Cognitive Psychology 109 (2019) 47–67

Contents lists available at ScienceDirect

Cognitive Psychology

journal homepage: www.elsevier.com/locate/cogpsych

Preference accumulation as a process model of desirability ratings

Sudeep Bhatia[a,*,1], Timothy J. Pleskac[b,c]

[a] University of Pennsylvania, United States
[b] University of Kansas, United States
[c] Max Planck Institute for Human Development, Germany

ARTICLE INFO

Keywords:
Decision making
Desirability ratings
Accumulation
Sequential sampling
Computational models

ABSTRACT

In desirability rating tasks, decision makers evaluate objects on a continuous response scale. Despite their prominence, full process models of these rating tasks have not been developed. We investigated whether a preference accumulation process, a process often used to model discrete choice, might explain ratings as well. According to our model, attributes from each option are sampled and evaluated stochastically. The evaluations are integrated over time, forming a preference. Preferences for options compete with each other, and accumulated preferences can decay. The model makes precise predictions regarding the statistical distribution of desirability ratings, as well as their dependence on deliberation time and on context. We test and confirm these predictions in two experimental studies. Additionally, quantitative model fits indicate that participants are better described by our proposed model, relative to a model without dynamism, competition, or stochastic attribute sampling. Our results show that the descriptive power of models of preference accumulation extends beyond discrete choice, and that the assumptions of this framework accurately characterize the core cognitive processes at play in the construction of preference and the evaluation of objects.

1. Introduction

Preference has sometimes been thought of as synonymous with choice. This assumption is the cornerstone of revealed preference theory, and a guiding principle in economics (Becker, 1976; Samuelson, 1947). However, theories based on revealed preference have repeatedly been shown to have a limited ability to predict the choices that people actually make (Camerer, Loewenstein & Rabin, 2011; Weber & Johnson, 2009). Instead, many of the more successful descriptive theories of decision making predict choice by specifying the cognitive processes involved in constructing preference, and in mapping constructed preferences onto choice.

One popular approach to specifying these cognitive processes is known as the preference accumulation framework (e.g., Bhatia, 2013; Bhatia & Mullett, 2016; Diederich, 1997; Krajbich, Armel & Rangel, 2010; Pleskac, Yu, Hopwood, & Liu, in press; Roe, Busemeyer, & Townsend, 2001; Trueblood, Brown, & Heathcote, 2014; Tsetsos, Chater & Usher, 2012; Turner, Schley, Muller & Tstsos, 2018; Usher & McClelland, 2004). Models within this framework typically posit that attributes are sampled stochastically, that attribute values are dynamically and competitively aggregated into preferences, and that this accumulated preference then determines the choices people make. Yet, choices are one of several possible ways by which preferences can be elicited. For instance, be it in psychology or in other fields, preferences are often elicited using desirability ratings (e.g. Green & Srinivasan, 1978; Likert, 1932; Nunnally, 1967) and are in fact being used more and more to study the nature of preferences (e.g., Betsch, Plessner, Schwieren, &

* Corresponding author at: Department of Psychology, University of Pennsylvania, Philadelphia, PA, United States.
E-mail address: bhatiasu@sas.upenn.edu (S. Bhatia).
[1] Funding for Sudeep Bhatia was received from the National Science Foundation grant SES-1626825.

https://doi.org/10.1016/j.cogpsych.2018.12.003
Accepted 20 December 2018

Gütig , 2001; Betsch, Kaufmann, Lindow, Plessner, & Hoffmann, 2006; Brusovansky, Vanunu, & Usher, in press; Vanunu, Pachur, & Usher, in press). When people are asked to provide desirability ratings they are given one or more objects to evaluate, and instead of indicating their preferences through a discrete choice they are asked to rate the objects on a continuous numeric scale. The higher they rate each object, the more favorable they are toward that object. To the extent that choice and desirability ratings both rely on common evaluative processes (i.e. common processes for constructing preferences), we should expect theories of choice to extend to tasks eliciting desirability ratings. Thus, if desirability ratings are determined by a preference accumulation process, then individuals in a desirability rating task should sample attributes and accumulate attribute values into preferences, just as they would in a discrete choice task. However, instead of selecting between one of two options based on the preference state, individuals should map the accumulated preference onto a response scale.

This extension of using a preference accumulation framework to model the process of selecting a desirability rating gives rise to three unique predictions about the properties of desirability ratings. The first property is *preference correlation:* Stochastic attribute sampling causes preferences to become correlated, with higher correlations between preferences for similar objects relative to dissimilar objects (even if object desirability is kept constant). The second property is *preference change:* Preferences are accumulated, and thus evolve over the time course of the decision, with preferences for more desirable objects increasing more quickly (or decreasing less quickly) than preferences for less desirable objects. The third property is *preference competition:* Preferences compete with each other, so that objects evaluated alongside highly desirable competitors (i.e. competitors whose preferences increase at a fast rate and reach high values) have lower preferences than objects evaluated alongside less desirable competitors (i.e. competitors whose preferences increase at slow rate, or decrease, and do not reach high values).

These three properties have been used to explain a wide range of behavioral effects in preferential choice. For example, preference correlation, generated by stochastic attribute sampling, can lead to differences in choice probabilities depending on the similarity between objects, explaining violations of the proportionality and weak stochastic transitivity axioms in multiattribute choice. This type of preference correlation is also necessary to account for similarity-based moderators of context effects and choice deferral effects (Bhatia, 2013; Bhatia & Mullett, 2016; Diederich, 1997, 2003; Roe et al., 2001). In related domains, such as risky choice, preference correlations explain observed violations of stochastic dominance, and the sensitivity of choice to the covariance in the payoffs of the available gambles (Andraszewicz, Rieskamp, & Scheibehenne, 2015; Diederich & Busemeyer, 1999; Busemeyer & Townsend, 1993).

Relatedly, preference change implies dynamic dependencies in choice that are necessary to explain why harder decisions take longer. Such dependencies also allow preference accumulation models to account for deliberation time patterns involving context effects, attention effects, and choice deferral effects (Bhatia, 2013; Bhatia & Mullett, 2016; Krajbich et al., 2010; Roe et al., 2001; Trueblood et al., 2014; Tsetsos et al., 2012; Usher & McClelland, 2004; Zeigenfuse, Pleskac, & Liu, 2014). More generally, preference change helps explain the systematic relationship between deliberation time and choice probability, such that with increasing time choice probabilities become more extreme (see also Dai, Pleskac, & Pachur, 2018). This relationship between deliberation time and choice probability is present not only in preferential choice, but also in non-evaluative, low-level, (typically perceptual) decision domains (e.g. Bogacz et al., 2006; Gold & Shadlen, 2007; Pleskac & Busemeyer, 2010; Ratcliff & Rouder, 1998; Usher & McClelland, 2001).

Finally, preference competition allows for preferences to display a sensitivity to the other objects in the choice set. This can generate a type of context dependence, and subsequently account for the influence of irrelevant decoys on multiattribute choice (Roe et al., 2001; Usher & McClelland, 2004). Competition can also explain why choice deferral gets more likely as additional objects are added to the choice set (Bhatia & Mullett, 2016). Additionally, in low-level decision domains, competitive accumulation has been shown to explain a number of key behavioral patterns involving decision difficulty, accuracy, and deliberation time (e.g. Teodorescu & Usher, 2013; Usher & McClelland, 2001).

All three of the above properties can also interact, and a number of models use all three properties to account for observed findings in multiattribute choice. These models include multialternative decision field theory (Berkowitsch, Scheibehenne & Rieskamp, 2014; Roe et al., 2001), the loss averse leaky competitive accumulation model (Usher & McClelland, 2004), and the associative accumulation model (Bhatia, 2013; Bhatia & Mullett, 2016). More recently, quantitative model testing by Turner et al. (2018) has provided strong evidence in favor of the core assumptions of the above models, and thus evidence in favor of preference correlation, change, and competition in multiattribute decision making.

Although the three properties discussed here have been examined primarily in discrete choice tasks, they also make predictions in continuous desirability ratings tasks. These predictions pertain to the statistical distributions of desirability ratings, their dependence on deliberation time, and their dependence on the other objects being evaluated by the decision maker. If these stochastic, dynamic, and context-dependent predictions do hold, then it would appear that theories of preference accumulation can be successfully extended beyond discrete choice to model continuous responses. This would, in turn, indicate that the key assumptions of these theories—the stochastic sampling of attributes and the dynamic and competitive accumulation of attribute values into preference—provide an accurate characterization of the core cognitive processes responsible for the construction of preference and the evaluation of multiattribute objects.

2. Model

The task we wish to model involves subjective ratings for multiattribute objects. We denote a multiattribute object i as a vector of m attributes, $\boldsymbol{x_i} = (x_{i1}, x_{i2}, \ldots, x_{im})$, with x_{ij} corresponding to the amount of attribute j in the object. Individuals are asked to evaluate the attributes of the objects in consideration, and assign to each object i a continuous real-valued desirability rating representing their

preference for the object. This is similar to standard choice tasks in which individuals are asked to evaluate a set of objects and select the object they find to be the most preferable. Choice tasks are commonly modelled using preference accumulators, but they can easily be adopted to model desirability ratings.

To specify the model, let us assume that the individual is asked to evaluate n objects, $\boldsymbol{x_1}$, $\boldsymbol{x_2}$, …, $\boldsymbol{x_n}$, and thus accumulates preferences for each of these objects simultaneously. These preferences are updated based on the stochastic sampling of attributes. We assume that at each point in time, the decision maker randomly attends to the m underlying attributes. The attention weights for the attributes (which could reflect attribute importance) are described by the m-dimensional vector $\boldsymbol{w}(t)$. We assume the attention weights are distributed according to a stationary multivariate normal $\boldsymbol{w}(t) \sim N(\boldsymbol{\mu_w}, \boldsymbol{\Sigma_w})$. Here $\boldsymbol{\mu_w} = \mathrm{E}[\boldsymbol{w}(t)]$ is the m-dimensional mean weight vector for the m attributes, and $\boldsymbol{\Sigma_w}$ is an $m \times m$ covariance matrix. As this paper considers settings with desirable attributes we restrict $\boldsymbol{\mu_w} > \mathbf{0}$. For convenience, we assume that weights are not correlated so that $\boldsymbol{\Sigma_w}$ is a diagonal matrix with constant variance ${\sigma_w}^2$ for all attributes.

The decision maker, at each time t, calculates the subjective values of the n objects, based on attention weights, $\boldsymbol{w}(t)$. We assume that these subjective values are further perturbed by additive errors. We write the errors at a given time t using the n-dimensional vector $\boldsymbol{\varepsilon}(t)$, which we assume is distributed according to a stationary multivariate normal $\boldsymbol{\varepsilon}(t) \sim N(\boldsymbol{\mu_\varepsilon}, \boldsymbol{\Sigma_\varepsilon})$. The errors have a mean of zero, with $\boldsymbol{\mu_\varepsilon} = \mathrm{E}[\boldsymbol{\varepsilon}(t)] = \mathbf{0}$. They are also uncorrelated across objects, so that $\boldsymbol{\Sigma_w}$ is a diagonal matrix with a constant ${\sigma_\varepsilon}^2$ on the diagonal. We can write the n objects being evaluated in an $n \times m$ matrix $\mathbf{X} = [\boldsymbol{x_1}; \boldsymbol{x_2}; \ldots; \boldsymbol{x_n}]$, with row i corresponding to object i. Thus, the subjective values, $\boldsymbol{v}(t)$, at time t can be written as: $\boldsymbol{v}(t) = \mathbf{X}\cdot\boldsymbol{w}(t) + \boldsymbol{\varepsilon}(t)$.

Thus far we have outlined the sequential attribute sampling processes that determine moment-by-moment evaluations of each object's desirability. The preference accumulation approach assumes that these valuations are aggregated into preferences that change dynamically over time. We write the preferences for the objects at time t using the n-dimensional vector $\boldsymbol{p}(t)$, so that the preferences for the objects are given by: $\boldsymbol{p}(t) = \mathbf{S}\cdot\boldsymbol{p}(t\text{-}1) + \boldsymbol{v}(t)$. Here $\mathbf{S}$ is an $n \times n$ feedback matrix that describes the influence of $\boldsymbol{p}(t\text{-}1)$ on $\boldsymbol{p}(t)$. We assume that preferences decay over time (at an identical rate for all objects). To do so, the diagonal entries of $\mathbf{S}$ feature a decay constant, δ, with $0 \leq \delta \leq 1$. When $\delta = 0$ preference does not accumulate and reflects the current sample, when $\delta = 1$ there is perfect accumulation over time, and intermediate values of δ result in intermediate decay in the accumulated preference (for this reason δ is perhaps better thought of as a self-feedback parameter). Additionally, the preferences for the objects are allowed to inhibit each other. This inhibition is done by setting the off-diagonal entries of $\mathbf{S}$ to a lateral inhibition constant, $-\lambda$, with $0 \leq \lambda \leq 1$. When $\lambda = 0$ there is no lateral inhibition, when $\lambda = 1$ there is perfect lateral inhibition such that an increment in preference on one accumulator decrements preference on the other accumulator by an equal amount, and intermediate levels of λ result in intermediate levels of inhibition. Note that as the attention weights correspond to attribute importance, we can use the expected weight vector, multiplied by the attributes of the objects, as a proxy for the overall desirability of each object. This corresponds to the expected accumulation rate for the objects, $\mathrm{E}[\boldsymbol{v}(t)] = \mathbf{X}\cdot\mathrm{E}[\boldsymbol{w}] = \mathbf{X}\cdot\boldsymbol{\mu_w}$.

Fig. 1 illustrates the general structure of the model with a connectionist framework where the decision maker is evaluating two objects each with three attributes. This framework has typically been applied to predict choices. In fact, the exposition here is very similar to that provided in Roe et al. (2001) (see Bhatia, 2013; Usher & McClelland, 2004 for variants; Turner et al., 2018, provide a cohesive overview of core assumption; also see Berkowitsch et al., 2014 for a useful summary). Note our claim that attributes are sampled according to a multivariate normal weight vector is slightly different to the Bernoulli sampling assumption sometimes used to predict choice. This normality assumption simplifies the sampling process so as to have simple analytical solutions for the distribution of preferences at each time point (as we fit our model to ratings data obtained at different time points). A Bernoulli sampling assumption does yield a multivariate normal distribution for preferences, but does so only in limit. In contrast, our assumption yields a normal distribution for preferences even when $t = 3, 5, 7, 9, 11$ etc. (as in our experiments). Crucially, the Bernoulli and normal sampling assumptions generate similar types of preference correlations, and thus make similar predictions in our experimental task. In contrast, models without stochastic attribute sampling (i.e. models that do not permit variability in $\boldsymbol{w}(t)$, and thus yield a single deterministic weight vector at each time period – e.g. Glöckner et al., 2014; Trueblood et al., 2014) would not, without additional assumptions, generate similarity-based correlations in preference states.

