumer Research*, *11*, 542-550.

Jones, M., & Love, B. C. (2011). Bayesian fundamentalism or enlightenment? On the explanatory status and theoretical contributions of Bayesian models of cognition. *Behavioral and Brain Sciences*, *34*, 169-231.

Kahneman, D., & Tversky, A. (1979). Prospect theory: An analysis of decision under risk. *Econometrica*, *47*, 263-291.

Krajbich, I., & Rangel, A. (2011). Multialternative drift-diffusion model predicts the relationship between visual fixations and choice in value-based decisions. *Proceedings of the National Academy of Sciences of the United States*, *108*, 13852-13857.

Krantz, D. H. (1989). Conjoint measurement: The luce-tukey axiomatization and some extensions. *Journal of Mathematical Psychology*, *16*, 158-174.

Lee, M. D., & Wagenmakers, E.-J. (2013). *Bayesian modeling for cognitive science: A practical course*. Cambridge University Press.

Liew, S. X., Howe, P. D., & Little, D. R. (2016). The appropriacy of averaging in the study of context effects. *Psychonomic Bulletin and Review*, *23*, 1639-1646.

Luce, R. D. (1959). *Individual choice behavior: A theoretical analysis.* New York: Wiley Press.

Medin, D. L., & Schaffer, M. M. (1978). Context theory of classification learning. *Psychological Review*, *85*, 207-238.

Montenegro, M., Myung, J. I., & Pitt, M. A. (2011). *REM integral expressions.* (Unpublished manuscript.)

Myung, I. J. (2000). The importance of complexity in model selection. *Journal of Mathematical Psychology*, *44*, 190-204.

Myung, I. J., Forster, M., & Browne, M. W. (2000). Special issue on model selection. *Journal of Mathematical Psychology*, *44*, 1-2.

Myung, I. J., & Pitt, M. A. (1997). Applying Occam's razor in modeling cognition: A Bayesian approach. *Psychonomic Bulletin and Review*, *4*, 79-95.

Myung, J. I., Montenegro, M., & Pitt, M. A. (2007). Analytic expressions for the BCDMEM model of recognition memory. *Journal of Mathematical Psychology*, *51*, 198-204.

Navarro, D. J., Griffiths, T. L., Steyvers, M., & Lee, M. D. (2006). Modeling individual differences using Dirichlet processes. *Journal of Mathematical Psychology*, *50*, 101-122.

Noguchi, T., & Stewart, N. (2014). In the attraction, compromise, and similarity effects, alternatives are repeatedly compared in pairs on single dimensions. *Cognition*, *132*(1), 44–56.

Palestro, J. J., Sederberg, P. B., Osth, A., Van Zandt, T., & Turner, B. M. (2016). *Likelihood-free techniques for cognitive science.* (Manuscript submitted for publication)

Parducci, A. (1965). Category judgment: a range-frequency model. *Psychological Review*, *72*, 407-418.

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

Pitt, M. A., Myung, I. J., & Zhang, S. (2002). Toward a method of selecting among computational models of cognition. *Psychological Review*, *109*, 472-491.

Rae, B., Heathcote, A., Donkin, C., Averell, L., & Brown, S. (2014). The hare and the tortoise: Emphasizing speed can change the evidence used to make decisions. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *40*, 1226-1243.

Ratcliff, R. (1978). A theory of memory retrieval. *Psychological Review*, *85*, 59-108.

Reddi, B. A. J., & Carpenter, R. H. S. (2000). The influence of urgency on decision time. *Nature Neuroscience*, *3*, 827-830.

Regenwetter, M., Dana, J., & Davis-Stober, C. P. (2011). Transitivity of preferences. *Psychological Review*, *118*, 42-56.

Roberts, S., & Pashler, H. (2000). How persuasive is a good fit? *Psychological Review*, *107*, 358-367.

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

Rouder, J., Morey, R., Gomez, P., & Heathcote, A. (2014). The lognormal race: A cognitive-process model of choice and latency with desirable psychometric properties. *Psychometrika*, *1*, 1-23.

Schley, D. R., & Peters, E. (2014). Assessing 'economic value': Symbolic number mapping predicts risky and riskless valuations. *Psychological Science*, *25*, 753-761.

Schwarz, G. (1978). Estimating the dimension of a model. *Annals of Statistics*, *6*, 461-464.

Shadlen, M. N., & Newsome, W. T. (2001). Neural basis of a perceptual decision in the parietal cortex (area LIP) of the rhesus monkey. *Journal of Neurophysiology*, *86*, 1916-1936.

Shepard, R. N. (1987). Toward a universal law of generalization for psychological science. *Science*, *237*, 1317-1323.

Shiffrin, R. M., Lee, M. D., Kim, W., & Wagenmakers, E.-J. (2008). A survey of model evaluation approaches with a tutorial on hierarchical Bayesian methods. *Cognitive Science*, *32*, 1248-1284.

Simon, H. A. (1982). *Models of bounded rationality: Empirically grounded economic reason.* Cambridge, MA: MIT Press.

Simonson, I. (1989). Choice based on reasons: The case of attraction and compromise effects. *Journal of Consumer Research*, *16*, 158-174.

Simonson, I., & Tversky, A. (1992). Choice in context: Tradeoff contrast and extremeness aversion. *Journal of Marketing Research*, *29*, 281-295.

Slovic, P. (1995). The construction of preference. *American Psychologist*, *50*, 364-371.

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

Stone, M. (1960). Models for choice reaction time. *Psychometrika*, *25*, 251-260.

Summerfield, C., & Tsetsos, K. (2012). Building bridges between perceptual and economic decision-making: neural and computational mechanisms. *Frontiers in Neuroscience*, *6*, 70.

Teodorescu, A. R., & Usher, M. (2013). Disentangling decision models: From independence to competition. *Psychological Review*, *120*, 1-38.

ter Braak, C. J. F. (2006). A Markov chain Monte Carlo version of the genetic algorithm Differential Evolution: easy Bayesian computing for real parameter spaces. *Statistics and Computing*, *16*, 239-249.

Trueblood, J. S., Brown, S. D., & Heathcote, A. (2014). The multiattribute linear ballistic accumulator model of context effects in multialternative choice. *Psychological Review*, *121*, 179-205.

Trueblood, J. S., Brown, S. D., & Heathcote, A. (2015). The fragile nature of contextual preference reversals: Reply to tsetsos, chater, and usher (2015). *Psychological Review*, *122*, 848-853.

Trueblood, J. S., Brown, S. D., Heathcote, A., & Busemeyer, J. R. (2013). Not just for consumers: context effects are fundamental to decision making. *Psychological Science*, *24*, 901-908.

Tsetsos, K., Chater, N., & Usher, M. (2012). Salience driven value integration explains decision biases and preference reversal. *Proceedings of the National Academy of Sciences*, *109*, 9659-9664.

Tsetsos, K., Chater, N., & Usher, M. (2015). Examining the mechanisms underlying contextual preference reversal: Comment on mlba (trueblood, brown, heathcote, 2014). *Psychological Review*, *122*, 838-847.

Tsetsos, K., Gao, J., & McClelland, J. L. (2012). Using time-varying evidence to test models of decision dynamics: Bounded diffusion vs. the leaky competing accumulator model. *Frontiers in Neuroscience*, *6*, 76-79.

Tsetsos, K., Usher, M., & Chater, N. (2010). Preference reversal in multiattribute choice. *Psychological Review*, *117*, 1275-1293.

Tsetsos, K., Usher, M., & McClelland, J. L. (2011). Testing multi-alternative decision models with non-stationary evidence. *Frontiers in Neuroscience*, *5*, 1-18.

Turner, B. M., Dennis, S., & Van Zandt, T. (2013). Bayesian analysis of memory models. *Psychological Review*, *120*, 667-678.

Turner, B. M., Forstmann, B. U., Love, B. C., Palmeri, T. J., & van Maanen, L. (2017). Approaches of analysis in model-based cognitive neuroscience. *Journal of Mathematical Psychology*, *76*, 65-79.

