Evidence Accumulates for Individual Attributes during Value-Based Decisions

Douglas G. Lee [1] Todd A. Hare [2, 3]

[1] Institute of Cognitive Sciences and Technologies, National Research Council of Italy

[2] Zurich Center for Neuroeconomics, Department of Economics, University of Zurich

[3] Neuroscience Center Zurich, University of Zurich and ETH Zurich

Email for correspondence:

DouglasGLee@gmail.com

Abstract

When choosing between different options, we tend to consider specific attribute qualities rather than deliberating over some general sense of the objects' overall values. The importance of each attribute together with its quality will determine our preference rankings over the available alternatives. Here, we show that the relative importance of the latent attributes within food rewards reliably differs when the items are evaluated in isolation compared to when binary choices are made between them. Specifically, we used standard regression and sequential sampling models to examine six datasets in which participants evaluated, and chose between, multi-attribute snack foods. We show that models that assume that attribute importance remains constant across evaluation and choice contexts fail to reproduce fundamental patterns in the choice data and provide quantitatively worse fits to the choice outcomes, response times, and confidence reports compared to models that allow for attribute importance to vary across preference elicitation methods. Our results provide important evidence that incorporating attribute-level information into computational models helps us to better understand the cognitive processes involved in value-based decision-making.

Keywords: multi-attribute choice, value-based choice, preferential choice, drift-diffusion model, DDM

Introduction

Most decisions that we make are based on information about a variety of relevant features of the available options. Theories and mathematical models of multi-attribute choice generally agree that, in principle, the decision system in our brains should compare options based on how well they score across all relevant attribute dimensions (Bettman et al., 1998; Gigerenzer & Gaissmaier, 2011; E. J. Johnson & Payne, 1985; Keeney et al., 1993; Levav et al., 2010; Payne et al., 1988, 1993; Russo et al., 1996; Shah & Oppenheimer, 2008). These overall scores, be they based on subjective valuations or more objective features, are typically thought to be calculated as the weighted sums of sub-scores across all dimensions (Bettman et al., 1998; Bhatia & Stewart, 2018; Gigerenzer & Gaissmaier, 2011; E. J. Johnson & Payne, 1985; Levav et al., 2010; Payne et al., 1988, 1993; Russo et al., 1996; Shah & Oppenheimer, 2008). Specifically, each option will be assigned a score along each attribute dimension, and each dimension will be given some weight according to how relevant or important it is to the decision. Simplifications of this strategy that assign equal weights to all attributes (Dawes, 1979; Dawes & Corrigan, 1974), reduce attribute scores to binary better/worse rankings (Russo & Dosher, 1983), or only consider a subset of the attributes have been proposed as well (Fishburn, 1974; Tversky, 1972). How well these simpler strategies perform depends the choice context (Bettman et al., 1998; Gigerenzer & Gaissmaier, 2011; E. J. Johnson & Payne, 1985; Levav et al., 2010; Payne et al., 1988, 1993; Russo et al., 1996; Shah & Oppenheimer, 2008). Regardless of precisely how they are combined, almost every choice is determined by an assessment of multiple attributes. Thus, it is important for both basic and applied researchers to better understand how the attribute composition of choice options (and not just their overall values) influences the decision-making process in the brain.

It is known that individuals' preferences over options that combine monetary gains and losses with probabilities or time delays may reverse when different methods are used to elicit those preferences. For example, preferences revealed through choices have been shown to reverse compared to those elicited by matching, pricing, or rating procedures (Alós-Ferrer et al., 2016; Alos-Ferrer et al., 2020, 2021; Fischer et al., 1999; Grether & Plott, 1979; Lichtenstein & Slovic, 1971; Seidl, 2002; Tversky et al., 1988, 1990; Weber & Johnson, 2009). A leading explanation for these preference reversals is that the weights on the risk, time, and/or money dimensions differ across the preference elicitation procedures (Seidl, 2002; Tversky et al., 1988). Eye-tracking experiments have shown that changes in the proportion of visual fixations to a lottery's potential monetary outcome relative to its probability across choice and pricing trials are associated with the differences in the relative weight given to outcomes versus probabilities when choosing versus setting a price (Alos-Ferrer et al., 2021). This influence of visual attention on context dependent weighting is consistent with sequential sampling models that predict that the effects of overall value and attribute differences on choices are determined in part by the amount of attention paid to each option or attribute (Busemeyer & Townsend, 1993; Diederich, 1997; Krajbich et al., 2010; Roe et al., 2001). Together these theories and data form the basis of our hypothesis that decision values in naturalistic multi-attribute choices will also be constructed at the time of choice from the options' basic attributes in a context-dependent manner, rather than being compared as a unitary overall value aggregated across all attributes in a constant fashion.