A much more important difference between prior work and our paper is that we wish to use the model to predict continuous desirability ratings rather than discrete choices. Thus, we do not assume that the object with the highest preference is chosen at a given decision based on either optional stopping or interrogation. Rather, in line with the assumptions of Pleskac and Busemeyer (2010) (also see Busemeyer & Diederich, 2002, p. 368; Yu, Pleskac, & Zeigenfuse, 2015), we assume that desirability ratings for the objects, at the time of rating T, are determined by mapping $\boldsymbol{p}(T)$ directly onto the response scale. Objects with larger values of $p_i(T)$ are given higher desirability ratings. For simplicity, we use an identity function such that the accumulated preference maps directly onto the observed ratings. This procedure works well for the continuous rating scale participants used in our experiments, but other mapping functions are possible such as with criteria that bin the preference ratings into discrete categories (see for example Pleskac & Busemeyer, 2010).

In the experiments below, T is determined by an experimentally imposed deliberation time limit. Thus, our paper restricts the model to settings in which the time of the desirability rating is exogenous to the evaluation process (analogous to the interrogation procedure used to model discrete choice). Models with an optional stopping mechanism, where it is up to the individual when to select a rating, are also feasible (e.g., Kvam, 2018; Ratcliff, in press; Smith, 2016). One way to conceptualize how this can occur is to model preference accumulation as a Markov process where there is a discrete set of preference states and the accumulation process moves across these states. Each preference state can be mapped to a desirability rating, and when preference reaches each state there can be a small probability of stopping and selecting the rating that corresponds to the preference state. Such a process then predicts when people stop and what rating they give (for the formal solutions see Pleskac & Busemeyer, 2010). Regardless, with either

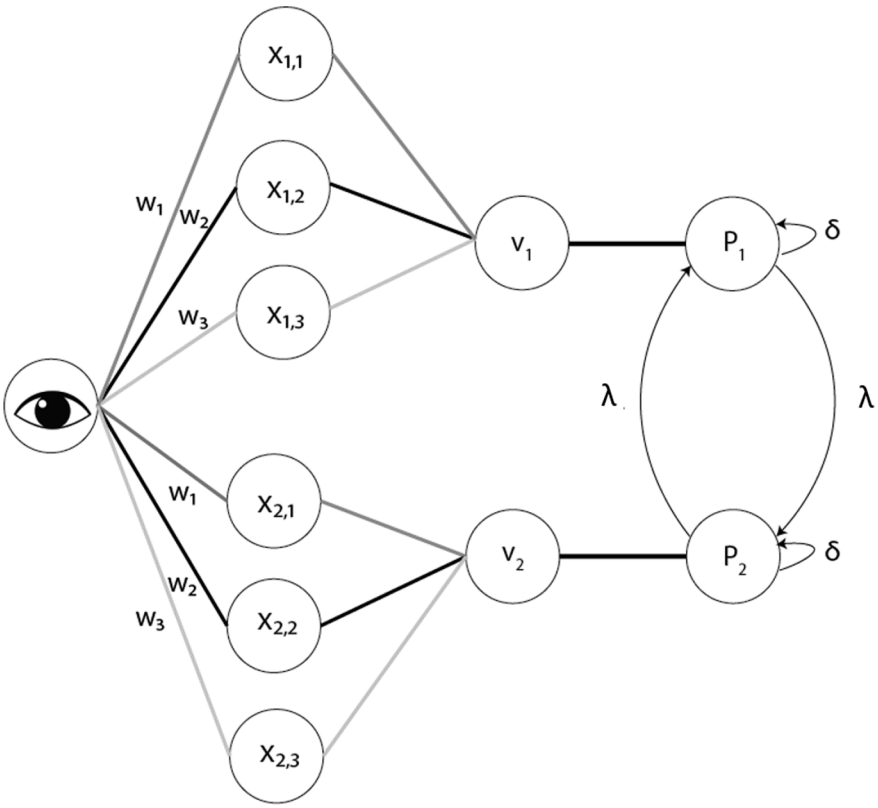


Fig. 1. A connectionist diagram of the preference accumulation model. According to the model at each time point t, the decision maker calculates a subjective value of each object based on the attentional weights allocated to each attribute. The weights themselves are random variables whose value changes at each time point. In the illustration above, at this particular time point, the second attribute has the highest attentional weight as denoted by the darker line. These valuations are aggregated into a preference, which can decay at rate δ. Additionally, the preferences for the objects are allowed to inhibit each other. The degree of inhibition is controlled by the parameter λ.

response procedure the predictions of preference accumulation for desirability ratings should be the same, and thus for simplicity our studies focus on an interrogation procedure.

Additionally, for computational tractability, we assume that the time scale of the model's attentional fluctuation and accumulation is in seconds (that is, one second corresponds to one iteration of the preference update equation), and thus the deliberation time limit is imposed after relatively few preference updates. Similarly, to facilitate analytical solutions for preference distributions we do not place upper or lower bounds on the preference states, and thus theoretically, our model can give any ratings (even outside the range of [0,1] permitted in the experiments). Finally, we assume that the evaluation process starts with an unbiased accumulator $p_i(0) = 0.5$ for all i.

In some cases, it may be necessary to make more complex assumptions about the time limit, the starting point, the time scale of the model's attentional fluctuation and accumulation, maximum or minimum feasible values for $\boldsymbol{p}(t)$, or the mapping of $\boldsymbol{p}(t)$ to the response scale. However, for the purposes of this paper the above assumptions are sufficient. This is because our model structure allows $\boldsymbol{p}(t)$ to be relatively flexible: Recall that the mean weight vector, $\boldsymbol{\mu_w}$ is unbounded, and thus the subjective values and preferences can, with the appropriate weights and model parameters, be scaled up or down, as necessary.

The vector of preferences $\boldsymbol{p}(t)$ is a stochastic and dynamic vector. As it is a linear function of $\boldsymbol{w}(t)$ and $\boldsymbol{\varepsilon}(t)$, which are distributed according to multivariate normal, $\boldsymbol{p}(t)$ is also distributed according to the multivariate normal. Thus, the distribution of $\boldsymbol{p}(t)$ at time t can be written as: $\boldsymbol{p}(t) \sim N(\boldsymbol{\mu}_{\boldsymbol{p}(t)}, \boldsymbol{\Sigma}_{\boldsymbol{p}(t)})$, with $\boldsymbol{\mu}_{\boldsymbol{p}(t)}$ corresponding to the expected preference vector at time t, and $\boldsymbol{\Sigma}_{\boldsymbol{p}(t)}$ corresponding to covariance matrix for preferences at time t. We assume that desirability ratings for the objects, at the time of rating T, are determined by mapping $\boldsymbol{p}(T)$ directly onto the response scale. Thus, the predicted distribution of desirability ratings is given by $\boldsymbol{p}(T) \sim N(\boldsymbol{\mu}_{\boldsymbol{p}(T)}, \boldsymbol{\Sigma}_{\boldsymbol{p}(T)})$. This distribution is a function of the attribute values of the objects ($\mathbf{X}$), the deliberation time (T), and the parameters of the model (the decision weights, $\boldsymbol{\mu_w}$; decay, δ; inhibition, λ; and noise terms σ_w and σ_ε). The derivation for the values of $\boldsymbol{\mu}_{\boldsymbol{p}(t)}$ and $\boldsymbol{\Sigma}_{\boldsymbol{p}(t)}$ at time t is provided in Appendix B of Roe et al. (2001) (though note that unlike Roe et al., we are not allowing for a contrast matrix, and that our feedback matrix does not feature distant dependent lateral inhibition). Fig. 2 provides a realization of the accumulated preferences that can occur as a decision maker evaluates two objects.

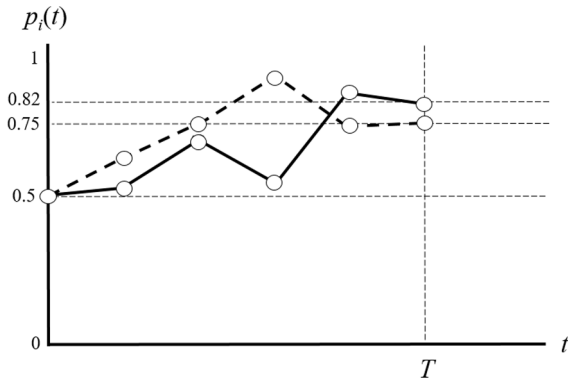


Fig. 2. Illustration of a hypothetical decision. The decision maker is assumed to accumulate preferences, $p_i(t)$, for the choice objects in consideration, and respond with a continuous rating corresponding to the accumulated preferences at the deliberation time limit, T. In this example the object corresponding to the solid line is given a rating of 0.82 whereas the object corresponding to the dashed line is given a rating of 0.75.

3. Properties and predictions

The m mean attribute weights that make up $\boldsymbol{\mu_w}$, and the standard deviation σ_ε for the additive error, are parameters that are present in nearly all models of multiattribute preference (e.g. Keeney & Raiffa, 1993). It is the remaining three parameters, the standard deviation σ_w for the attentional weight vector, the decay term, δ, and the lateral inhibition term, λ, that are unique to the accumulator approach, and that give this model its stochastic, dynamic, and context-dependent properties.

The first of these properties, which we label *preference correlation*, is the result of the variability in the weights $\sigma_w > 0$. In this model, not only do we have stochasticity generated by an independent additive error term capturing noise in the valuations of each option, but we also have stochasticity generated by fluctuating weights in terms of what attributes are being sampled or attended to. This latter type of stochasticity leads to correlations between the valuations, and subsequently preferences, of the choice objects, that depend crucially on the overlap in attributes across objects. Ultimately objects that have similar attributes are also likely to have similar valuations and preferences. More formally, whenever $\sigma_w > 0$, changing the attributes of the objects, even when keeping their overall desirabilities ($\mathrm{E}[\boldsymbol{v}(t)] = \mathbf{X}\cdot\mathrm{E}[\boldsymbol{w}] = \mathbf{X}\cdot\boldsymbol{\mu_w}$) constant, changes in the off-diagonal entries in the preference covariance matrix $\boldsymbol{\Sigma}_{\boldsymbol{p}(t)}$. If $\sigma_w = 0$ and the attribute weights are fixed, the preference covariance matrix would be unaffected by changes in the attributes of the objects (keeping their desirabilities constant). Note that positive inhibition, $\lambda > 0$, can lead to preference correlations, in the form of non-zero off-diagonal entries in the preference covariance matrix $\boldsymbol{\Sigma}_{\boldsymbol{p}(t)}$, however, these entries would only be affected by object desirability, and not by object similarity.

The second property, which we call *preference change*, emerges due to the decay term $\delta > 0$. This property works as follows. When $\delta > 0$ preferences add up over time and thus the distribution of $\boldsymbol{p}(t)$ varies with t. If an object i is highly desirable, with a large $\mathrm{E}[v_i(t)] = \boldsymbol{x_i}\cdot\mathrm{E}[\boldsymbol{w}] = \boldsymbol{x_i}\cdot\boldsymbol{\mu_w}$, the expectation of $p_i(t)$ is likely to increase over time. Overall, more desirable objects see quicker increases (or slower decreases) in their preferences over time, relative to less desirable objects. Note that positive lateral inhibition ($\lambda > 0$) can lead to changes in preferences even if there is no decay ($\delta = 0$), though such a model would not easily be able to account for high desirability ratings, and desirability ratings that increase over time. Thus, intuitively, preference change can be understood as largely due to decay ($\delta > 0$), and if there is no decay ($\delta = 0$) our model would fail to display reasonable changes in preference over time. Formally, if both $\delta = 0$ and $\lambda = 0$, the expected value of $\boldsymbol{p}(t)$ would remain perfectly constant over time, i.e. $\mathrm{E}[\boldsymbol{p}(t)] = \mathrm{E}[\boldsymbol{v}(t)] = \mathbf{X}\cdot\mathrm{E}[\boldsymbol{w}] = \mathbf{X}\cdot\boldsymbol{\mu_w}$.

The final property, which we label *preference competition*, is exclusively a product of the lateral inhibition term $\lambda > 0$. Not only do preferences accumulate over time, but the preferences also interact. Thus, for an object i, the expected preference, $\mathrm{E}[p_i(t)]$, depends not only on its own desirability, $\mathrm{E}[v_i(t)]$, but also on the desirability of its competitor, $\mathrm{E}[v_j(t)]$. One implication of this is that highly desirable objects have a strong inhibitory effect on the preferences of other objects, and thus dampen these preferences. In contrast, the inhibitory effect of less desirable objects is much weaker. Ultimately the preferences for a given object are likely to be higher when it is evaluated alongside an undesirable object compared to when it is evaluated alongside a highly desirable object. Note that stochastic attention weights, $\sigma_w > 0$, can lead to correlations between preferences (i.e. non-zero off-diagonal entries in the preference covariance matrix $\boldsymbol{\Sigma}_{\boldsymbol{p}(t)}$), but cannot, by themselves, lead to changes in the expected preferences, $\mathrm{E}[\boldsymbol{p}(t)]$.

In Figs. 3–5 we illustrate the above three effects. Each figure is based on a situation that involves rating the desirability of two objects, each defined on three binary-valued attributes. The attributes are assumed to have equal weights, with $\boldsymbol{\mu_w} = [0.1,0.1,0.1]$, and the process is assumed to display additive error, with $\sigma_\varepsilon = 0.05$. Unless otherwise specified, we also assume $T = 5$, $\delta = 0.6$, $\lambda = 0.01$, and $\sigma_w = 0.01$. These parameter values approximate average participant parameter values recovered in our model fits to Experiment 2 data (see Table 2 below).