Turner, B. M., Gao, J., Koenig, S., Palfy, D., & McClelland, J. L. (2017). *The dynamics* *of multimodal integration: The averaging diffusion model.* (In press)

Turner, B. M., & Schley, D. R. (2016). The anchor integration model: A descriptive model of anchoring effects. *Cognitive Psychology*, *90*, 1-47.

Turner, B. M., & Sederberg, P. B. (2012). Approximate Bayesian computation with Differential Evolution. *Journal of Mathematical Psychology*, *56*, 375-385.

Turner, B. M., & Sederberg, P. B. (2014). A generalized, likelihood-free method for parameter estimation. *Psychonomic Bulletin and Review*, *21*, 227-250.

Turner, B. M., Sederberg, P. B., Brown, S., & Steyvers, M. (2013). A method for efficiently sampling from distributions with correlated dimensions. *Psychological Methods*.

Turner, B. M., Sederberg, P. B., & McClelland, J. L. (2016). Bayesian analysis of simulation-based models. *Journal of Mathematical Psychology*, *72*, 191-199.

Turner, B. M., & Van Zandt, T. (2012). A tutorial on approximate Bayesian computation. *Journal of Mathematical Psychology*, *56*, 69-85.

Turner, B. M., & Van Zandt, T. (2014). Hierarchical approximate Bayesian computation. *Psychometrika*, *79*, 185-209.

Tversky, A. (1972). Elimination by aspects: A theory of choice. *Psychological Review*, *79*, 281-299.

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

Tversky, A., & Simonson, I. (1993). Context-dependent preferences. *Management Science*, *39*, 1179-1189.

Usher, M., & McClelland, J. L. (2001). On the time course of perceptual choice: The leaky competing accumulator model. *Psychological Review*, *108*, 550-592.

Usher, M., & McClelland, J. L. (2004). Loss aversion and inihibition in dynamical models of multialternative choice. *Psychological Review*, *111*, 757-769.

van Ravenzwaaij, D., van der Maas, H. L. J., & Wagenmakers, E. J. (2012). Optimal decision making in neural inhibition models. *Psychological Review*, *119*, 201-215.

Van den Berg, R., Awh, E., & Ma, W. (2014). Factorial comparison of working memory models. *Psychological Review*, *121*, 124-149.

Mathematical Details of the Models

Here we present the technical details for the base version of each model we fit to the data from the main article. For more details on the motivation of the particular mechanisms used by each model, we encourage the reader to consult the original articles. Although the equations describing each process model are shown here, the prior specifications can be found in the main text.

Multiattribute Linear Ballistic Accumulator (MLBA) model

Consider a multi-alternative choice experiment in which an alternative i has two attributes associated with it, denoted P_i and Q_i. For example, in perceptual experiments, these two attributes could be the width or length of an object; while in consumer choice experiments, these two attributes could indicate price and quality. For ease of notation, assume that three alternatives are presented in the experiment, such that $i = \{1, 2, 3\}$. In the Multiattribute Linear Ballistic Accumulator (MLBA; Trueblood et al., 2014) model, the mean drift rate d_i for alternative i is defined by the following set of equations:

$$\begin{aligned} d_1 &= V_{12} + V_{13} + I_0, \\ d_2 &= V_{21} + V_{23} + I_0, \\ d_3 &= V_{31} + V_{32} + I_0, \end{aligned} \tag{6}$$

where I_0 is a positive baseline input parameter. The values V_{ij} represent a comparison between alternative i and alternative j, expressed by a function of attribute difference. Formally,

$$V_{ij} = W_{P_{ij}}(u_{P_i} - u_{P_j}) + W_{Q_{ij}}(u_{Q_i} - u_{Q_j}), \tag{7}$$

where (u_{P_i}, u_{Q_i}) and (u_{P_j}, u_{Q_j}) represent the subjective mapping of alternative i and j's objective attribute values (P_i, Q_i) and (P_j, Q_j). The mapping function is determined by parameter m in the following way:

$$\begin{aligned} u_{P_i} &= \frac{b}{[\tan^m(\theta) + (\frac{b}{a})^m]^{\frac{1}{m}}}, \\ u_{Q_i} &= \frac{a\tan(\theta)}{[1 + (\frac{a}{b})^m \tan^m(\theta)]^{\frac{1}{m}}}, \end{aligned}$$

where a, b and θ are expressed as:

$$\begin{aligned} a &= P_i - \frac{Q_i(P_j - P_i)}{Q_j - Q_i}, \\ b &= Q_i - \frac{P_i(Q_j - Q_i)}{P_j - P_i}, \\ \theta &= \arctan\left(\frac{Q_i}{P_i}\right). \end{aligned}$$

The values u_{P_j} and u_{Q_j} are calculated in an equivalent way, where the index i is replaced by j in the equations above.

In Equation 7, $W_{P_{ij}}$ and $W_{Q_{ij}}$ reflect the amount of weight given to a particular attribute comparison, which is expressed as:

$$\begin{aligned} W_{P_{ij}} &= \exp\left(-\lambda|u_{P_i} - u_{P_j}|\right), \\ W_{Q_{ij}} &= \exp\left(-\lambda\beta|u_{Q_i} - u_{Q_j}|\right). \end{aligned}$$

If the subjective attribute value difference is non-negative (e.g. $u_{P_i} - u_{P_j} \geq 0$), $\lambda = \lambda_1$, otherwise $\lambda = \lambda_2$. The parameter β is an attribute bias parameter, such that when $\beta = 1$, attribute P and attribute Q are considered equally. When $\beta > 1$, there is a bias toward attribute Q, and when $0 < \beta < 1$, the bias is toward attribute P.

In summary, five parameters are estimated in MLBA model: one constant baseline input parameter I_0, one mapping parameter m, two decay parameters λ_1 and λ_2, and one attribute bias parameter β. Once the mean drift rates have been calculated in Equation 6, they are passed through to the LBA model (Brown & Heathcote, 2008) to generate a response. This LBA process requires another set of parameters χ, A, and s, but are typically fixed for model identifiability purposes.

Multialternative Decision Field Theory (MDFT) model

Maintaining the same scenario as the section above, consider three alternatives such as $\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z}$, possessing two attributes for each alternative, such that $\boldsymbol{x}$, $\boldsymbol{y}$, and $\boldsymbol{z}$ are vectors with two components P and Q. In the Multialternative Decision Field Theory (MDFT; Roe et al., 2001) model, these attribute values are arranged in the information matrix M such that

$$M = \begin{bmatrix} x_P & x_Q \\ y_P & y_Q \\ z_P & z_Q \end{bmatrix}.$$

The MDFT model assumes a set of attention weights W for attributes, which can be written as a vector evolving over time t:

$$W(t) = \begin{bmatrix} w_P(t) & w_Q(t) \end{bmatrix}^{\intercal}.$$

In the MDFT model, attention for each attribute P and Q follows a Bernoulli process such that the probability of attending to dimension P is ω, and the probability of attending to dimension Q is $(1 - \omega)$. To implement this process, at each moment in time, a variable p is sampled from a Bernoulli distribution with probability parameter ω.