Consistent with this idea, recent work has shown that the disparity of the options' attribute compositions affects multi-attribute decision making (Lee & Holyoak, 2021). A pair of options has high disparity if, for example, one option scores high in the first attribute dimension but low in the second, while the other option scores high in the second dimension but low in the first

(Figure 1a, left panel). On the contrary, a pair of options has low disparity if both options have similar scores along each attribute dimension (Figure 1a, middle panel). Notably, two decisions could be equally difficult in the traditional sense that the overall value ratings of the choice options are equally close together, yet have very different levels of disparity (Figure 1a, right panel). In multiple independent experiments, Lee & Holyoak (Lee & Holyoak, 2021) found that choice behavior differs as a function of disparity, such that higher disparity corresponds to higher choice consistency (i.e., a choice in favor of the option that was previously rated as having the higher overall value) and lower response time, even after accounting for differences in overall values. Here, we show that computational models of value comparison that assume an immutable combination of attributes into the overall option value cannot account for this pattern of results.

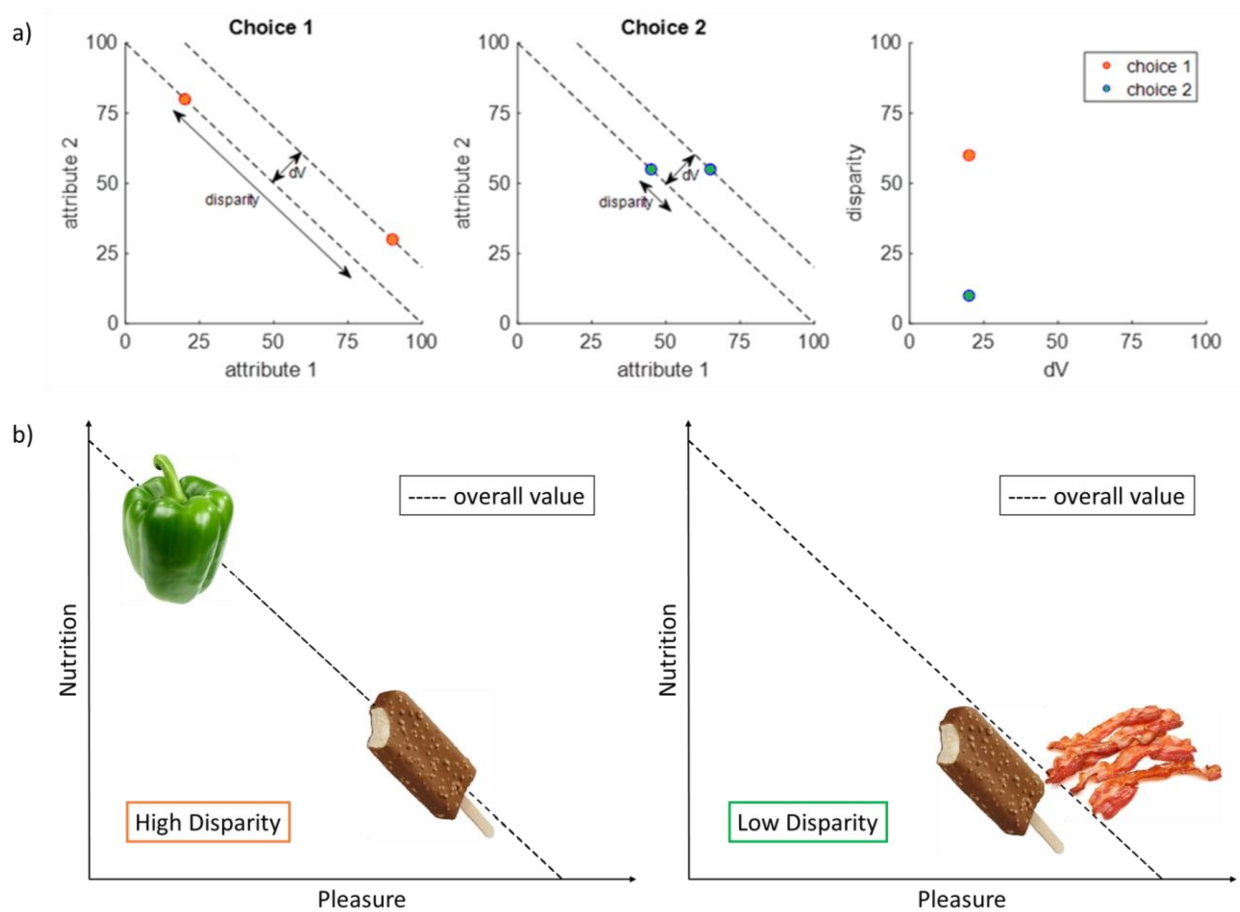


Figure 1. Choice disparity. a) A schematic illustration of orthogonal components of choice difficulty: dV and disparity. *The left plot illustrates a "high disparity" choice, and the middle plot illustrates a "low disparity" choice. The orange and green dots represent the alternative options for each choice, each plotted according to its measurements on two attribute dimensions. The example assumes equal importance weights for each attribute, so the iso-value curves are represented by parallel lines with slope -1. The difference in overall value of the options, dV, is the distance between the iso-value curves on which the options lie. Disparity is the distance between the options in the dimension orthogonal to overall value (see Equation 1 below for a mathematical formulation). The right plot shows the location of each choice pair in the transformed dV-disparity space.* **b)** *An example illustration of two choice sets for snack foods, one high disparity (left plot), one low disparity (right plot). As shown by the dashed iso-value lines, all of the available snacks are of comparable overall value (and thus each choice pair is of comparable low dV). However, the two choice pairs are of very different disparity. In the high disparity pair (left), one option scores high on pleasure but low on nutrition, while the other option scores low on pleasure but high on nutrition. In the low disparity pair (right), both options score high on pleasure and low on nutrition.*