We begin with the preference correlation property, for which we consider objects $\boldsymbol{x_1} = [1,1,0]$, $\boldsymbol{x_2} = [1,0,0]$, and $\boldsymbol{x_3} = [0,0,1]$. Note that $\boldsymbol{x_2}$ and $\boldsymbol{x_3}$ have identical desirabilities, but differ in terms of their attribute overlap with $\boldsymbol{x_1}$. Thus, whenever there is stochasticity in the attribute weights $\sigma_w > 0$, we should expect to observe higher preference correlations in ratings tasks involving $\boldsymbol{x_1}$ and $\boldsymbol{x_2}$, compared to ratings tasks involving $\boldsymbol{x_1}$ and $\boldsymbol{x_3}$. This property is shown in Fig. 3, which plots the difference in the covariance of

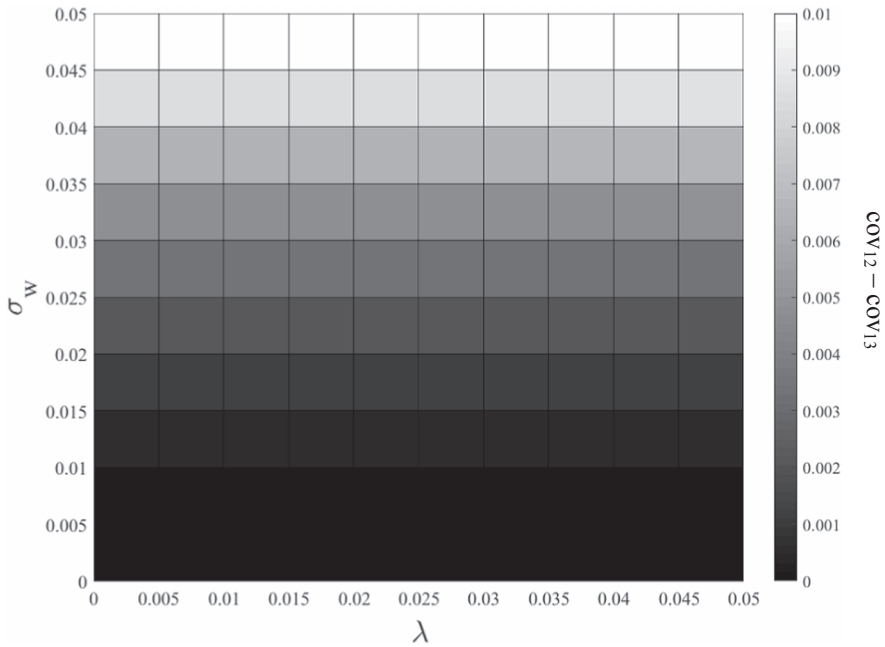


Fig. 3. Illustration of the preference correlation property, and its dependence on parameters σ_w and λ. Here we display the differences in the covariance of the preferences for $\boldsymbol{x_1}$ = [1,1,0] and $\boldsymbol{x_2}$ = [1,0,0], compared to preferences for $\boldsymbol{x_1}$ = [1,1,0] and $\boldsymbol{x_3}$ = [0,0,1].

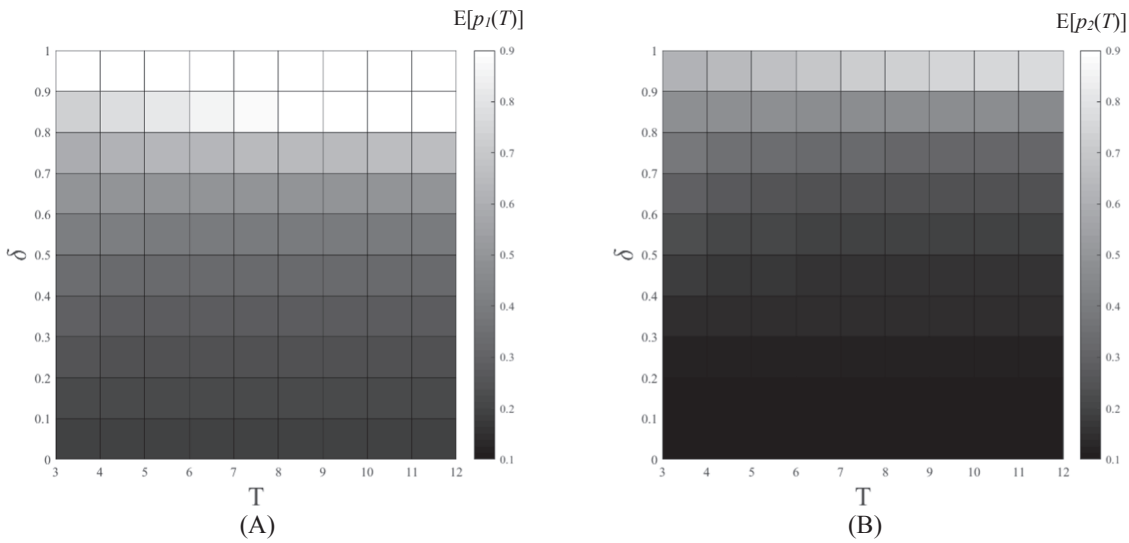


Fig. 4. Illustration of the preference change property, and its dependence on parameters δ and T. Here we display the expected preferences for $\boldsymbol{x_1}$ = [1,1,0] (A) and $\boldsymbol{x_2}$ = [0,0,1] (B).

the preferences of $\boldsymbol{x_1}$ and $\boldsymbol{x_2}$ relative to $\boldsymbol{x_1}$ and $\boldsymbol{x_3}$, i.e. $cov_{12} - cov_{13}$, where cov_{12} is the off-diagonal entry of $\boldsymbol{\Sigma_{p(t)}}$ with $\boldsymbol{x_1}$ and $\boldsymbol{x_2}$, and cov_{13} is the off-diagonal entry of $\boldsymbol{\Sigma_{p(t)}}$ with $\boldsymbol{x_1}$ and $\boldsymbol{x_3}$. We plot $cov_{12} - cov_{13}$ with σ_w in the set {0, 0.001, 0.01, 0.015, 0.02, 0.025, 0.03, 0.035, 0.04, 0.045, 0.05}. As λ can also influence covariance, we also consider λ in the set {0, 0.005, 0.01, 0.015, 0.02, 0.025, 0.03, 0.035, 0.04, 0.045, 0.05}. We keep all other parameters fixed at the values specified above. Darker shades correspond to lower values of $cov_{12} - cov_{13}$, with the darkest shade indicating $cov_{12} - cov_{13} = 0$.

There are a number of key patterns that can be seen in Fig. 3. First, $cov_{12} - cov_{13} = 0$ when $\sigma_w = 0$, but $cov_{12} - cov_{13} > 0$ when $\sigma_w > 0$. Additionally, increasing stochasticity in the weights σ_w increases $cov_{12} - cov_{13}$. Thus we get differences in preference covariance only in the presence of stochastic attribute weights, with higher stochasticity in attribute weights (higher σ_w) leading to higher differences in covariance. Finally, these properties are largely insensitive to the inhibition λ. Although inhibition (λ > 0) can lead to preference covariance (in the form of non-zero values of cov_{12} and cov_{13}), it does not influence differences in covariance ($cov_{12} - cov_{13}$) keeping desirabilities fixed, as is the case here. These patterns show that the preference correlation property discussed

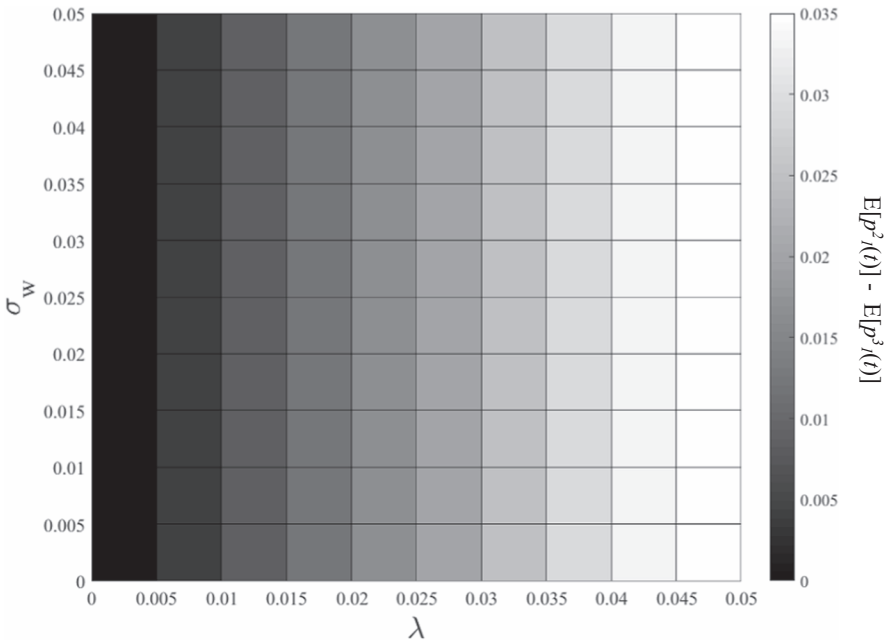


Fig. 5. Illustration of the preference competition property, and its dependence on parameters σ_w and λ. Here we display the difference in the expected preference of $\boldsymbol{x_1}$ = [1,1,0] when evaluated alongside $\boldsymbol{x_2}$ = [0,0,1], relative to the expected preference of $\boldsymbol{x_1}$ when evaluated alongside $\boldsymbol{x_3}$ = [1,1,1].

above depends fundamentally on stochasticity in attribute sampling, that is, on the parameter σ_w.

To examine the preference change property, we use objects $\boldsymbol{x_1}$ = [1,1,0] and $\boldsymbol{x_2}$ = [0,0,1]. Note that $\boldsymbol{x_1}$ is more desirable than $\boldsymbol{x_2}$. Thus, whenever there is decay ($\delta > 0$) we should expect to observe a quicker increase (or slower decrease) in preferences for $\boldsymbol{x_1}$ than $\boldsymbol{x_2}$. Fig. 4A and 4B demonstrate this property, with values of δ in the range {0, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1.0} and values of T in the range {3, 4, 5, 6, 7, 8, 9, 10, 11, 12}. All other parameters are set to the values defined above. Fig. 4A plots the expected preference for $\boldsymbol{x_1}$, E[$p_1(t)$], and Fig. 4B plots the expected preference for $\boldsymbol{x_2}$, E[$p_2(t)$]. Again, darker shades indicate lower preferences.

As expected, we observe higher preferences (corresponding to lighter shades) for Fig. 4A relative to 4B. Thus, for a given level of decay δ and time T, preferences for the more desirable object are, on average, higher than preferences for the less desirable object. More interestingly, we note that preferences are changing systematically with time T, with preferences for $\boldsymbol{x_1}$ increasing at a quicker rate than preferences for $\boldsymbol{x_2}$. When there is no accumulation due to perfect decay ($\delta = 0$), there is no observable shift in preference with varying T, showcasing the dependence of the preference change property on δ (technically, small values of λ, as is the case here, can alter preferences even without accumulation, but the effect is too mild to be visible in these plots). Overall, the highest preferences, in both figures, are observed when δ and T are at set at their maximum values. These patterns show that the preference change property depends fundamentally on the accumulation of preferences, that is, on the parameter δ. Another illustration of the preference change property generated by our model can be found in Usher and McClelland (2001, Fig. 1).

In Fig. 5 we examine the preference competition property. Here we consider $\boldsymbol{x_1}$ = [1,1,0], $\boldsymbol{x_2}$ = [0,0,1], and $\boldsymbol{x_3}$ = [1,1,1]. In this setting, $\boldsymbol{x_3}$ is more desirable than $\boldsymbol{x_2}$. Thus we should expect the preference for $\boldsymbol{x_1}$ to be higher in the presence of $\boldsymbol{x_2}$ compared to in the presence of $\boldsymbol{x_3}$. Fig. 5 demonstrates this property by displaying the difference in the expected preference for $\boldsymbol{x_1}$ when evaluated alongside $\boldsymbol{x_2}$ relative to the expected preference for $\boldsymbol{x_1}$ when evaluated alongside $\boldsymbol{x_3}$, that is E[$p_1^2(t)$] - E[$p_1^3(t)$] (with the superscript indicating the competitor object). We consider values of λ in the set {0, 0.005, 0.01, 0.015, 0.02, 0.025, 0.03, 0.035, 0.04, 0.045, 0.05}. To contrast the effect of λ against the effect of σ_w, we also consider σ_w in the set {0, 0.005, 0.01, 0.015, 0.02, 0.025, 0.03, 0.035, 0.04, 0.045, 0.05}. Other parameters are fixed at the values defined above. Again darker shades indicate smaller differences, with the darkest shade corresponding to E[$p_1^2(t)$] - E[$p_1^3(t)$] = 0.

Fig. 5 shows a number of important patterns. First, E[$p_1^2(t)$] - E[$p_1^3(t)$] = 0 when λ = 0, but E[$p_1^2(t)$] - E[$p_1^3(t)$] > 0 when λ > 0. Additionally, increasing λ increases E[$p_1^2(t)$] - E[$p_1^3(t)$]. Thus we get differences in preference only in the presence of inhibition, with higher inhibition leading to higher differences in preference. Finally, these properties are completely insensitive to the stochasticity in the attention weights σ_w. Although σ_w can lead to preference correlation, it has no effect on mean preference. These patterns show that the preference competition property discussed above depends fundamentally on lateral inhibition, that is, on the parameter λ.

In this paper, we present the results of two experiments designed to test for the preference correlation, preference change, and preference competition properties in desirability ratings tasks. These experiments present participants with two multiattribute objects, and elicit desirability ratings for one or both of the objects after an experimentally imposed deliberation time limit. If the model of multiattribute preference accumulation outlined above accurately describes the evaluative processes responsible for desirability ratings, then we expect the following predictions to hold: First, in accordance with the preference correlation property, there should

be a higher correlation between the ratings of similar objects than between the ratings of dissimilar objects, controlling for the desirabilities of the objects. Second, in accordance with the preference change property, ratings should systematically change with deliberation time, with ratings for desirable objects more likely to increase with longer deliberation times than ratings for undesirable objects. Third and finally, in accord with the preference competition property, when objects are presented alongside desirable objects their ratings should be suppressed. Similarly, when objects are presented alongside undesirable objects their ratings should be boosted. To our knowledge, these three properties have not been systematically examined in desirability ratings task, though some prior work does provide preliminary evidence for preference competition (Brenner, Rottenstreich & Sood, 1999).