In the three alternative case, the comparison between the alternatives is represented in the contrast matrix

$$C = \begin{bmatrix} 1 & -\frac{1}{2} & -\frac{1}{2} \\ -\frac{1}{2} & 1 & -\frac{1}{2} \\ -\frac{1}{2} & -\frac{1}{2} & 1 \end{bmatrix}.$$

The "valence" of each alternative at time t is then evaluated through a product of the attribute values of the stimuli (i.e., M), the attention weight matrix W, and the contrast matrix C:

$$V(t) = CMW(t) + \boldsymbol{\epsilon}(t),$$

where $\boldsymbol{\epsilon}(t)$ is zero-centered Gaussian noise at time t with standard deviation Σ (i.e., $\boldsymbol{\epsilon}(t) \sim \mathcal{N}(0, \Sigma)$). The overall preference state $P(t) = [P_x(t), P_y(t), P_z(t)]$ at time t for each of the alternatives is then determined by multiplying the preference states by a "feedback" matrix S and adding the valance matrix from above, according to the following equation:

$$P(t) = SP(t-1) + V(t).$$

The feedback matrix S is determined by differences in the attribute dimensions across alternatives $\Delta P = P_i - P_j$ and $\Delta Q = Q_i - Q_j$. Specifically, the element S_{ij} represents a comparison between alternatives i and j, such that

$$S_{ij} = \begin{cases} 1-\phi_2 & \text{if } i = j \\ -\phi_2 \exp\left(-\phi_1 Dist_{ij}^2\right) & \text{if } i \neq j, \end{cases}$$

where

$$Dist_{ij} = (\Delta I)^2 + \beta(\Delta D)^2,$$

and

$$\begin{aligned} \Delta I &= \frac{1}{\sqrt{2}}(\Delta P - \Delta Q) \\ \Delta D &= \frac{1}{\sqrt{2}}(\Delta P + \Delta Q) \end{aligned}$$

In summary, five parameters are estimated in MDFT model: three parameters related to the feedback matrix S: ϕ_1, ϕ_2 and β, one error term Σ, and one attention weight parameter ω.

Associations and Accumulation model (AAM)

Using similar notation as in the MDFT model above, the Associations and Accumulation model (AAM; Bhatia, 2013) starts by calculating attribute activation. Specifically, for attribute j,

$$a_j = x_j + y_j + z_j + a_0^j$$

where a_0^j is a basis parameter for each attribute dimension j. The attribute weight or accessibility pa_j is calculated as the proportion of a_j:

$$pa_j = \frac{a_j}{\sum_{j=1}^{2} a_j}$$

This normalization step ensures that the attribute weights are constrained to sum to one. The valence matrix V is calculated with the parameter α such that

$$V = \begin{bmatrix} \text{sign}(x_1)|x_1|^\alpha & \text{sign}(x_2)|x_2|^\alpha \\ \text{sign}(y_1)|y_1|^\alpha & \text{sign}(y_2)|y_2|^\alpha \\ \text{sign}(z_1)|z_1|^\alpha & \text{sign}(z_2)|z_2|^\alpha \end{bmatrix}$$

The preference vector is determined in a similar fashion as in the MDFT model above, by multiplying the previous state of the preference by a feedback matrix S, adding the valence matrix V, and a noise term $\boldsymbol{\epsilon}(t)$, such that

$$P_i(t) = SP_i(t-1) + V_i[w, 1-w]^\intercal + \boldsymbol{\epsilon}(t),$$

where $\boldsymbol{\epsilon}(t) \sim \mathcal{N}(0, e)$, and $w \sim \text{Bernoulli}(pa_1)$. The feedback matrix is again constructed with a combination of the leakage or decay parameter d, and the lateral inhibition parameter l, such that

$$S = \begin{bmatrix} d & -l & -l \\ -l & d & -l \\ -l & -l & d \end{bmatrix}.$$

In summary, AAM estimates five parameters: two parameters in feedback matrix d, l, two attribute related parameters a_0^j (i.e., one per attribute dimension j), α, and one error term e.

Multiattribute Leaky Competing Accumulator model (MLCA)

The Multiattribute Leaky Competing Accumulator model (MLCA; Usher & McClelland, 2004) model is structurally very similar to the MDFT model. The MLCA model assumes that preferences fluctuate stochastically over time, and this process can be represented mathematically by multiplying the previous preference state by a feedback matrix S, an indicator function $\boldsymbol{I}(t)$, and adding random noise $\epsilon(t)$ according to the following equation:

$$P(t+1) = SP(t) + (1-k)\left[\boldsymbol{I}(t) + \boldsymbol{\epsilon}(t)\right], \tag{8}$$

where k is the leakage parameter, $\boldsymbol{\epsilon}(t)$ is zero-centered Gaussian noise with standard deviation η (i.e., $\boldsymbol{\epsilon}(t) \sim \mathcal{N}(0, \eta)$). The feedback matrix S consists of a lateral inhibition parameter L and leakage k, such that

$$S = \begin{bmatrix} k & -L(1-k) & -L(1-k) \\ -L(1-k) & k & -L(1-k) \\ -L(1-k) & -L(1-k) & k \end{bmatrix}.$$

In Equation 8, $\boldsymbol{I}(t)$ is a vector with components $I_i(t)$ corresponding to each of the choice alternatives such that

$$I_i(t) = I_0 + \sum_{j \neq i}^{n-1} V(d_{ij}(t)),$$

where I_0 is a positive baseline input parameter. The variable d_{ij} represents the pairwise differences between options i and j on the dimension P or Q, depending on which dimension is being sampled at a given moment in time. To implement this sampling process, at each moment in time, a random variable p is sampled from a Bernoulli distribution with probability ω. If $p = 1$, then attribute dimension P is considered at that moment, and if $p = 0$, attribute dimension Q is considered. For example, when dimension P is sampled (i.e., $p = 1$), $d_{ij} = P_i - P_j = \Delta P$. The value function $V(x)$ is used by the MLCA model to convert the objective quantities to subjective values that are used in the evidence accumulation process. The MLCA model departs from the other models because it uses the concept of loss aversion (Kahneman & Tversky, 1979; Usher & McClelland, 2004) to filter the pairwise differences. Explicitly, the value function

$$V(x) = \begin{cases} z(x) & \text{if } x \geq 0 \\ -\left(z(|x|) + z(|x|)^2\right) & \text{if } x < 0, \end{cases},$$

where

$$z(x) = \log(1 + x).$$

For this model only, we rescaled the attribute values from the $(0, 10)$ scale to the $(0, 1)$ scale, as used in the original paper (Usher & McClelland, 2004).

Author Note

This research was supported by National Science Foundation grant 1358507. The authors would like to thank Benjamin Scheibehenne for the data in our Study 1, and Ting Wang for help with some initial simulations and technical summaries found in the appendix.

Footnotes

[1]Note that in Usher and McClelland (2004), this model was simply referred to as the Leaky Competing Accumulator (LCA) model, building off the original model developed in Usher and McClelland (2001). However, we use the term “multiattribute” to dissociate it from the LCA model because it makes different assumptions about how options are subjectively evaluated (i.e., the “front end” portion of the model). This naming convention is analogous to the Multiattribute Linear Ballistic Accumulator model (Trueblood et al., 2014).

[2]However, the differences are transformed in a nonlinear fashion in both the MLBA and MLCA models, as we discuss in the next section.

[3]We also evaluated the Bayesian predictive information criterion (BPIC; Ando, 2007), but found that the results were nearly identical for every analysis in the article. As such, we do not present these measures.

[4]In some unreported analyses, we found that including filler trials when fitting a model to data can sometimes dramatically alter the shape of the posterior distributions in ways that suggest the models are better constrained by these additional trials.

[5]Study 2 also contained a consumer choice task analogous to the reported perceptual choice task. The consumer task included 450 binary, and 495 ternary, choice trials counter-balanced with the perceptual task. Participants chose between products presented in a table format (2×2 in binary and 2×3 in ternary choices), with attributes in rows and alternatives in columns. Including both perceptual and consumer sessions, participants completed 1890 experimental trials over the two days and four tasks. For reasons of space and parsimony, we decided not to report the consumer choice task. Fitting the consumer data to the models would require estimating the attribute weights for each participant and each product category, introducing additional free parameters in all models. Initial explorations found that some hierarchical models were unidentifiable without fixing certain parameters across the consumer goods. As we felt that this choice

could be seen as a contentious one, we elected to save the discussion of this modeling problem for future work.

[6]While most context effects experiments deal with only two attributes, multivariate attribute spaces are a natural extension in which case a Dirichlet process would be used rather than a Bernoulli process.