Instead, choices between naturalistic multi-attribute stimuli and subsequent confidence ratings for those choices are best explained by models in which individual attributes are actively (re)weighted during the comparison process. In our tests, we focus on multi-attribute choices between naturalistic, unitary options with multiple inherent, latent features as opposed to bundled or conjoint options made up of multiple components (e.g., probability + amount for risky choice; delay + amount for inter-temporal choice; effort or pain + reward for cost-benefit tradeoffs; different items for bundled choices). We believe that this type of naturalistic reward, which could plausibly be treated as an integrated whole, provides a stronger test of whether items are compared based on fixed overall values or values constructed from flexible attribute weights during decisions. We find that models that allow context-dependent attribute weights during decisions best explain the outcome and response time data. However, we also show that using a subset of attribute-specific ratings together with overall value ratings helps to better explain choice behavior (when obtaining ratings for the full set of individual attributes is impractical).

Methods

Data

We analyzed the data from six previously published experiments (Experiments 1-5 in (Lee & Holyoak, 2021), plus one unpublished pilot experiment that we label Experiment 0). The total number of participants across the six datasets was 307, using the same exclusion criteria from the original study (41 for Experiment 0, 50 for experiment 1, 48 for experiment 2, 54 for experiment 3, 60 for experiment 4, and 54 for experiment 5).

In each experiment, participants completed several distinct phases. They first passively observed images of individual snack foods (100 in Experiments 0 and 1, 60 in Experiments 2-5). Second, they provided overall value ratings for each individual snack food. Next, they rated the pleasure they expected to derive from each food and its nutritional value, in separate experimental sections (the order of the pleasure and nutrition rating tasks was counterbalanced across participants). Following all three rating phases, participants completed a choice task in which they chose their preferred snack from pairs of options (50 choice trials for Experiments 0 and 1, 30 choice trials for Experiments 2-5). During the choice section, after each choice, participants also rated their confidence that the option they chose was indeed their preferred option on that trial.

In addition to the aforementioned datasets, we also examined data from an unpublished pilot auxiliary task that were originally collected along with the primary data reported in (Lee & Daunizeau, 2021). In the main study, participants provided overall value ratings for 148 food options, then made 74 choices between pairs of options. Out of the main group of participants, 17 completed the auxiliary task, in which they rated each of the food options in terms of “taste”, “health”, “texture”, and “appearance”. We are thus able to compare versions of our models (described below) that incorporate four attributes rather than two (the third and fourth attributes enter the models in the same way as the first and second attributes).