4. Model fitting

In addition to testing for these stochastic, dynamic, and context-dependent properties using qualitative statistical methods, we will fit the model outlined above to participant ratings. This model permits non-zero values of the three main parameters: the standard deviation σ_w for the attentional weight vector, the decay term, δ, and the lateral inhibition term, λ (as well as non-zero values for m mean attribute weights that make up $\boldsymbol{\mu_w}$, and the standard deviation σ_ε for the additive error), and thus will be referred to as the *full* model. Due to the direct relationship between the three properties and the three unique parameters of the model we will also consider nested variants of this model that individually restrict the parameters and provide a model-based approach to test for the presence of these properties. Thus, in addition to our full model, we will examine a constrained model which restricts $\sigma_w = 0$ (but permits flexible values for the remaining parameters), a second constrained model which restricts $\delta = 0$ (but permits flexible values for the remaining parameters), and a third constrained model which restricts $\lambda = 0$ (but permits flexible values for the remaining parameters). We refer to these models via the parameter they restrict: i.e. σ_w-*constrained* model, δ-*constrained* model, and λ-*constrained* model respectively. Comparisons of the full model against the three constrained models test whether each of the three parameters are necessary for good model fits, and thus whether the cognitive mechanisms corresponding to these parameters are at play in the ratings task. For example, if the full model outperforms the σ_w-constrained model, we can conclude that $\sigma_w > 0$ is necessary to adequately fit data, and that stochastic attribute sampling plays an important role in multiattribute desirability ratings.

Finally, we will also fit a model that has no stochasticity in the attention weights, no decay, and no inhibition ($\sigma_w = \lambda = \delta = 0$), but permits flexible values for the remaining parameters. This model features none of the fundamental properties of the accumulator approach and is identical to a simple linear multiattribute utility function combined with additive independent normally distributed noise. For this reason, it can be seen as a *baseline* model that is consistent with conventional models of preference (e.g. Keeney & Raiffa, 1993). Note that this model is also nested in the full model.

In the experiments below we fit our models on the participant level using maximum likelihood estimation applied to the analytical solutions for the distributions of desirability ratings. Thus, these fits attempt to find the parameters that maximize the likelihood of observing a participant's ratings data, given the predicted distribution of preferences at the time of rating T, $\boldsymbol{p}(T) \sim N(\boldsymbol{\mu}_{\boldsymbol{p}(T)}, \boldsymbol{\Sigma}_{\boldsymbol{p}(T)})$. Our fits are implemented in MATLAB using the simplex function. Although the analytical solutions for preference distributions makes it easy to find optimal parameters, the simplex routine can get stuck at local maxima. Thus, we repeat the fits 200 times for each participant, with a randomly determined starting point, and select the parameters with the overall highest likelihood. Note again, that we permit our models to form preference states outside the range [0,1] (which was is range of ratings permissible in our experiments). That said, when using best-fit model parameters to predict responses and evaluate the R^2 of fits, we truncate model predictions to 0 if $\mathrm{E}[p_i(T)] < 0$, or 1 if $\mathrm{E}[p_i(T)] > 1$.

5. Experiment 1

5.1. Overview

Our first experiment offered participants pairs of multiattribute objects. After an experimentally controlled deliberation time, participants were asked to rate one of the two objects. The attributes of the objects and the deliberation time limit were manipulated across trials, thereby permitting tests of the preference change and preference competition properties of the model. As only one object was rated in each trial, this design was not able to test the preference correlation property of the model.

5.2. Methods

5.2.1. Participants

A total of 91 participants performed this experiment in a behavioral laboratory, for course credit. Participants were members of a University of Pennsylvania experimental subjects pool, and were asked to rate the desirability of apartments around campus.

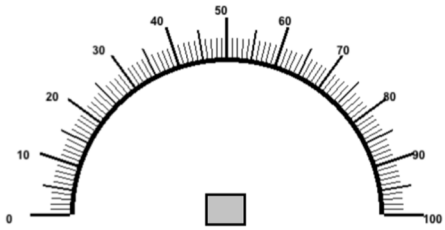
5.2.2. Materials and procedures

In each trial, participants were shown two apartments, which differed on three attributes: proximity to campus, size, and affordability. The attributes were binary valued, so that the apartments could either be proximate or not proximate, large or not large, and affordable or not affordable. An example of the stimuli presentation is provided in Fig. 6A. In this example both Apartment X and Apartment Y are close to campus, but Apartment X is large in size, whereas Apartment Y is affordable.

After being shown two available apartments, participants were taken to a second screen which asked them to rate the desirability of one of the two apartments (the target). The information about the attributes of the apartments was not shown to participants on the

	Option X	Option Y
Proximity	+	+
Size	+	
Affordability		+

(A)



(B)

Fig. 6. Example of experimental stimuli (A) and desirability rating interface (B) in Experiment 1. The presence of an attribute is indicated with a "+", and participants have to use a mouse click to rate one of the apartments after the two apartments have been displayed. The grey box in B indicates the starting position of the mouse in each trial. Note that it is equidistant to all the points on the rating scale.

ratings screen, and conversely information about which of the apartments was the target was not shown to participants on the apartment screen. Additionally, participants were explicitly asked to assess the absolute desirability of target apartment, rather than the desirability of the target relative to the other apartment (the competitor). Importantly we varied the length of time the apartments were shown on the screen prior to the participant being prompted to make the rating. These deliberation time limits were taken from the set {3 s, 5 s, 7 s, 9 s, 11 s}.

The ratings were made on a continuous scale ranging from 0 to 100, with 0 indicating extremely undesirable and 100 indicating extremely desirable (in the analysis below we rescale this to 0–1). The participants indicated their ratings with a computer mouse using the interface shown in Fig. 6B. Here the cursor of the mouse returned to its origin-position after every rating. Additionally, the scale was presented in a semi-circle so that each point on the scale was equidistant to the origin-position of the mouse.

There were 8 possible apartments for the three-attribute setting considered here. All possible combinations of pairs of apartments were presented to participants, leading to a total of 28 unique stimuli pairs (which is the number of entries above or below the main diagonal in the 8×8 matrix formed by comparing all 8 versions of the two apartments). Each of the 28 pairs were presented twice for each deliberation time condition, corresponding to whether the first or the second apartment was the target. As there were 5 deliberation time conditions, this led to a total of $28 \times 2 \times 5 = 280$ trials. The order of presentation for these 280 trials, as well as whether the target was presented on the left (as Apartment X) or on the right (as Apartment Y), was randomized.

We excluded all trials in which participants took more than 5 s to respond after being prompted. We also excluded 4 participants who did not finish the entire experiment.

5.3. Results

5.3.1. Qualitative tests

In our final data, we had a total of 22,694 observations across 87 participants. In each of the observations, the participant rated the target, after being shown the target alongside a competitor for a fixed deliberation time (ranging from 3 to 11 s). The mean rating across all participants was 0.49, with a maximum rating of 1.00, a minimum rating of 0.00, and a median rating of 0.50 (note that here, as in the model fits, we have rescaled the ratings to be in the range [0,1]).

We first qualitatively examined whether the preference change and preference competition properties of the model were obtained in the data. Again, the preference change property involves the role of deliberation time, with the prediction that increases in deliberation time should be associated with larger increases in the ratings for desirable targets compared to the ratings for undesirable targets. Likewise, the preference competition property involves the role of the competitor. Despite participants being asked to rate the absolute desirability of the target, the model outlined above predicts that highly desirable (undesirable) competitors should decrease (increase) the ratings for the target.

For the purposes of this preliminary test we used the sum of the attributes in a given apartment as a proxy for the overall desirability of the apartment. As there are a total of three binary valued attributes, the target and the competitor can each take on desirability values of 0, 1, 2 or 3. We first examined the relationship between the participant rating and the desirability (i.e. sum of all attributes) of the target. Unsurprisingly, these two variables were highly correlated ($\rho = 0.72$, $p < 0.001$). A linear regression with ratings as the dependent variable and the desirability of the target as the independent variable, as well as participant-level random intercepts, found a strong significant relationship ($\beta = 0.26$, $z = 175.85$, $p < 0.001$). We ran a variant of this regression including the desirability (i.e. sum of all attributes) of the competitor as an additional independent variable and noted a significant negative effect of this variable on ratings of the target ($\beta = -0.03$, $z = -19.36$, $p < 0.001$). Thus, not only were desirable targets likely to have high ratings, as would be expected, but so were targets with undesirable competitors. Likewise, both undesirable targets and targets with desirable competitors were likely to have low ratings. This provides evidence for the preference competition property of the proposed model. The relationship between ratings and the desirability of the target and the desirability of the competitor is illustrated in Fig. 7A and B respectively.

In order to test the property of preference change, we again used a linear regression with ratings as the dependent variable, the desirability of the target as the independent variable, and participant-level random intercepts. We also included a main effect for the

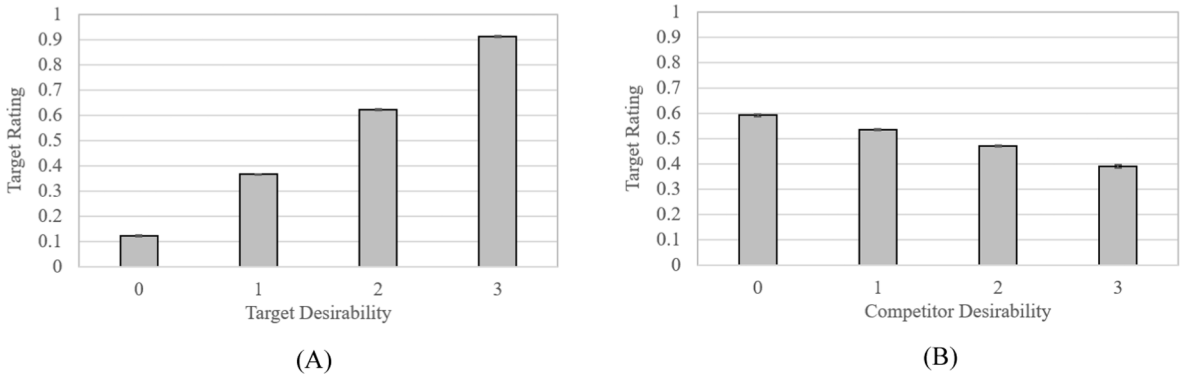


Fig. 7. Effect of target desirability (A) and competitor desirability (B) on target ratings in Experiment 1. The target ratings are averaged across pooled participant ratings for each level of desirability, and error bars indicate ± 1 SE.

deliberation time limit, as well as an interaction effect between deliberation time and target desirability. Our regression revealed a significant positive interaction effect ($\beta = 0.002$, $z = 3.43$, $p < 0.001$), supporting our prediction that ratings for desirable targets increase with deliberation time relative to ratings for undesirable targets.

We also tested for the effect of deliberation time on ratings separately for our four levels of desirability. Thus, a regression featuring ratings as the dependent variable and deliberation time as an independent variable (as well as participant-level random intercepts) for highly desirable targets with all three attributes (i.e. sum of attributes = 3) revealed a positive significant effect of deliberation time ($\beta = 0.003$, $z = 3.09$, $p < 0.01$). We observed a similar positive significant effect ($\beta = 0.002$, $z = 2.67$, $p < 0.01$) for desirable targets with two out of the three attributes. This effect, however, did not emerge for less desirable targets with only one of out of the three attributes ($\beta = -0.001$, $z = -1.21$, $p = 0.23$) or for highly undesirable targets with none of the three attributes ($\beta = -0.002$, $z = 1.38$, $p = 0.17$). The relationship between deliberation time and participant ratings, as a function of target desirability, is shown in Fig. 8A and B.

5.3.2. Quantitative tests

The qualitative tests showed that the preference change and preference competition properties of the proposed model are reflected in the data. Particularly we found that the length of time that the target is displayed (i.e. the deliberation time), prior to the ratings prompt, influenced the rating of the target. Additionally, ratings were not only affected by the desirability of the target, but also by the desirability of the competitor

In this section, we sought to more rigorously study these dynamic and context-dependent effects. Thus, we fit an accumulator model capable of predicting these effects, as well as nested variants of these models that are time and context independent. Note that in this experiment we obtained ratings for only one object per trial, and were thus unable to test the preference correlation property. For this reason, the standard deviation σ_w for the attentional weight vector does not influence the model's predictions, and we set $\sigma_w = 0$ for all models in this section.

Again, the specific models we consider in this section are the full model, which allows for flexible $\boldsymbol{\mu_w}$, δ, λ, and σ_ε; the λ-constrained model that allows for flexible $\boldsymbol{\mu_w}$, δ, and σ_ε, but sets $\lambda = 0$ so there is no inhibition; the δ-constrained model that allows for flexible $\boldsymbol{\mu_w}$, λ, and σ_ε, but sets $\delta = 0$ so that there is no accumulation (perfect decay); and a baseline model that allows for flexible $\boldsymbol{\mu_w}$ and σ_ε, but sets $\delta = \lambda = 0$. The full model features both preference change and preference competition, and thus is capable of

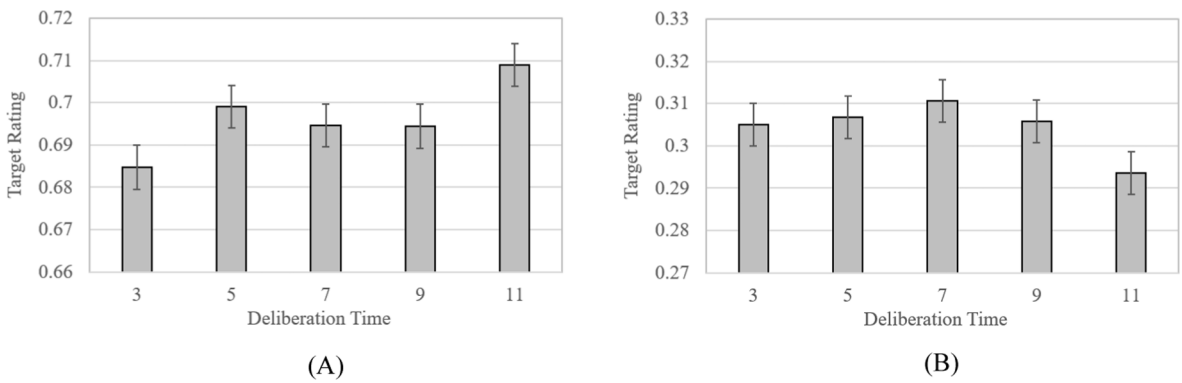


Fig. 8. Effect of deliberation time (in seconds) on target ratings for desirable targets with a desirability of 2 or 3 (A) and undesirable targets with a disability of 0 or 1 (B), in Experiment 1. The target ratings are averaged across pooled participant ratings for each deliberation time condition, and error bars indicate +/- 1 SE.