Table 2

Fit statistics for each variant of the four models. Each value represents the mean statistic obtained across all chains in the sampling algorithm. The deviance information criterion (DIC), the effective number of parameters p_D, the average deviance $\bar{D}$, and the best deviance value obtained $\hat{D}$ are shown in the third, fourth, fifth and sixth columns, respectively. Deviance is defined as negative two times the log likelihood value (i.e., lower values are preferred). The best-fitting model variant is shown in bold-face type.

Model	Features	DIC	p_D	$\bar{D}$	$\hat{D}$
MDFT 1.0	base	114.56	6.46	101.65	95.19
MDFT 2.0	attribute bias	**100.41**	11.44	77.53	66.09
MLCA 1.0	base	170.13	13.22	143.69	130.47
MLCA 2.0	attribute bias	**148.09**	12.03	124.04	112.01
AAM 1.0	base	356.54	23.98	308.58	284.59
AAM 2.0	lateral inhibition	362.90	25.26	312.38	287.12
AAM 3.0	attribute bias	158.98	16.32	126.34	110.02
AAM 4.0	attribute bias, inhibition	**112.06**	14.53	83.00	68.48
MLBA 1.0	base	255.88	18.83	218.22	199.40
MLBA 2.0	with attribute bias	**181.84**	16.81	148.22	131.41

Table 3

Fit statistics for each variant of the four hierarchal models applied to Study 1. Each value represents the mean statistic obtained across all chains in the sampling algorithm. The deviance information criterion (DIC), the effective number of parameters p_D*, the average deviance* $\bar{D}$*, and the best deviance value obtained* $\hat{D}$ *are shown in the third, fourth, fifth and sixth columns, respectively. Deviance is defined as negative two times the log likelihood value (i.e., lower values are preferred). The best-fitting model variant is shown in bold-face type.*

Model	DIC	p_D	$\bar{D}$	$\hat{D}$
HMDFT 1.0	**2219.12**	183.66	2035.46	1851.80
HMDFT 2.0	2948.56	224.87	2723.69	2498.83
HMLCA 1.0	**1782.40**	152.37	1630.03	1477.65
HMLCA 2.0	1976.80	190.47	1786.33	1595.85
HAAM 1.0	**1628.90**	137.66	1491.24	1353.59
HAAM 2.0	1642.45	144.25	1498.20	1353.95
HMLBA 1.0	**1837.43**	177.20	1660.23	1483.04
HMLBA 2.0	2033.31	161.43	1871.88	1710.45

Table 4

Experimental conditions in binary and ternary perceptual trials. The attribute values in the two dimensions (D1 and D2 for width and height, respectively) are here given in a 0-10 virtual scale. The virtual values were exponentiated in order to derive the rectangles' dimensions in pixels. After the exponentiation random jitters drawn uniformly from the 0-3 interval (i.e. in units of pixels) were added on each dimension of each rectangle.

		A		B		C		
Choice Set	Condition	D1	D2	D1	D2	D1	D2	Trials
Binary	B1	4.5	5.5	5.5	4.5			90
Binary	B2	4.5	5.5	5	5			90
Binary	B3	5.5	4.5	5	5			90
Binary	B4	4.5	5.5	4.36	5.36			45
Binary	B5	5.5	4.5	5.36	4.36			45
Binary	B6	5	5	4.36	5.36			45
Binary	B7	5	5	5.36	4.36			45
Ternary	C1	4.5	5.5	5.5	4.5	4.36	5.36	60
Ternary	C2	4.5	5.5	5.5	4.5	5.36	4.36	60
Ternary	C3	4.5	5.5	5.5	4.5	4.64	5.36	60
Ternary	C4	4.5	5.5	5.5	4.5	5.36	4.64	60
Ternary	C5	4.5	5.5	5.5	4.5	5	5	120
Ternary	C6	4.5	5.5	5.36	4.36	4.36	5.36	45
Ternary	C7	5.5	4.5	6.43	5.36	4.36	4.36	45
Ternary	C8	5	5	4.36	5.36	5.36	4.36	45

Table 5

Fit statistics for each variant of the four models from Study 2. Each value represents the mean statistic obtained across all chains in the sampling algorithm. The deviance information criterion (DIC), the effective number of parameters p_D*, the average deviance* $\bar{D}$*, and the best deviance value obtained* $\hat{D}$ *are shown in the third, fourth, fifth and sixth columns, respectively. Deviance is defined as negative two times the log likelihood value (i.e., lower values are preferred). The best-fitting model variant is shown in bold-face type.*

Model	DIC	p_D	$\bar{D}$	$\hat{D}$
MDFT 1.0	**3760.78**	536.06	3224.72	2688.67
MDFT 2.0	3880.53	582.56	3297.97	2715.41
MLCA 1.0	2255.91	373.80	1882.10	1508.30
MLCA 2.0	**1921.40**	361.86	1559.55	1197.69
AAM 2.0	1912.31	387.56	1524.75	1137.18
AAM 4.0	**1620.50**	328.93	1291.57	962.65
MLBA 1.0	**2406.44**	321.43	2085.01	1763.58
MLBA 2.0	2470.51	385.76	2084.75	1698.99

Table 6

Fit statistics for each variant of the four hierarchal models applied to Study 2. Each value represents the mean statistic obtained across all chains in the sampling algorithm. The deviance information criterion (DIC), effective number of parameters p_D*, average of the deviances* $\bar{D}$*, and highest deviance value obtained* $\hat{D}$ *are shown in the second, third, fourth, and fifth columns, respectively. Deviance is defined as negative two times the log likelihood value (i.e., lower values are preferred).*

Model	DIC	p_D	$\bar{D}$	$\hat{D}$
HMDFT 1.0	16314.42	925.45	15388.98	14463.53
HMDFT 2.0	**13858.75**	865.02	12993.73	12128.7
HMLCA 1.0	16125.25	926.75	15198.50	14271.74
HMLCA 2.0	**12889.65**	772.88	12116.78	11343.90
HAAM 1.0	13284.60	824.41	12460.18	11635.77
HAAM 2.0	**10875.25**	693.96	10181.29	9487.33
HMLBA 1.0	15088.36	861.69	14226.67	13364.98
HMLBA 2.0	**13295.92**	779.62	12516.3	11736.68

Table 7

Summary of extant model fits to data. The ranking of the best-fitting model variant for each of the extant models (rows) is shown for each data analysis (columns) we performed.

	Consumer Goods		Perceptual		
Model	Aggregated	Hierarchical	Aggregated	Hierarchical	Switchboard
MDFT	1	4	4	4	3
MLCA	3	2	2	2	1
AAM	2	1	1	1	2
MLBA	4	3	3	3	4

Figure Captions

Figure 1. Illustrations of the architecture of each of the four extant models of context effects. Green nodes correspond to attribute values of the stimuli, blue nodes correspond to the integrated representations of the stimuli, and red nodes correspond to the preference states for the stimuli. Arrows indicate directions of influence in the diagram, where double-headed arrows indicate bi-directional influence (e.g., at the decision process).

Figure 2. Model predictions (gray contours) against the observed data (black "+" sign). Each panel is a ternary plot that shows the relative probabilities of choosing the target (T), distractor (D), and competitor (C) options, where the indifference curves corresponding to each of the three options are shown as the dashed gray lines. The rows correspond to the four best-fitting version of each model: MDFT (first), MLCA (second), AAM (third), and MLBA (fourth). The columns correspond to the three context effects: attraction (left), compromise (middle), and similarity (right).

Figure 3. Model predictions (yellow clouds) against the observed data (black circles). Each panel is a ternary plot that shows the relative probabilities of choosing the target (T), distractor (D), and competitor (C) options, where the indifference curves corresponding to each of the three options are shown as the dashed gray lines. The rows correspond to the four best-fitting version of each model: AAM (first), MDFT (second), MLBA (third), and MLCA (fourth). The columns correspond to the three context effects: attraction (left), compromise (middle), and similarity (right).