Table 1
Average best-fitting parameter values, log-likelihood values, and R^2 values, for fits in Experiment 1 (brackets indicate standard errors). "G^2/DF" corresponds to the test statistic and degrees of freedom for the likelihood ratio test comparison against the full model, on the group level. This statistic is significantly different to 0 for all models ($p < 0.001$). "% Better Fit by Full" indicates the proportion of participants significantly ($p < 0.05$) better fit by the full model relative to the nested model according to the likelihood ratio test on the individual level. Note the G^2 for each model can be found by multiplying the difference between the log-likelihood of the full model and the more constrained model for each participant by 2 and then summing across the participants.

	Full	λ-constrained	δ-constrained	Baseline
δ	0.44 (0.03)	0.43 (0.03)	[0]	[0]
λ	0.03 (0.01)	[0]	0.03 (0.01)	[0]
σ_ε	0.09 (0.01)	0.09 (0.01)	0.17 (0.01)	0.17 (0.01)
μ_{w1}	0.18 (0.01)	0.18 (0.01)	0.33 (0.01)	0.32 (0.01)
μ_{w2}	0.15 (0.01)	0.14 (0.01)	0.26 (0.01)	0.26 (0.01)
μ_{w3}	0.22 (0.01)	0.21 (0.01)	0.38 (0.01)	0.37 (0.01)
Log-likelihood	124.17 (11.32)	120.98 (11.55)	113.52 (11.94)	110.37 (12.13)
R^2	0.65 (0.03)	0.64 (0.03)	0.56 (0.05)	0.55 (0.05)
G^2/DF	–	554.70/87	1852.10/87	2400.73/174
% Better Fit by Full	–	21%	48%	61%

accommodating both qualitative findings discussed in the prior section. The λ-constrained model, in contrast, without inhibition only allows preference change, and cannot account for the dependence of ratings on the desirability of the competitor. Likewise, the δ-constrained model only allows preference competition, and cannot account for the increase in ratings over time for desirable targets (though it can, with appropriate values of λ predict a drop in these ratings for undesirable targets). Finally, the baseline model does not feature either of these two properties.

A quantitative comparison of the constrained and baseline models against the full model would tell us whether the δ and λ parameters (and thus the accumulation and inhibition assumptions of the full model) are necessary to quantitatively account for the data. For example, bad fits of the full model relative to the λ-constrained model would indicate that the full model does not capture the observed preference competition patterns in an appropriate manner. This is also the case with the δ-constrained model and the preference change patterns. If the full model is outperformed by the baseline model, it would seem that none of the unique assumptions of the preference accumulation framework are necessary to quantitatively capture data. Conversely, good fits for the full model relative to the remaining models would constitute strong support for the unique assumptions of this model.

We fit each of these models on the participant level using the methods outlined earlier in this paper. The model fits are summarized in Table 1. This table displays parameter values and log-likelihood values for the models, averaged across participants, with standard errors in brackets. The table also displays the statistic, G^2 and degrees of freedom, for the likelihood ratio test (LRT) on the group level. This test uses the sum of the participant-level log-likelihood values to measure how many times more likely the data are under the full model relative to a nested model that sets one or more parameters to a fixed value. The degrees of freedom used in the group-level LRT correspond to the total additional number of parameters used to fit the data for all participants in the full vs. nested model (for example, when the nested model contains one less parameter per participant, the degree of freedom for the LRT is 87, which is the total number of participants for which we fit the models). This test can be seen as comparing a group-level full model that permits completely flexible parameters for each participant, against a group-level nested model that permits some flexible parameters for each participant but sets the remaining parameters to a constrained value ($\delta = 0$, $\lambda = 0$, or both). Table 1 also displays the proportion of participants significantly ($p < 0.05$) better fit by the full model relative to the remaining three models, according the LRT applied separately for each participant. Lastly, Table 1 displays average R^2 values across participants, for each of the models. These values are calculated by taking the model predictions for each participant's choice (based on the expected preferences generated by the best-fitting parameters for that participant) and comparing these predictions against the participant's actual choices.

Table 1 shows that the full model was able to describe the data well by achieving an average R^2 values of 0.65. This slightly outperformed the λ-constrained model and greatly outperformed the δ-constrained and baseline models. However, R^2 does not control for the increased flexibility of the full model, for which we need to use the likelihood ratio test (LRT). When applied to the group level, in the manner described in the prior paragraph, this test showed that the full model significantly ($p < 0.001$) outperformed the three remaining models. Thus, on aggregate, allowing for both flexible δ and flexible λ is necessary to describe behavior.

When the LRT was applied on the individual level, we also found that the full model significantly ($p < 0.05$) outperformed the baseline model for the majority of participants. The comparisons between the full model and the λ-constrained model and the δ-constrained models identified more individual level variability. For both models, the full model provided a better fit only for a substantial minority of participants particularly for the λ-constrained model. Yet, taken together with the fits at the aggregate level, we take this as supporting evidence that a flexible δ and flexible λ are necessary to adequately describe behavior (though not all participants necessarily have non-zero δ and λ).

It is also useful to compare the λ-constrained model and the δ-constrained against each other to test which of the two parameters is more important for describing behavior. These two models have an equivalent number of parameters can thus can be compared using only their log-likelihoods and R^2 statistics. As shown in Table 1, the λ-constrained model has a higher average log-likelihood and R^2 than the δ-constrained model, indicating that a model with no accumulation ($\delta = 0$) provides a worse fit than a model without

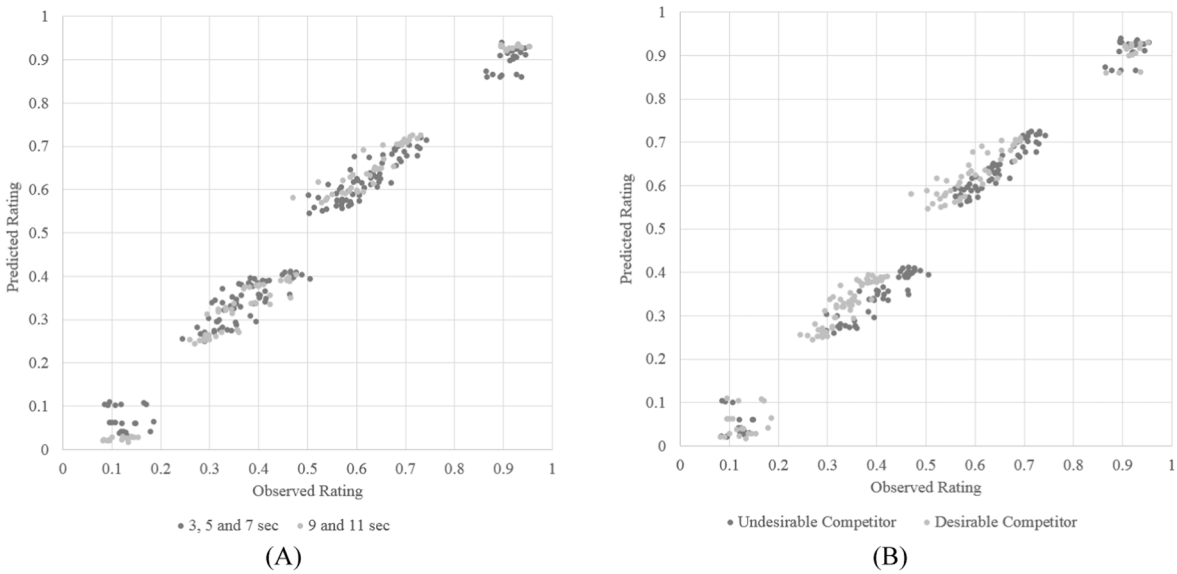


Fig. 9. A scatter plot showing mean observed participant ratings vs. mean predicted participant ratings (based on best fit individual-level models) for each of the 280 unique trials in Experiment 1. (A) shades markers based on deliberation time, whereas (B) shades the markers based on the competitor's desirability ("undesirable" if total attributes in competitor is 0 or 1; "desirable" if total attributes in competitor is 2 or 3).

competition ($\lambda = 0$).

A final analysis examined the average best fitting model parameters across participants. Overall, for the full model, we found relatively small values of λ and moderate values of δ. In terms of the average weights $\boldsymbol{\mu_w} = [\mu_{w1}, \mu_{w2}, \mu_{w3}]$, we also found higher values on the third attribute, affordability, compared to the first two attributes, proximity to campus and size. Comparing parameters across the models, we found that the full model and the λ-constrained model had similar values for nearly all flexible parameters. The full model was best fit with a slightly larger δ, most likely to compensate for the small amount of inhibition.

The parameters of the full model and λ-constrained model were, however, different to the parameters of the δ-constrained and baseline models, which had much larger attribute weights and noise standard deviations. This is likely due to the fact that the first two models accumulate preferences over time, and thus require relatively small weights (and corresponding small error terms) at each time step, in order to approximate the desirability ratings made after a number of time steps. In contrast, the δ-constrained model features only a small amount of inhibition, that results in gradually decreasing preferences over time, and the baseline model features no dynamism whatsoever, and thus fluctuates around a single stationary preference state. Both of these models thus require larger attribute weights and errors to capture participant desirability ratings.

In Fig. 9A and B we show a scatter plot comparing the predictions of the best-fit full model for each participant against observed participant ratings. Here each point corresponds to one of the 280 unique trials used in our experiment. For each trial, we computed both the average rating made by the participants for the trial, as well as the average rating (i.e. expected preference) predicted for the trial based on the best fitting parameters for each participant. As can be seen in these figures, there is a strong positive correlation between the predicted and observed ratings ($r = 0.99$, $p < 0.001$). Note that the four clusters of points visible in the figures correspond to trials in which the target object has zero attributes, one attribute, two attributes, and three attributes (each additional attribute led to a discrete jump in the desirability of the target, which is why the four clusters are so cleanly differentiated). Fig. 9A shades the points based on deliberation time (3, 5 or 7 s vs. 9 or 11 s), whereas Fig. 9B shades the points based on the desirability of the competitor (whether it has 0 or 1 vs. 2 or 3 attributes). The preference change and competition properties of the data and the model can be observed in these figures (which show that the dark and light points are shifted both on the x and the y axes).

5.4. Discussion

In Experiment 1 we tested the preference change and preference competition predictions of dynamic competitive multiattribute preference accumulators. Our experiment involved evaluations of hypothetical apartments defined on three attributes, and found qualitative and quantitative evidence in support of these predictions. Particularly, desirability ratings for desirable objects elicited after longer periods of deliberation were more positive than those elicited after short periods of deliberation. This pattern disappeared for undesirable objects. Likewise, desirability ratings for objects were higher if they were evaluated alongside undesirable competitors compared to if they were evaluated alongside desirable competitors.

In terms of model fitting, we compared the fits of the full model, which permitted both the dynamic accumulation of preferences as well as inhibition between preferences, with fits for two constrained models which individually restricted either accumulation (setting $\delta = 0$) or inhibition (setting $\lambda = 0$), as well as with fits for a baseline model which permitted neither accumulation nor inhibition ($\delta = \lambda = 0$). We found that the full model significantly outperformed the three other nested models, when evaluated on

the group level. The full model also outperformed the other three models for a substantial subset of participants, when evaluated on the individual level (though there were individual differences in the degree to which inhibition was needed to describe behavior). This shows that dynamic and competitive accumulation (corresponding to $\delta > 0$ and $\lambda > 0$) is necessary to account for behavior in the ratings task. This is likely due to the observed qualitative trends in the data: Our findings regarding preference change and preference competition cannot be captured without these critical assumptions, as demonstrated in the model properties and predictions section above.

6. Experiment 2

6.1. Overview

Although the results of Experiment 1 support the predictions of the multiattribute preference accumulator in desirability ratings tasks, it only examined two of three core properties outlined in this paper. Particularly, participants were asked to evaluate only one of the two presented objects, making it impossible to test the preference correlation property, generated by sequential attribute sampling. The goal of Experiment 2 was to provide a test of all three of the key properties of the model, while also replicating the results of Experiment 1 in a different choice domain. For this purpose, it again presented participants with pairs of multiattribute objects, and elicited desirability rating after an experimentally controlled deliberation time limit. However, unlike Experiment 1, participants were asked to rate both of the presented objects, allowing us to test whether there are significant correlations between the preferences for similar items.

6.2. Methods

6.2.1. Participants

A total of 120 participants performed this experiment in a behavioral laboratory, for monetary payment. Participants were members of a University of Pennsylvania experimental subjects pool.

6.2.2. Materials and procedures

As in Experiment 1, participants were asked to give desirability ratings, however, the objects in consideration were cars and not apartments. In each trial, participants were shown two cars, which differed on three attributes: mileage, size, and affordability (in a manner analogous to Figs. 5A and 6A). The attributes again were binary valued, so that the cars could have good mileage or not, be large or not, and be affordable or not.

After being shown two available cars, participants were asked to rate the desirability of both the cars. These ratings were elicited one after the other, on separate screens, and the information about the cars was not shown to participants on either of the ratings screen. As in Experiment 1, participants were asked to assess the absolute desirability of the car being rated (the target), rather than the desirability of the target relative to the other car (the competitor). Additionally, we varied the length of time the cars were displayed to the participants prior to the participant being prompted to make the rating. These deliberation times were either 5 s or 9 s. Once again, the ratings were made on a continuous scale, as shown in Fig. 6B.

This three-attribute setting again generates 8 possible cars and therefore 28 unique stimuli pairs. Each of the 28 pairs were presented twice for each display time condition. Participants rated both the cars in each trial, and the two presentations of a given stimuli pair varied which of the two cars in the pair was evaluated first. As there are 2 display time conditions, this resulted in a total of $28 \times 2 \times 2 = 112$ trials, with two ratings per trial (for a total of 224 ratings per participant). The order of presentation for these 112 trials, as well as whether the target was presented on the left (as Car X) or on the right (as Car Y), was randomized.

As in Experiment 1, we excluded all trials in which participants took more than 5 s to respond after being prompted. Additionally, we excluded 5 participants who did not complete the entire experiment.