Figure 4. Model predictions (y-axis) against the observed data (x-axis) for each of the four models: MDFT (top left), MLCA (top right), AAM (bottom left), and MLBA (bottom right) models. Correlation values for each model are reported in the corresponding panel. Each point corresponds to a particular response probability for a

particular stimulus, aggregated across subjects.

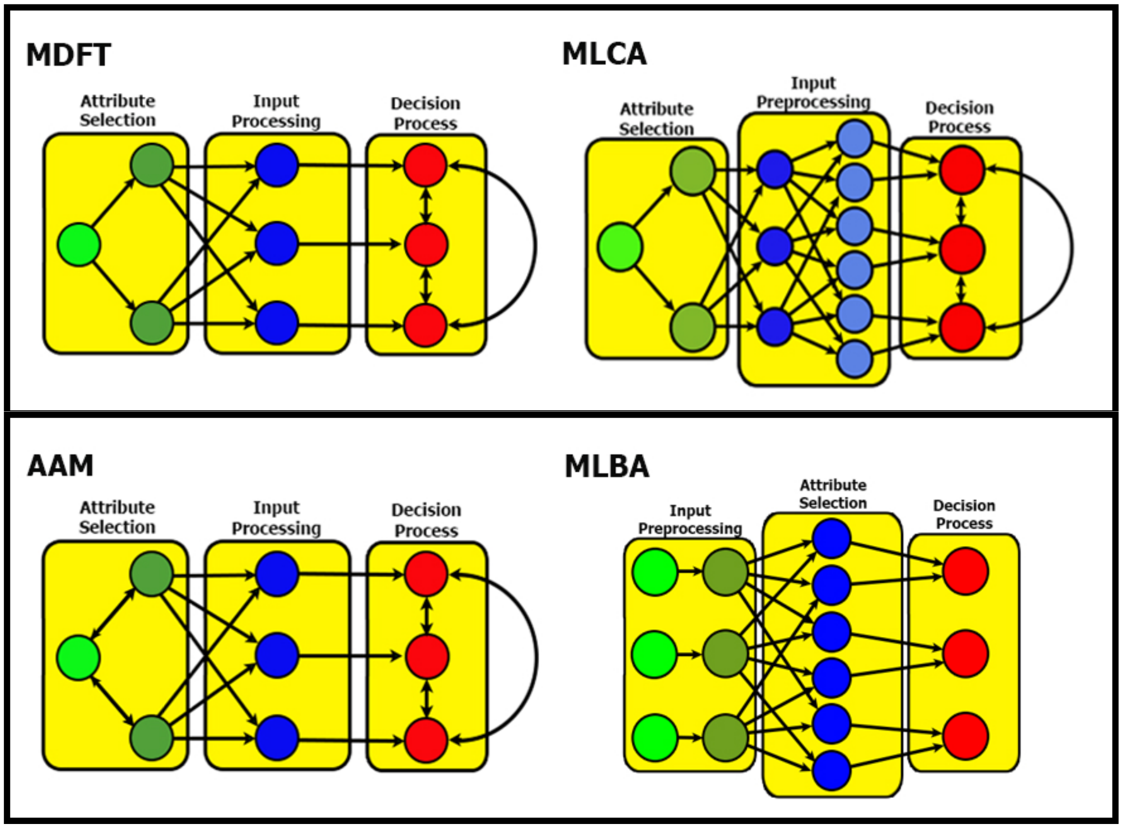
Figure 5. Model predictions (y-axis) against the observed data (x-axis) for each of the four models: MDFT (top left), MLCA (top right), AAM (bottom left), and MLBA (bottom right) models. Predictions for each subject's response on each of the 15 stimuli are shown. Correlation values are shown for each model in the corresponding panel. Each point corresponds to a particular response probability from a particular subject for a particular stimulus.

Figure 6. Magnitude of context effects in the data (black dots) compared to predictions from the four best-fitting hierarchical versions of the models: HMDFT 2.0 (first row), HMLCA 2.0 (second row), HAAM 2.0 (third row), and HMLBA 2.0 (fourth row). The columns correspond to the various joint distributions of context effects: similarity vs. compromise (first column), similarity vs. attraction (second column), and attraction vs. compromise (third column). In each panel, dashed lines represent points of zero effect.

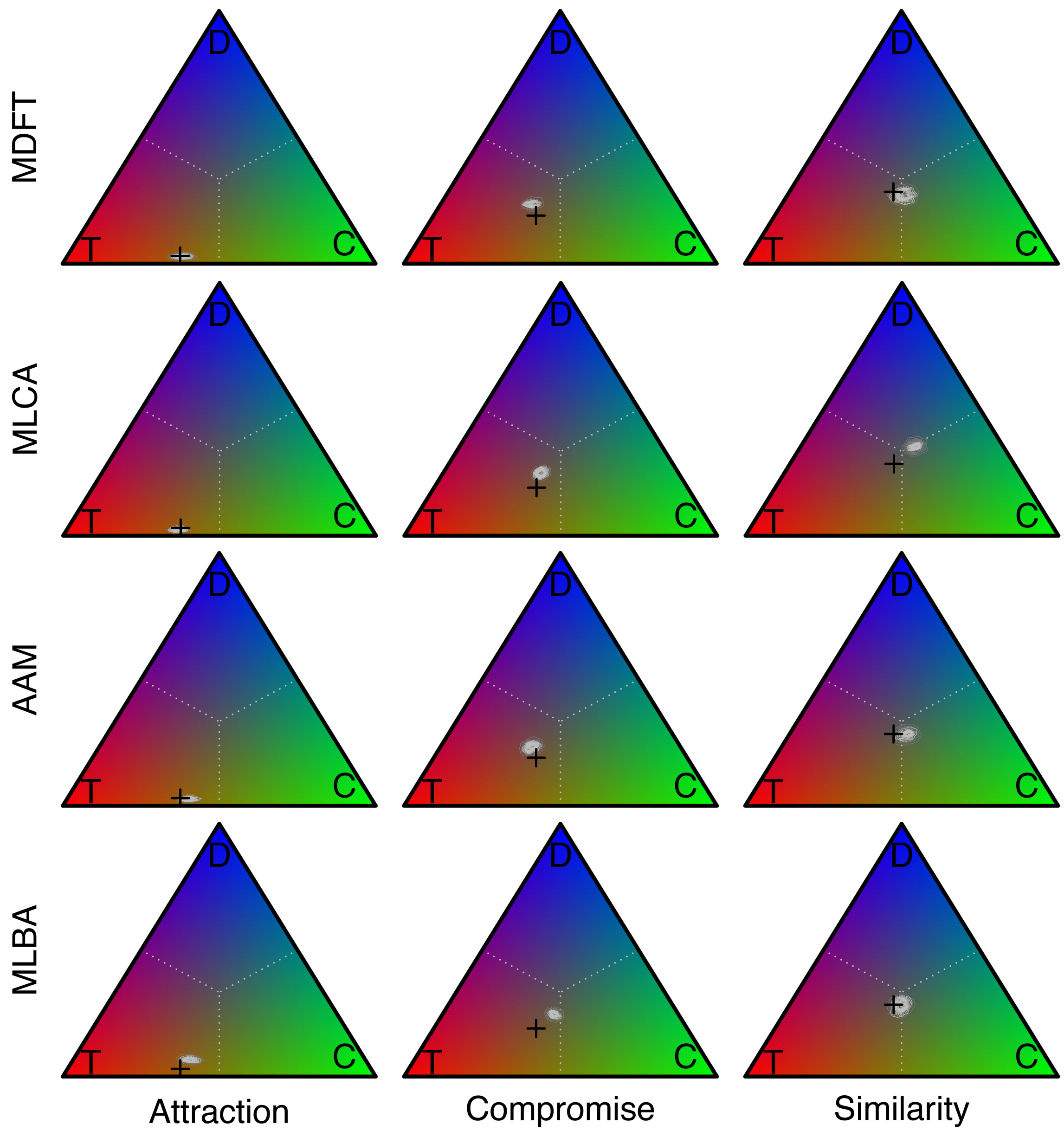
Figure 7. Relative model fits for the switchboard analysis. The performance of each model is assessed via the BIC statistic, which is color coded according to the legend on the right-hand side (i.e., lower values of BIC are better). The plot is organized according to the particular value of the "switch" at each node in the model-making path. Along the rows, the outermost factor is competition, followed by the type of valuation noise, followed by the type of processing (i.e., subjective mapping of attribute values). Along the columns, the outermost factor is attribute integration, followed by the type of attention, followed by distances computed, followed by filtration. Each factor is labeled according to the seven nodes discussed in the text.

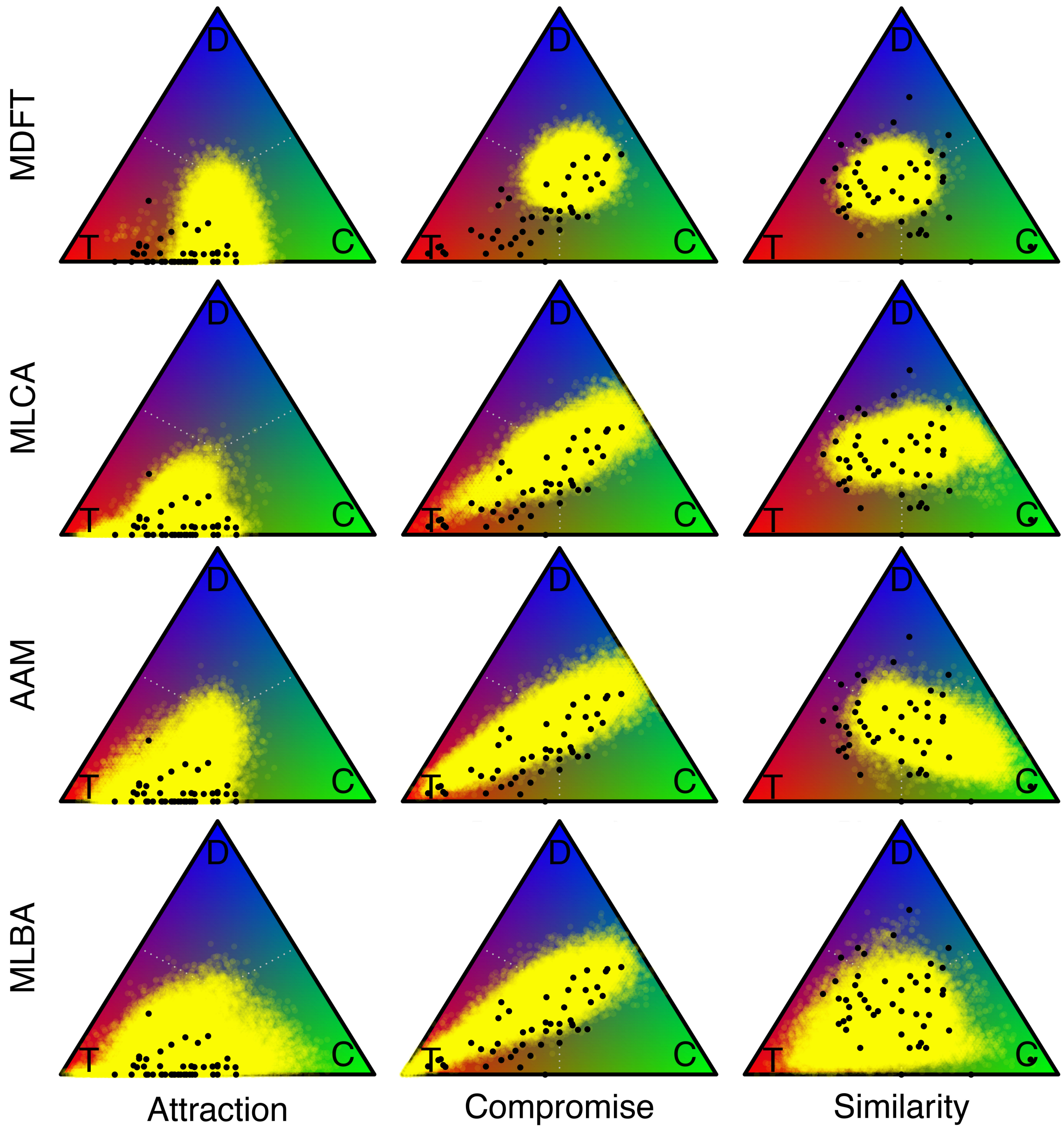
Figure 8. Evaluating specific model mechanisms. Each bar shows the median BIC statistic across all models for every node by switch combination. The dashed vertical line

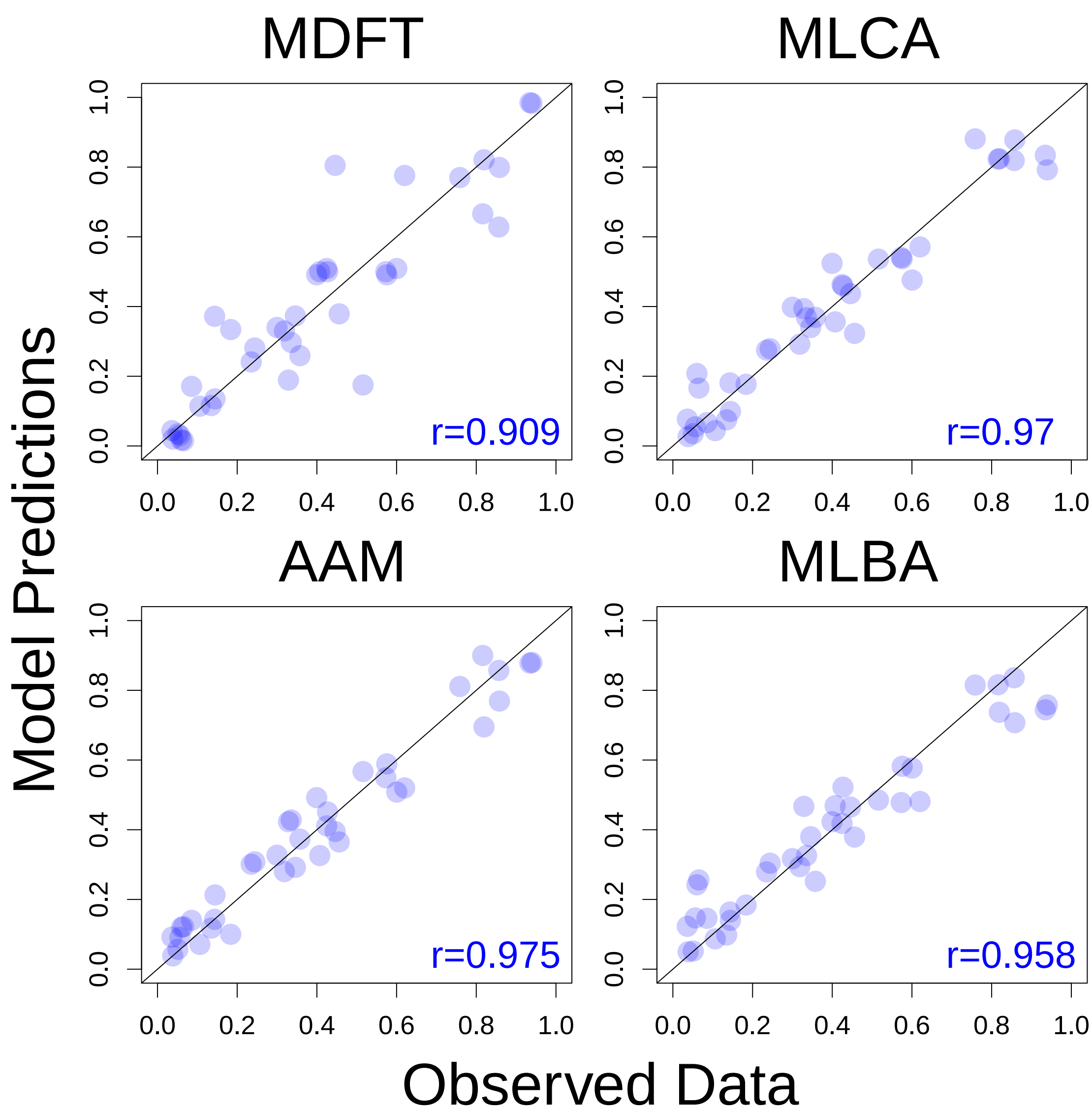
represents the global median BIC value as a visual guide. Nodes are ordered to correspond to the description of the model-making path in the main text.