6.3. Results

6.3.1. Qualitative tests

In our final data, we had a total of 24,276 observations across 115 participants. In each of the observations the participant rated the target, after being shown the target alongside a competitor for a fixed period of time. Note that unlike in Experiment 1, two observations were collected within each trial. Ultimately, the mean rating across all participants was 0.51, with a maximum rating of 1.00, a minimum rating of 0.00, and a median rating of 0.51.

As in Experiment 1, we first tested for the relationship between the desirability of the target (measured using the sum of its attributes) and the rating for the target. Again, we found that these two variables were highly correlated, and in fact the correlation coefficient for this relationship was nearly identical to that documented in Experiment 1 ($\rho = 0.73$, $p < 0.001$). A linear regression with ratings as the dependent variable and desirability of the target as the independent variable, as well as participant-level random intercepts, again found a strong significant relationship ($\beta = 0.25$, $z = 170.04$, $p < 0.001$). We ran a variant of this regression including the desirability (i.e. sum of all attributes) of the competitor as an additional independent variable and noted a significant negative effect of this variable on ratings of the target ($\beta = -0.02$, $z = -12.83$, $p < 0.001$). This once again shows that desirable competitors reduce the ratings of the target, as predicted by the preference competition property of the proposed accumulator model. The relationship between participant ratings and the desirability of the target and the competitor is illustrated in Fig. 10A and B

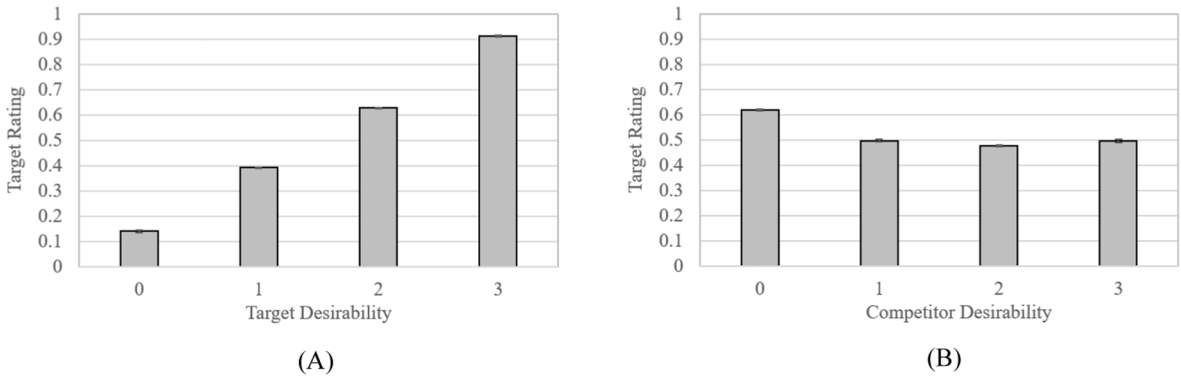


Fig. 10. Effect of target desirability (A) and competitor desirability (B) on target ratings in Experiment 2. The target ratings are averaged across pooled participant ratings for each level of desirability, and error bars indicate ± 1 SE.

respectively.

We also ran a linear regression to examine the effect of the deliberation time limit. This regression had ratings as the dependent variable, the desirability of the target, the deliberation time, and the interaction between the two, as the independent variables, and participant-level random intercepts. Our regression revealed a significant positive interaction effect ($\beta = 0.002$, $z = 2.61$, $p < 0.01$), again supporting our prediction that ratings for desirable targets increase with deliberation time relative to ratings for undesirable targets, consistent with the preference change property of the proposed model.

We also tested for the effect of deliberation time on ratings separately for our four levels of desirability. A regression featuring ratings as the dependent variable and deliberation time as an independent variable (as well as participant-level random intercepts) for highly desirable targets with all three attributes (i.e. sum of attributes = 3) revealed a positive significant effect of deliberation time ($\beta = 0.004$, $z = 2.77$, $p < 0.01$). We observed a similar positive significant effect ($\beta = 0.003$, $z = 3.59$, $p < 0.001$) for desirable targets with two out of the three attributes. This effect did not emerge for less undesirable targets with only one of out of the three attributes ($\beta = -0.001$, $z = -0.86$, $p = 0.39$) or for highly undesirable targets with none of the three attributes ($\beta = 0.001$, $z = 0.53$, $p = 0.60$). The relationship between deliberation time and participant ratings, as a function of target desirability, is shown in Fig. 11A and B.

Finally, we tested for the effect of object similarity on ratings. Based on the preference correlation property of the proposed model, we predicted that ratings for similar cars should be correlated, controlling for their desirability. In contrast, the ratings for dissimilar cars should be uncorrelated or negatively correlated. For the purposes of this test, we formalized similarity by computing the total number of attributes on which the two cars were identical. Thus, if two cars were described by attribute vectors [0,1,1] and [0,1,0], we obtained a similarity measure of $1 + 1 + 0 = 2$, corresponding to the fact that these vectors have the same values for the first and second attributes. The inverse of this metric, corresponding to dissimilarity, is known as Hamming distance (and can be interpreted as the minimum number of substitutions required to change one binary vector into the other). In our study, the maximum similarity between two cars in a trial was 2 (as we did not present pairs of identical cars), whereas the minimum similarity was 0.

Overall, we found a strong positive correlation between the ratings of two cars across the trials in this experiment, if the two cars were identical on two attributes and had a similarity of 2 ($r = 0.48$, $p < 0.001$). There was a moderate negative correlation between the ratings if the two cars were identical on only one attribute and had a similarity of 1 ($r = -0.18$, $p < 0.001$), and a strong

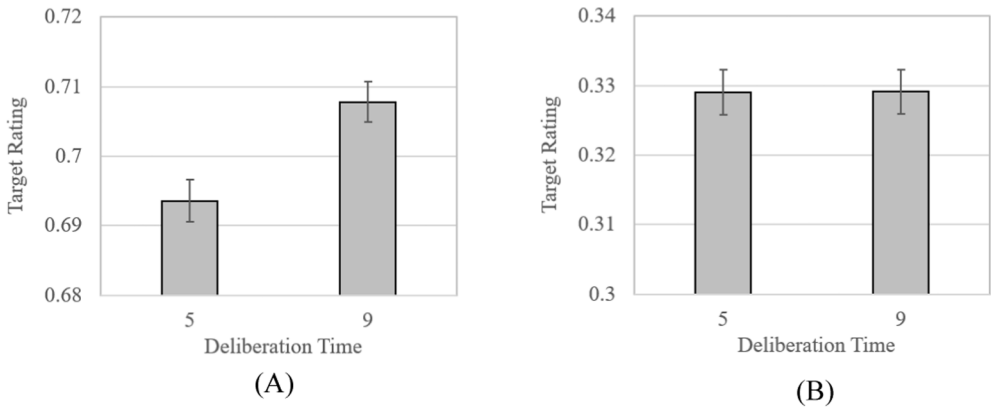


Fig. 11. Effect of deliberation time (in seconds) on target ratings for desirable targets with a desirability of 2 or 3 (A) and undesirable targets with a disability of 0 or 1 (B), in Experiment 2. The target ratings are averaged across pooled participant ratings for each deliberation time condition, and error bars indicate ± 1 SE.

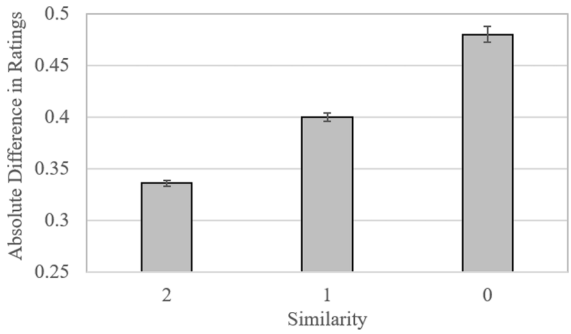


Fig. 12. Effect of object similarity on the absolute difference in ratings for the two objects in Experiment 2. Here the absolute difference in the ratings for the two objects are averaged across pooled participant ratings for each level of similarity, and error bars indicate ± SE.

negative correlation if the two cars were identical on none of the attributes and had a similarity of 0 ($r = -0.50$, $p < 0.001$).

Although these results suggest that our predictions are reflected in the data, it is important to note that this analysis does not control for object desirability: It is possible that similar cars have highly correlated ratings because they are similarly desirable. It is possible to control for desirability by comparing trials that had the same objects in terms of desirability but not in terms of relative similarity. Again, for this purpose, we approximated the desirability of a car using the total number of attributes in the car. With this measure, there were only two types of trials meeting our criterion: Trials in which one car had two attributes and the other car had the third attribute (generating a similarity of 0 and a desirability difference of 1 attribute), and trials in which one car had two attributes and the other car had one of these two attributes (generating a similarity of 2 and a desirability difference of 1 attribute). An example of the former trial would involve attribute vectors [1,1,0] and [0,0,1], whereas an example of the latter trial would involve attribute vectors [1,1,0] and [1,0,0]. Overall, we found the correlations between the responses in the first type of trial to be significantly negative ($r = -0.21$, $p < 0.001$) and the correlations between the responses in the second type of trial to be significantly positive ($r = 0.05$, $p < 0.01$), providing a more rigorous demonstration of our prediction.

Yet another way to test this relationship involves calculating the absolute difference in the ratings of the two cars in a trial and regressing this on the absolute difference in desirability and the similarity of the cars. Defining desirability using the total number of attributes in a car, and allowing for participant-level random intercepts, this regression revealed not only a positive effect of the desirability difference ($\beta = 0.20$, $z = 83.51$, $p < 0.001$), but also a strong negative effect of similarity ($\beta = -0.03$, $z = -12.73$, $p < 0.001$). Thus, participants were likely to give similar cars similar ratings, controlling for the relative desirabilities of the cars. The relationship between object similarity and participant ratings is shown in Fig. 12.

6.3.2. Quantitative tests

Consistent with Experiment 1, the above analysis shows that the preference change and preference competition properties of the proposed model are reflected in the data. It also finds support for the preference correlation property in the data, with similar objects being given more similar ratings than dissimilar objects (keeping desirability differences fixed). As in Experiment 1, we also tested these properties using a model-based approach. In this experiment, we obtained ratings for both the objects in a given trial and thus were able to fit our full model (with a flexible standard deviation σ_w for the attentional weight vector), as well as constrained variants of this model, using the methods outlined earlier in this paper.

Overall, we considered five models: the full model, which allows for flexible $\boldsymbol{\mu_w}$, δ, λ, σ_w, and σ_ε and thus displays preference correlation, preference change, and preference competition; a λ-constrained model that allows for flexible $\boldsymbol{\mu_w}$, δ, σ_w, and σ_ε, but has no inhibition ($\lambda = 0$) and thus displays only preference correlation and preference change but not preference competition; a δ-constrained model that allows for flexible $\boldsymbol{\mu_w}$, λ, σ_w, and σ_ε, but has no accumulation ($\delta = 0$) and thus displays only preference correlation and preference competition but not preference change; a σ_w-constrained model that allows for flexible $\boldsymbol{\mu_w}$, δ, λ, and σ_ε, but has no variability in attention weights ($\sigma_w = 0$) and thus displays only preference competition and preference change but not preference correlation; and a baseline model that allows for flexible $\boldsymbol{\mu_w}$ and σ_ε, but sets $\sigma_w = \delta = \lambda = 0$ and thus displays none of the three key properties of preference accumulators. As in Experiment 1, superior fits for the full model relative to the remaining nested models would provide strong support for the unique assumptions of this model.

We fit each of these models on the participant level using the methods outlined earlier in this paper. The model fits are summarized in Table 2. This table displays parameter values, log-likelihood values, and R^2 values, for the models averaged across participants, with standard errors in brackets. The table also displays the statistic, G^2, and degrees of freedom, for the likelihood ratio test (LRT) contrasting the fits of the full model against the nested models on the group level, as well as the proportion of participants significantly ($p < 0.05$) better fit by the full model relative to the remaining models, according the LRT applied separately for each participant.

Table 2 shows that the full model was able to describe the data with a high level of accuracy by achieving an average R^2 values of 0.62. This was only slightly larger than that of the λ-constrained model, though it substantially outperformed the δ-constrained, σ_w-constrained, and baseline models. To more formally control for the flexibility of the full model, we need to use the likelihood ratio test (LRT). When applied to the group level this test showed that the full model significantly ($p < 0.001$) outperformed the four

Table 2
Average best-fitting parameter values, log-likelihood values, and R^2 values, for fits in Experiment 2 (brackets indicate standard errors). "G^2/DF" corresponds to the test statistic and degrees of freedom for the likelihood ratio test comparison against the full model, on the group level. This statistic is significantly different to 0 for all models ($p < 0.001$). "% Better Fit by Full" indicates the proportion of participants significantly ($p < 0.05$) better fit by the full model relative to the nested model according to the likelihood ratio test on the individual level. Note the G^2 for each model can be found by multiplying the difference between the log-likelihood of the full model and the more constrained model for each participant by 2 and then summing across the participants.

	Full	λ-constrained	δ-constrained	σ_w-constrained	Baseline
δ	0.66 (0.02)	0.63 (0.03)	[0]	0.59 (0.03)	[0]
λ	0.01 (0.00)	[0]	0.02 (0.00)	0.01 (0.00)	[0]
σ_w	0.01 (0.00)	0.02 (0.00)	0.02 (0.00)	[0]	[0]
σ_ε	0.05 (0.00)	0.06 (0.00)	0.17 (0.01)	0.07 (0.00)	0.19 (0.01)
μ_{w1}	0.12 (0.01)	0.13 (0.01)	0.34 (0.01)	0.14 (0.01)	0.34 (0.01)
μ_{w2}	0.08 (0.01)	0.09 (0.01)	0.26 (0.01)	0.10 (0.01)	0.25 (0.01)
μ_{w3}	0.13 (0.01)	0.14 (0.01)	0.37 (0.01)	0.16 (0.01)	0.37 (0.01)
Log-likelihood	93.53 (8.13)	91.46 (8.24)	76.55 (8.67)	84.09 (7.87)	72.93 (8.27)
R^2	0.62 (0.03)	0.62 (0.03)	0.44 (0.09)	0.57 (0.04)	0.44 (0.09)
ΔG^2 / DF	–	482.52/117	3971.66/117	2208.71/117	4819.63/351
% Better Fit by Full	–	26%	73%	50%	74%

remaining models. Thus, on aggregate, allowing for flexible δ, flexible λ, and flexible σ_w is necessary to describe behavior.

Applying the LRT at the individual level, we also found that the full model significantly ($p < 0.05$) outperformed the baseline model and the δ-constrained for the majority of participants. The full model similarly outperformed the σ_w-constrained model for half the participants, and the λ-constrained model for a substantial minority of participants. Thus, taken together with the aggregate fits, the fits at the individual level suggest that a flexible δ, flexible λ, and flexible σ_w are necessary to adequately describe behavior. Though, they also reveal that there are individual differences in the degree to which each of these components is needed particularly for the lateral inhibition component (λ-constrained model). These results replicate and extend the findings of Experiment 1.

It is also useful to compare the λ-constrained, δ-constrained, and σ_w-constrained models against each other to test which of the three parameters is more important for describing behavior. These three models have an equivalent number of parameters can thus can be compared using only their log-likelihoods and R^2 statistics. As with Experiment 1, we found the λ-constrained model has the highest average log-likelihood and R^2. This is followed by the σ_w-constrained model. The δ-constrained model had the worst fits. Taken together, these model comparisons imply that the most important component is accumulation, as captured by the parameter δ.

A final analysis examined the average best fitting model parameters across participants. Overall, for the full model, we found relatively small values of λ and moderate values of δ, consistent with Experiment 1. We also noted fairly small values of σ_w. Comparing parameters across the models, we found that the full model, the λ-constrained model, and the σ_w-constrained model had similar values for nearly all flexible parameters. Again, the full model was best fit with a slightly larger δ, most likely to compensate for the small amount of inhibition. Likewise, the σ_w-constrained model was fit with a slightly larger value of σ_ε, most likely to compensate for the absence of noise in attribute sampling.

The parameters of these three models were, however, different to the parameters of the δ-constrained and baseline models, which, as in Experiment 1, had much larger attribute weights and noise standard deviations. Again, this is likely due to the fact that the first three models accumulate preferences over time, and thus require relatively small weights (and corresponding small error terms) at each time step, in order to approximate the desirability ratings made after a number of time steps. This is not the case for the δ-constrained and baseline models.

Fig. 13A and B show a scatter plot comparing the predictions of the best-fit full model for each participant against participant ratings. Here each point corresponds to one of the 112 unique trials used in our experiment. For each trial, we computed both the average rating made by the participants for the trial, as well as the average rating predicted for the trial based on the best fitting parameters for each participant. Note that each trial was presented twice (in order to counterbalance item ordering) for each participant, so these averages include trial repetitions. As can be seen in these figures, there is a strong positive correlation between the predicted and observed ratings ($r = 0.99$, $p < 0.001$). Again, the four clusters of points visible in the figures correspond to trials in which the target object has zero attributes, one attribute, two attributes, and three attributes. Fig. 13A shades the points based on deliberation time (5 vs. 9 s), whereas Fig. 13B shades the points based on the desirability of the competitor (whether it has 0 or 1, or 2 or 3 attributes). The preference change and competition properties of the data and the model can be observed in these figures (which show that the dark and light points are shifted both on the x and the y axes).

6.4. Discussion

Experiment 2 tested the preference correlation, preference change, and preference competition predictions of dynamic competitive multiattribute preference accumulators with stochastic attribute sampling. Unlike Experiment 1, it involved the evaluation of hypothetical cars and furthermore asked participants to rate both the presented cars. Despite these differences, this experiment replicated the findings of Experiment 1 regarding preference change and preference competition. Ratings for desirable objects elicited after longer periods of deliberation were more positive than those elicited after short periods of deliberation, relative to ratings for

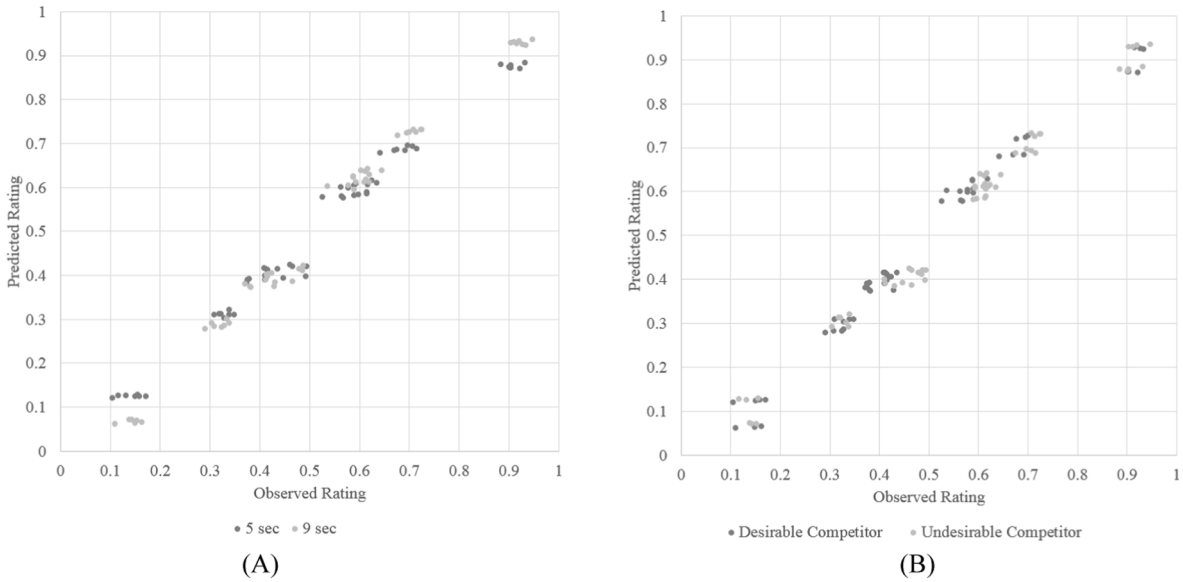


Fig. 13. A scatter plot showing mean observed participant ratings vs. mean predicted participant ratings (based on best fit individual-level models) for each of the 112 unique trials in Experiment 2. (A) shades markers based on deliberation time, whereas (B) shades the markers based on the competitor's desirability ("undesirable" if total attributes in competitor is 0 or 1; "desirable" if total attributes in competitor is 2 or 3).

undesirable objects. Additionally, desirability ratings for objects were higher if they were evaluated alongside undesirable competitors compared to if they were evaluated alongside desirable competitors. Experiment 2 also found novel evidence for preference correlation: The ratings for similar objects were positively correlated, whereas the ratings for dissimilar objects were negatively correlated (controlling for object desirability).

In terms of model fitting, we compared the fits of the full model, which permitted the dynamic competitive accumulation of preferences with stochastic attribute sampling, with fits for three constrained models which individually restricted either accumulation (setting $\delta = 0$), inhibition (setting $\lambda = 0$), or stochasticity in attribute sampling (setting $\sigma_w = 0$). We also compared the fits of the full model with fits for a baseline model which permitted none of these three properties ($\delta = \lambda = \sigma_w = 0$). We found that the full model significantly outperformed the four other nested models, when evaluated on the group level. Similar results were observed for a substantial subset of participants, when fits were evaluated on the individual level (though there were individual differences in the degree to which each of the components of the full model was needed). Overall, this shows that dynamic and competitive accumulation with stochastic attribute sampling (corresponding to $\delta > 0, \lambda > 0$, and $\sigma_w > 0$) is necessary to account for behavior in the ratings task. Again, this is likely due to the observed qualitative trends in the data: Our findings regarding preference change, preference competition, and preference correlation, cannot be captured without these critical assumptions, as demonstrated in the model properties and predictions section above.

7. General discussion

7.1. Predicting desirability ratings

What are the cognitive processes at play in the evaluation of objects and the construction of preferences? This question is at the heart of our understanding of human decision making and has received considerable attention in psychology and in a number of related fields. It is typically addressed using experimental tasks with discrete choices. However, the general cognitive processes at play in preference construction should extend beyond a given response mode, and theories formulated to describe choice should be able to make predictions and account for behavior in other types of preference-based tasks.

In this paper, we have attempted to extend the preference accumulation framework, initially formulated to describe choice behavior, to model desirability ratings. We adopted a multiattribute preference accumulator with decay, inhibition, and the stochastic sampling of attributes, and assumed that that desirability ratings for objects, at the time of rating, are determined by mapping accumulated preference states directly onto the response scale. Such a model makes three key predictions with regards to desirability ratings. The first prediction involves preference correlations, and states that the desirability ratings of objects elicited at a given point in time should be more correlated if the objects are highly similar than if they are dissimilar. The second prediction involves preference change over time, and states that the ratings of more desirable objects should increase with deliberation time relative to the ratings for less desirable objects. The third prediction involves preference competition, and states that desirable competitors should reduce the ratings of target objects, relative to undesirable competitors. These three predictions depend critically on the three unique assumptions of the model: the stochastic sampling of attributes, the accumulation of preferences over time, and inhibition

between preferences, so that a model without these three assumptions is unable to generate preference correlation, change, or competition.

We conducted two experiments to examine these predictions. Our experiments elicited desirability ratings for multiattribute objects, with trials varying in terms of the attributes of the rated object, the attributes of the competitor, and the amount of deliberation time permitted. The qualitative analysis of ratings data from these experiments strongly supported the preference correlation, change, and competition predictions of the preference accumulation model. Additionally, quantitative model fitting on these data compared the proposed model to nested variants, without preference accumulation, without inhibition, or without stochastic attribute sampling. The full model outperformed these nested variants, as these variants were unable to account for the preference change, competition, and correlation properties of the data. The full model also outperformed a baseline multiattribute linear model, which featured none of the dynamic, stochastic, and context-dependent properties of the proposed preference accumulator.

7.2. Alternate models

In this paper we adopted a multiattribute preference accumulator with decay, inhibition, and the stochastic sampling of attributes, as it is a core model used in numerous previous applications of the preference accumulation framework. For example, such a model has been shown to predict behavioral patterns pertaining to stochasticity, dynamism, and context dependence, in choice (Bhatia, 2013; Bhatia & Mullett, 2016; Diederich & Busemeyer, 1999; Roe et al., 2001; Usher & McClelland, 2004). Quantitative model fits also favor this model, relative to competitor models without these specific assumptions (Berkowitsch et al., 2014; Turner et al., 2018). Finally, this model can be shown to be a generalization of the standard weighted additive multiattribute model (Keeney & Raiffa, 1993), both for discrete choice and for desirability ratings (Roe et al., 2001).

That said, there are also other models of choice behavior that we could have considered for the purposes of this paper. One such model is the drift diffusion model (DDM) (Ratcliff, 1978; Ratcliff & Mckoon, 2008) (see also Ratcliff, in press; Smith, 2016). As with the preference accumulator model, the DDM aggregates preferences over time, with additive noise. When combined with the assumption that desirability ratings are determined by mapping preferences directly onto the response scale, this model can also be used to predict ratings behavior. However, one critical limitation of this model is that it assumes that preferences are purely relative, that is, drift rates are determined entirely by the differences in the desirabilities of the available objects. This causes the DDM to predict a very high degree of preference competition. Thus, for example, the simple instantiation of the DDM examined here, predicts that the rating for $\boldsymbol{x}_1 = [1,0,0]$ when presented alongside $\boldsymbol{x}_2 = [0,0,0]$, should be the same as the rating for $\boldsymbol{x}_1' = [1,1,1]$ when presented alongside $\boldsymbol{x}_2' = [0,1,1]$. A preliminary analysis of our results shows that this is not the case. For example, in Experiment 1, the average rating for $\boldsymbol{x}_1 = [1,0,0]$ when presented alongside $\boldsymbol{x}_2 = [0,0,0]$ is 0.41, whereas the average rating for $\boldsymbol{x}_1' = [1,1,1]$ when presented alongside $\boldsymbol{x}_2' = [0,1,1]$ is 0.95. Likewise, the regression analyses that include both the desirability of the target and the desirability of the competitor as independent variables find coefficients for these variables that differ by an order of magnitude (rather than being identical, as the DDM model suggests). Thus, in Experiment 1, increasing the desirability of the target by one unit increases ratings of the target by about 0.25 points. In contrast, decreasing the desirability of the competitor by one unit only increases the ratings of the target by 0.03 rating points. Now, it is useful to note that we did, implicitly, permit the DDM model in our fits. As our proposed preference accumulator mimics the DDM when $\delta = \lambda$ (e.g. Bogacz et al., 2006), our full model, with flexible δ and λ, can predict behavior generated by the DDM. Our fits reveal that the best fit δ is much larger than the best fit lambda λ implying that the DDM is not a good model for the type of ratings data we are studying.

The reason why the DDM model fails is quite intuitive. Participants are asked to rate the absolute desirabilities of the objects in our task. Although they are vulnerable to relativistic comparisons, the degree to which their ratings are relative (rather than absolute) is much weaker than would be predicted by a DDM with drift rates based on object utility differences. Note that this is not necessarily a problem for the DDM in discrete choice tasks, where participants have to explicitly compare objects against each other to determine the one they prefer the most.

Another model that could have been considered for the purposes of this paper is the accumulator model without preference decay (or inhibition) (Brown & Heathcote, 2008; Ratcliff & Smith, 2004). Again, such a model aggregates preferences over time, but assumes that accumulation is perfect. This model could be combined with the assumption that desirability ratings are determined by mapping preferences directly onto the response scale, and thus be used to predict ratings behavior. However, the critical limitation of this model is that it predicts that preferences are linear in time, and subsequently that observed ratings should increase linearly with the deliberation time limit. A quick examination of our behavioral data (e.g. Fig. 8) shows that this is not the case. Although ratings for desirable objects do increase with time, this pattern is certainly not linear. Rather it mimics the concave relationship predicted by the (imperfect) accumulator model with decay (a pattern that is also generated by the model proposed by Smith and Ratcliff (2009)). Again, note that we did, implicitly, permit the perfect accumulator model in our fits. As our proposed model mimics the perfect accumulator when $\delta = 1$ and $\lambda = 0$, our full model, with flexible δ and λ, can predict behavior generated by the perfect accumulator. Our fits reveal that the best fit δ is much smaller than 1, implying that the perfect accumulator is not a good model for the type of ratings data we are studying. Again, perfect accumulation is not necessarily a bad model in choice tasks, in which only relative preferences matter. The difference here is that we are attempting to predict continuous preference ratings, for which linearity in time is a very directly observable (and quite unreasonable) prediction.