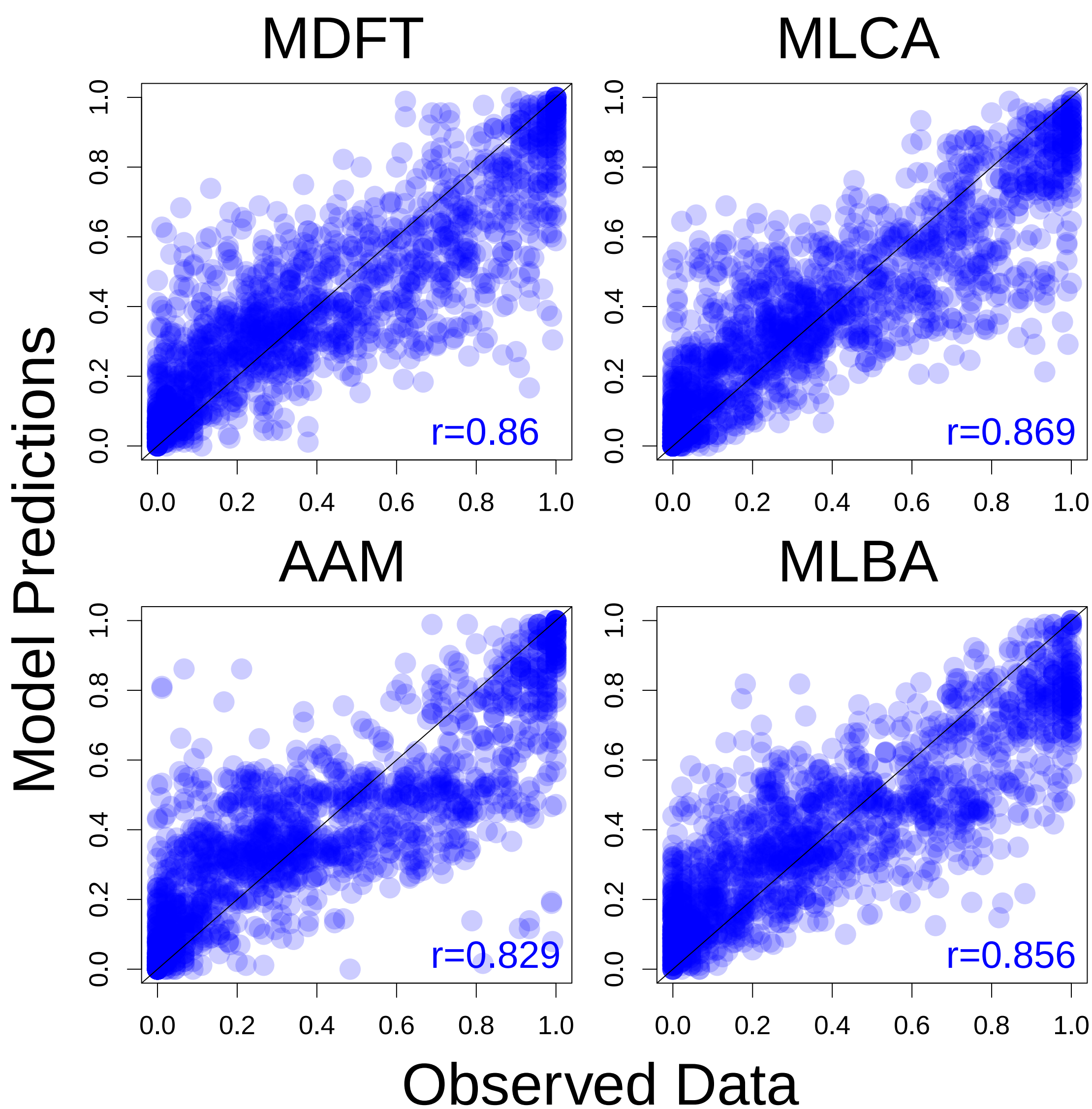
MDFT
Attribute Selection
Input Processing
Decision Process
MLCA
Attribute Selection
Input Preprocessing
Decision Process
AAM
Attribute Selection
Input Processing
Decision Process
MLBA
Input Preprocessing
Attribute Selection
Decision Process