7.3. Procedure invariance

Procedure invariance is the claim that the revealed preferences of individuals should be independent of the way in which they are

elicited (Slovic, 1995). Thus, for example, rankings observed through choice should be consistent with the implicit rankings of objects made in other preference elicitation tasks (such as desirability ratings, or willingness to pay assessments). There is, by now, considerable evidence that procedure invariance does not always hold, and that changing the response mode can alter how decision makers evaluate objects (Lichtenstein & Slovic, 1971; Tversky, Sattath & Slovic, 1988). Yet, our results that show that cognitive models initially developed to describe choice also describe behavior in continuous desirability tasks imply that procedure invariance does hold. Importantly, our results pertain primarily to the core cognitive process involved in preference construction, and do not extend to the parameters that determine the final output of this process. It is quite likely that parameters like preference decay and inhibition vary as a function of the response mode. Indeed, attribute weights may also vary, as different response modes could make different attributes appear more or less important (this is the explanation for violations of procedure invariance in risky choice, proposed by Tversky et al., 1988).

In fact, our finding that models of preference accumulation used in discrete choice also describe the broader processes underlying preference formation provides a formal approach to testing for and modelling violations of procedure invariance. By using the same core model to describe preference construction across multiple response modes, we can better understand the effects of these response modes on core decision parameters and resulting choice behavior. In fact, Hawkins et al. (2014a and 2014b), have already attempted such a test for violations of procedure invariance in accept vs. reject tasks. Hawkins et al. propose accumulation-based models that are able to predict responses in both types of tasks, and that can, in turn, be used to characterize differences in the cognitive processes involved in the two tasks. The results of their analysis complement the claims in this paper: Hawkins et al. find that the same core model is able to characterize behavior in both tasks, though the parameters of this model do vary based on response mode. We hope to adopt this approach in future work, and extend our understanding of the similarities and difference between desirability ratings and choice, by fitting our multiattribute accumulator to both choice and ratings data simultaneously.

7.4. Pricing tasks

By applying an accumulator model to describe responses made on a continuous scale, our paper shares the same goals as Johnson and Busemeyer (2005). Johnson and Busemeyer explain observed violations of procedure invariance in tasks involving pricing judgments, such as certainty equivalents for risky gambles. To do so, they use a sequential value matching process that first considers a candidate response value (e.g. a certainty equivalent), and then, with the use of an accumulation process, determines whether the candidate response is higher, lower, or equal in desirability to the object (e.g. gamble) being evaluated. The output of this accumulation process determines whether or not the decision maker reports the candidate response value, or continues deliberation with a second, higher or lower, candidate response.

Our approach differs quite substantially from that proposed by Johnson and Busemeyer. Crucially we do not assume a two-stage process. Rather, we assume that the preference state at the time of response maps directly onto the response scale. This difference is due largely to the fact that desirability ratings, unlike pricing judgments, do not involve responses that can be evaluated alongside the objects in consideration. For example, a candidate certainty equivalent can be compared against a focal gamble (both objects involve payoff and probability attributes), and thus be determined to be more desirable less desirable, or equally desirable to the gamble. This does not make sense for an abstract numerical rating.

This difference implies that although Johnson and Busemeyer (2005) also attempt to model a continuous response task using the accumulation framework, their model cannot directly be applied to the study of desirability ratings. Conversely, even though our approach could be used to model pricing judgments, it is unlikely that it would be able to account for the types of behavioral patterns studied by Johnson and Busemeyer (2005). Despite these differences, both our approach and Johnson and Busemeyer's approach can be seen as being part of the same core model, as both rely on an accumulation process to perform evaluations (see also Kvam & Busemeyer, 2018). This core model is, of course, the same model used by numerous scholars to describe behavior in discrete choice tasks. In future work, it would be useful to integrate these different models to build a single theoretical framework for studying multiattribute preferences in continuous response tasks.

7.5. Beyond multiattribute decision making

Our findings have implications for the study of behavior outside of multiattribute decision making. Indeed, we believe that our broader model and the effects that it uniquely describes apply to other desirability ratings tasks studied by cognitive psychologists and decision researchers. For example, we would expect preference change and competition in nearly any desirability rating task, and preference correlation in desirability ratings tasks that involve the stochastic sampling of option consequences (such as states of the world in risky choice and time periods in intertemporal choice).

Desirability ratings tasks are also commonly used in social and clinical psychology, as well as areas outside of psychology, such as political science and marketing. The psychometric properties of desirability ratings have been the subject of considerable research inquiry, yet there has been surprisingly little research investigating the cognitive processes at play in generating these desirability ratings. A formal characterization of these processes can provide a theoretically grounded account of when such ratings are most likely to be accurate, when they are likely to be distorted by contextual factors, and how this may vary as a function of deliberation time and other task-based constraints. The model proposed in this paper provides a framework for such an analysis, and can be applied without loss of generality to desirability ratings for friends and relatives, celebrities and politicians, brands and consumer products, as well as past experiences and future goals. Even though many of these naturalistic objects do not have the explicit multiattribute structure assumed in our experimental work, these objects are nonetheless composed of different features, and thus

subject to a sequential sampling process that accumulates information about these features over time (potentially with decay and inhibition). If this is the case, then we would expect to observe the types of stochasticity, dynamism, and context-dependence studied in this paper, in these diverse domains. Indeed, there is already some evidence for a type of context-dependence in social ratings (see e.g. Chang & Cikara, 2018, for a recent overview), illustrating the cross-disciplinary potential of further research on this topic.

8. Conclusion

The evaluation of objects involves two crucial steps: the construction of preference, and the mapping of constructed preference onto responses. In this paper we have proposed that the core mechanisms of models of multiattribute preference accumulation -the stochastic sampling of attributes and the aggregation of preferences with decay and inhibition- can be used describe the construction of preference in desirability ratings tasks. With the added assumption that these preferences are transformed directly onto ratings, we find that it is possible to predict desirability ratings with a high degree of accuracy. Such an approach also successfully describes the stochasticity, dynamism, and context-dependence inherent in desirability ratings behavior. Overall, our results show how models formulated for discrete choice can be successfully adapted to continuous ratings. By doing so they shed light on the core cognitive mechanisms at play in the construction of preference, and illustrate a cohesive approach to predicting behavior across different response modes.

Reference

Andraszewicz, S., Rieskamp, J., & Scheibehenne, B. (2015). How outcome dependencies affect decisions under risk. *Decision, 2*(2), 127–144.

Becker, G. S. (1976). *The economic approach to human behavior*. University of Chicago Press.

Berkowitsch, N. A., Scheibehenne, B., & Rieskamp, J. (2014). Rigorously testing multialternative decision field theory against random utility models. *Journal of Experimental Psychology: General, 143*(3), 1331–1348.

Betsch, T., Kaufmann, M., Lindow, F., Plessner, H., & Hoffmann, K. (2006). Different principles of information integration in implicit and explicit attitude formation. *European Journal of Social Psychology, 36*, 887–905.

Betsch, T., Plessner, H., Schwieren, C., & Gütig, R. (2001). I like it but I don't know why: A value-account approach to implicit attitude formation. *Personality and Social Psychology Bulletin, 27*, 242–253.

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

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

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. *Psychological Review, 113*(4), 700–765.

Brenner, L., Rottenstreich, Y., & Sood, S. (1999). Comparison, grouping, and preference. *Psychological Science, 10*(3), 225–229.

Brown, S. D., & Heathcote, A. (2008). The simplest complete model of choice response time: Linear ballistic accumulation. *Cognitive Psychology, 57*(3), 153–178. https://doi.org/10.1016/j.cogpsych.2007.12.002.

Brusovansky, M., Vanunu, Y., & Usher, M. (2018). Why we should quit while we're ahead: When do averages matter more than sums? *Decision* in press.

Busemeyer, J. R., & Diederich, A. (2002). Survey of decision field theory. *Mathematical Social Sciences, 43*(3), 345–370.

Busemeyer, J. R., & Townsend, J. T. (1993). Decision field theory: A dynamic-cognitive approach to decision making in an uncertain environment. *Psychological Review, 100*(3), 432–459.

Camerer, C. F., Loewenstein, G., & Rabin, M. (Eds.). (2011). *Advances in behavioral economics*. Princeton University Press.

Chang, L. W., & Cikara, M. (2018). Social decoys: Leveraging choice architecture to alter social preferences. *Journal of Personality and Social Psychology, 115*, 206–223.

Dai, J., Pleskac, T. J., & Pachur, T. (2018). Dynamic cognitive models of intertemporal choice. *Cognitive Psychology, 104*, 29–56.

Diederich, A. (1997). Dynamic stochastic models for decision making with time constraints. *Journal of Mathematical Psychology, 41*, 260–274.

Diederich, A. (2003). Decision making under conflict: Decision time as a measure of conflict strength. *Psychonomic Bulletin & Review, 10*(1), 167–176.

Diederich, A., & Busemeyer, J. R. (1999). Conflict and the stochastic-dominance principle of decision making. *Psychological Science, 10*(4), 353–359.

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

Gold, J. I., & Shadlen, M. N. (2007). The neural basis of decision making. *Annual Review of Neuroscience, 30*, 535–574.

Green, P. E., & Srinivasan, V. (1978). Conjoint analysis in consumer research: Issues and outlook. *Journal of Consumer Research, 5*(2), 103–123.

Hawkins, G. E., Marley, A. A. J., Heathcote, A., Flynn, T. N., Louviere, J. J., & Brown, S. D. (2014b). The best of times and the worst of times are interchangeable. *Decision, 1*(3), 192–214.

Hawkins, G. E., Marley, A. A. J., Heathcote, A., Flynn, T. N., Louviere, J. J., & Brown, S. D. (2014a). Integrating cognitive process and descriptive models of attitudes and preferences. *Cognitive Science, 38*(4), 701–735.

Johnson, J. G., & Busemeyer, J. R. (2005). A dynamic, stochastic, computational model of preference reversal phenomena. *Psychological Review, 112*(4), 841–861.

Keeney, R. L., & Raiffa, H. (1993). *Decisions with multiple objectives: Preferences and value trade-offs*. Cambridge University Press.

Krajbich, I., Armel, C., & Rangel, A. (2010). Visual fixations and the computation and comparison of value in simple choice. *Nature Neuroscience, 13*, 1292–1298.

Kvam, P. D., & Busemeyer, J. R. (2018). Dynamic and distributional properties of prices. In C. Kalish, M. Rau, J. Zhu, & T. T. Rogers (Eds.). *Proceedings of the 40th annual conference of the Cognitive Science Society* (pp. 653–658). Madison, WI: Cognitive Science Society.

Kvam, P. D. (2018). A geometric framework for modeling decisions among arbitrarily many alternatives. https://doi.org/10.31234/osf.io/7r8y9.

Lichtenstein, S., & Slovic, P. (1971). Reversals of preference between bids and choices in gambling decisions. *Journal of Experimental Psychology, 89*(1), 46–55.

Likert, R. (1932). A technique for the measurement of attitudes. *Archives of Psychology*.

Nunnally, J. C. (1967). *Psychometric theory*. New York: McGraw-Hill.

Pleskac, T. J., & Busemeyer, J. R. (2010). Two-stage dynamic signal detection: a theory of choice, decision time, and confidence. *Psychological Review, 117*(3), 864–901.

Pleskac, T. J., Yu, S., Hopwood, C., & Liu, T. (2018). Characterizing deliberation during preferential choice. *Decision* in press.

Ratcliff, R. (1978). A theory of memory retrieval. *Psychological Review, 85*(2), 59.

Ratcliff, R. (2018). Decision making on spatially continuous scales. *Psychological Review* in press.

Ratcliff, R., & McKoon, G. (2008). The diffusion decision model: Theory and data for two-choice decision tasks. *Neural Computation, 20*(4), 873–922.

Ratcliff, R., & Rouder, J. N. (1998). Modeling response times for two-choice decisions. *Psychological Science, 9*(5), 347–356.

Ratcliff, R., & Smith, P. L. (2004). A comparison of sequential sampling models for two-choice reaction time. *Psychological Review, 111*(2), 333.

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.

Samuelson, P. A. (1947). *Foundations of economic analysis*. Harvard University Press.

Slovic, P. (1995). The construction of preference. *American Psychologist, 50*(5), 364–371.

Smith, P. L. (2016). Diffusion theory of decision making in continuous report. *Psychological Review, 123*, 425–451.

Smith, P. L., & Ratcliff, R. (2009). An integrated theory of attention and decision making in visual signal detection. *Psychological Review, 116*(2), 283.

Teodorescu, A. R., & Usher, M. (2013). Disentangling decision models: From independence to competition. *Psychological Review, 120*(1), 1–38.
Trueblood, J. S., Brown, S. D., & Heathcote, A. (2014). The multiattribute linear ballistic accumulator model of context effects in multialternative choice. *Psychological Review, 121*(2), 179–205.
Tsetsos, K., Chater, N., & Usher, M. (2012). Salience driven value integration explains decision biases and preference reversal. *Proceedings of the National Academy of Sciences of the United States of America, 109*(24), 9659–9664.
Turner, B. M., Schley, D. R., Muller, C., & Tsetsos, K. (2018). Competing theories of multialternative, multiattribute preferential choice. *Psychological Review, 125*, 329–362.
Tversky, A., Sattath, S., & Slovic, P. (1988). Contingent weighting in judgment and choice. *Psychological Review, 95*(3), 371–384.
Usher, M., & McClelland, J. L. (2001). The time course of perceptual choice: the leaky, competing accumulator model. *Psychological Review, 108*(3), 550–592.
Usher, M., & McClelland, J. L. (2004). Loss aversion and inhibition in dynamical models of multialternative choice. *Psychological Review, 111*(3), 757–769.
Vanunu, Y., Pachur, T., & Usher, M. (2018). Constructing preference from sequential samples: The impact of evaluation format on risk attitudes. *Decision*. https://doi.org/10.1037/dec0000098 in press.
Weber, E. U., & Johnson, E. J. (2009). Mindful judgment and decision making. *Annual Review of Psychology, 60*, 53–85.
Yu, S., Pleskac, T. J., & Zeigenfuse, M. D. (2015). Dynamics of postdecisional processing of confidence. *Journal of Experimental Psychology: General, 144*(2), 489.
Zeigenfuse, M. D., Pleskac, T. J., & Liu, T. (2014). Rapid decisions from experience. *Cognition, 131*(2), 181–194.