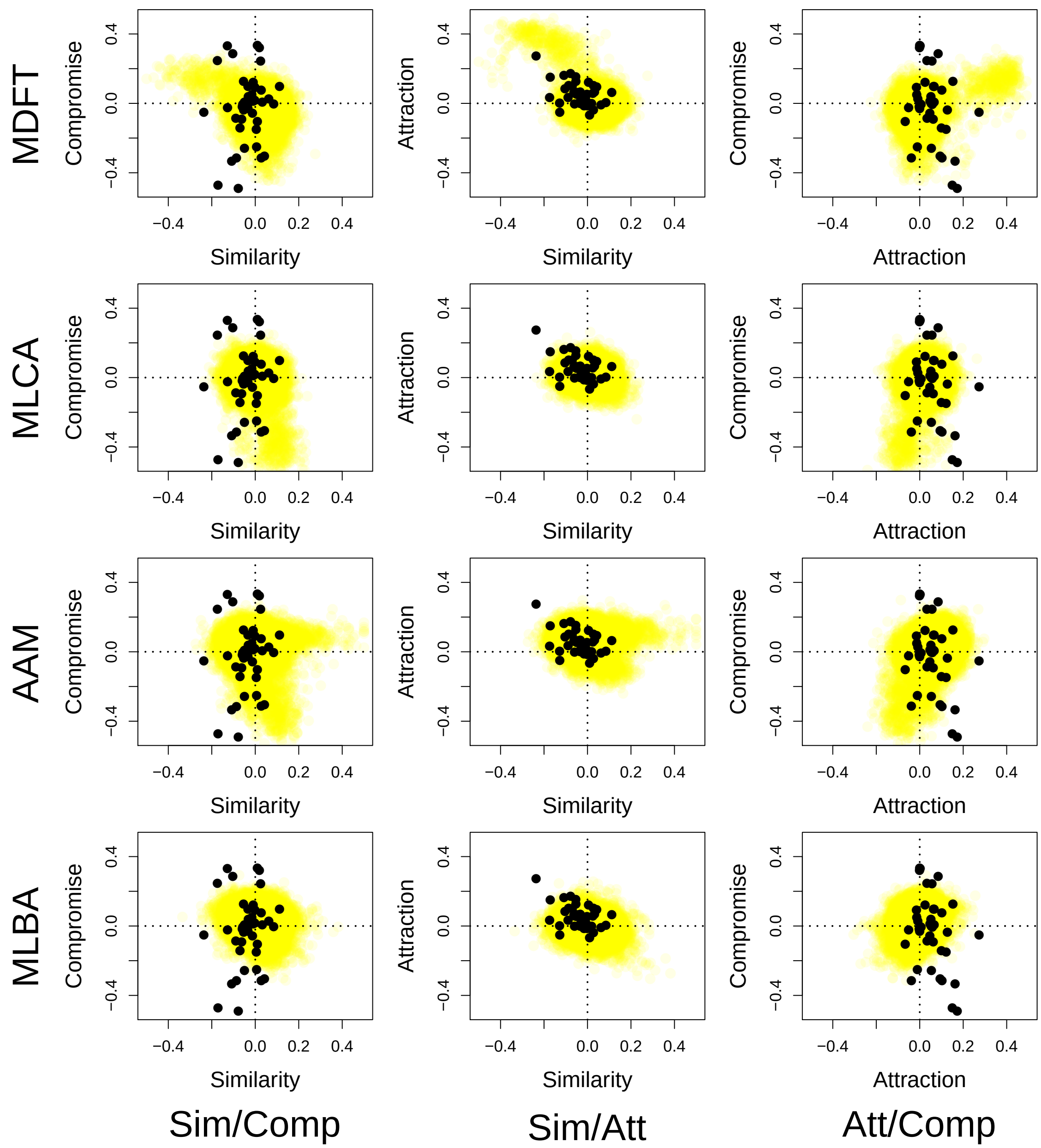

MDFT
MLCA
AAM
MLBA
D
T
C
Attraction
Compromise
Similarity

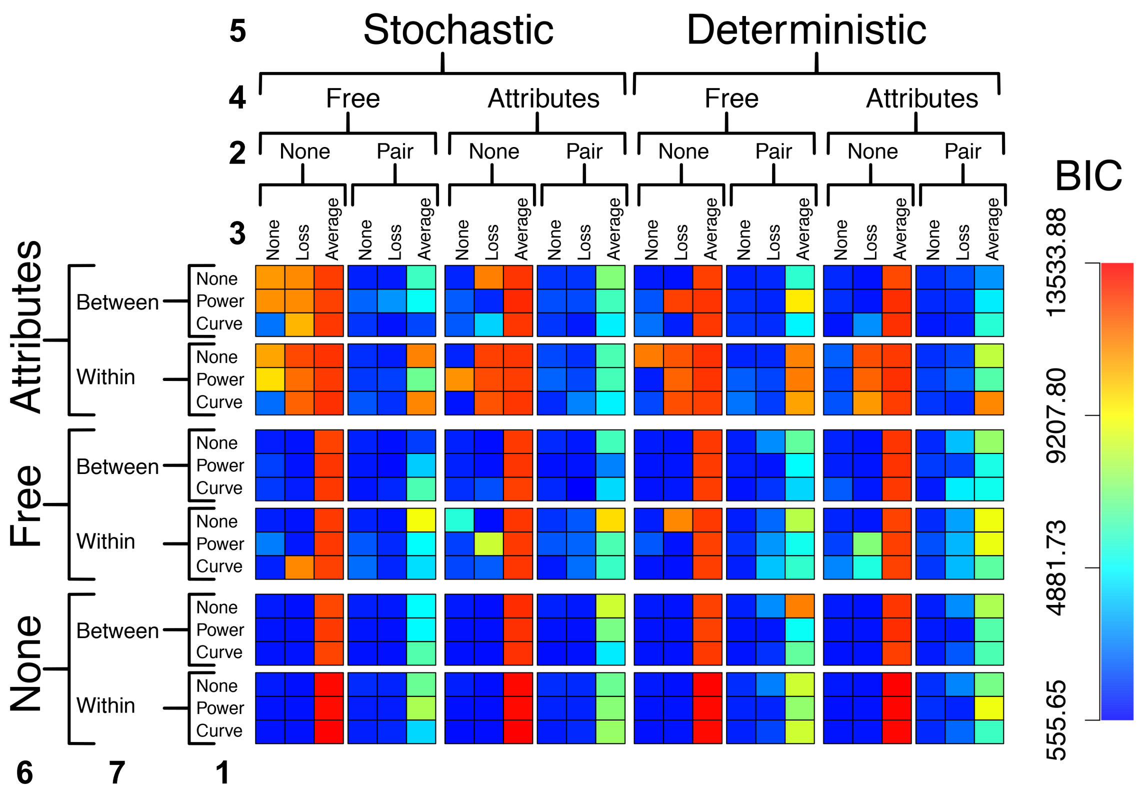

MDFT
MLCA
AAM
MLBA
D
T
C
Attraction
Compromise
Similarity

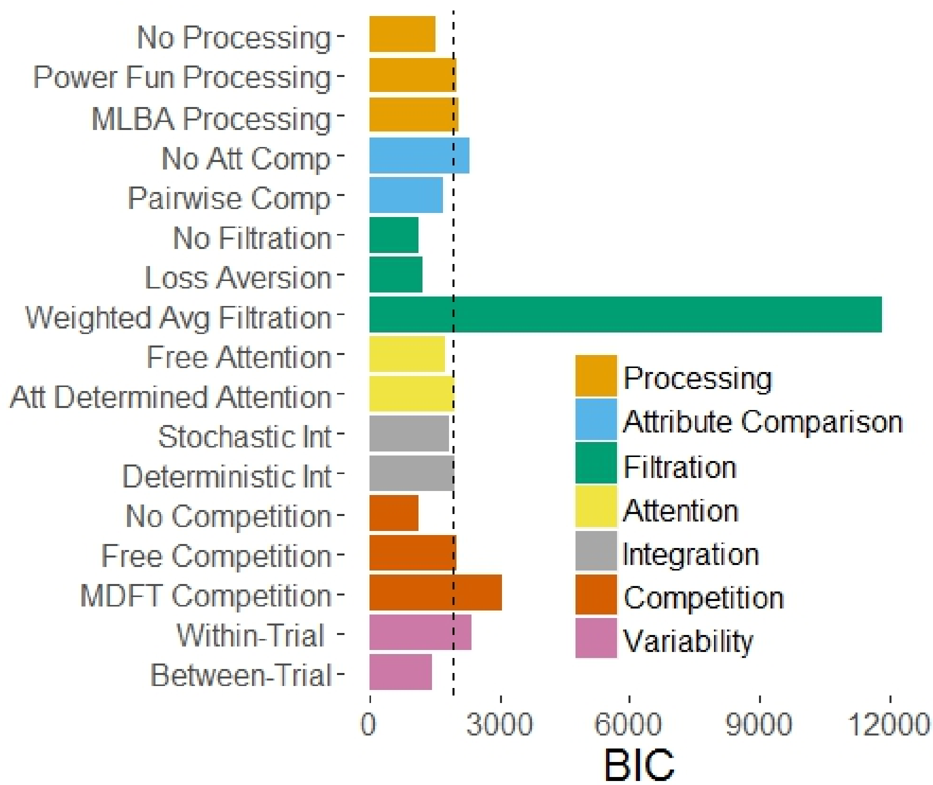


MDFT
MLCA
AAM
MLBA
r=0.909
r=0.97
r=0.975
r=0.958
0.0 0.2 0.4 0.6 0.8 1.0
Model Predictions
Observed Data


MDFT
MLCA
AAM
MLBA
r=0.86
r=0.869
r=0.829
r=0.856
Model Predictions
Observed Data
0.0
0.2
0.4
0.6
0.8
1.0


MDFT
MLCA
AAM
MLBA
Compromise
Attraction
Similarity
−0.4
0.0
0.2
0.4
Sim/Comp
Sim/Att
Att/Comp


5
Stochastic
Deterministic
4
Free
Attributes
Free
Attributes
2
None
Pair
None
Pair
None
Pair
None
Pair
3
None
Loss
Average
BIC
13533.88
9207.80
4881.73
555.65
Attributes
Free
None
Between
Within
None
Power
Curve
6
7
1


No Processing
Power Fun Processing
MLBA Processing
No Att Comp
Pairwise Comp
No Filtration
Loss Aversion
Weighted Avg Filtration
Free Attention
Att Determined Attention
Stochastic Int
Deterministic Int
No Competition
Free Competition
MDFT Competition
Within-Trial
Between-Trial
0
3000
6000
9000
12000
BIC
Processing
Attribute Comparison
Filtration
Attention
Integration
Competition
Variability