Models

In this study, we consider several variants of the drift-diffusion model (DDM; (Ratcliff, 1978; Ratcliff & McKoon, 2008; Ratcliff & Rouder, 1998)). The specifics of each model are detailed below. The DDM is a member of the evidence accumulation-to-bound class of models, wherein the values of the options (in simple two-alternative forced-choice tasks) are repeatedly compared across time. The so-called evidence that arises in favor of one option over the other is corrupted by processing (e.g., neural) noise, so repeated samples are accumulated to cancel out the noise. Once a sufficient amount of evidence has been accrued (i.e., the response threshold is reached), the process terminates and a choice is made. We fit variants of the DDM in which the evidence accumulation rate is proportional to the difference in overall value between the options (ovDDM) or to a weighted linear combination of the differences in individual attributes (maDDM) to six different empirical data sets. We demonstrate that the maDDM provides a better account of choice probabilities, mean response times (RTs), and the effects of attribute disparity on choice outcomes and RT, in terms of both the qualitative predictions and quantitative fits of the models.

Model 1: Overall value DDM (ovDDM)

The first model is a DDM in which only the aggregated, overall values of the two options influence the evidence accumulation or drift rate on each trial. Specifically, evidence about the overall value of each option is sampled at each time step, the evidence for the two options is compared, and the relative evidence in favor of option 1 over option 2 is added to the evidence accumulator. The cumulative evidence (*x*) evolves across deliberation time as follows:

$$x_t = x_{t-1} + \mu_1 - \mu_2 + \varepsilon$$

$$\varepsilon \sim N(0, \sigma^2)$$

$$x_0 = 0$$

where μ_i is the reported overall value of option $i \in \{1, 2\}$ and σ^2 is white noise in the accumulation process. Evidence sampling and accumulation proceeds until *x* reaches a response boundary $\in \{\theta, -\theta\}$, with the sign determining the chosen option (arbitrarily defined as positive for option 1, negative for option 2). Response time (RT) is equal to *t* at the moment a boundary is crossed. Choice probability (*p*, choice of option 1) and mean RT can be analytically derived (Alós-Ferrer, 2018) as a function of μ_1, μ_2, and σ^2, with θ being fixed (here, to $\theta = 1$ for simplicity):

$$DV = \mu_1 - \mu_2$$

$$p = \frac{1}{1 + e^{\left(\frac{-2dDV}{\sigma^2}\right)}}$$

$$RT = \frac{2p - 1}{dDV}$$

where μ_1 and μ_2 are independent variables, and *d* and σ^2 are free parameters to be estimated to capture the individual-specific mean rate of evidence accumulation (drift rate) and level of noise in the accumulation process, respectively.

Model 2: Multi-attribute DDM (maDDM)

The second model is of the same form as Model 1, except that the evidence accumulator is driven by two separate evidence streams (one for each attribute dimension: a, b). The process is otherwise identical, and it unfolds as follows:

$$x_t = x_{t-1} + \mu_1^a - \mu_2^a + \mu_1^b - \mu_2^b + \varepsilon$$

$$\varepsilon \sim N(0, \sigma^2)$$

$$x_0 = 0$$

where μ_i^j is the reported value of option i ∈ {1,2} along attribute dimension j ∈ {a,b}, and σ^2 is white noise common to the overall evidence accumulation process. Choice probability and mean RT are derived as:

$$DV^a = \mu_1^a - \mu_2^a$$

$$DV^b = \mu_1^b - \mu_2^b$$

$$p = \frac{1}{1 + e^{\left[\frac{-2(d^a DV^a + d^b DV^b)}{\sigma^2}\right]}}$$

$$RT = \frac{2p - 1}{d^a DV^a + d^b DV^b}$$

where d^j are independent free parameters that allow for different rates of evidence accumulation within each attribute dimension j∈ {a,b}.

Model 3: Multi-attribute DDM plus overall value (maDDM+)

Our third model assumes that the drift rate is driven by the separate values of the individual attributes that were explicitly evaluated (in these experiments, pleasure and nutrition), but that it is also influenced by other attributes that were not explicitly evaluated. Thus, if the overall value ratings contain information about the attributes that were rated individually as well as other attributes that were not rated, including overall value should enhance the model fit. Therefore, in this model, the evidence accumulator is driven by evidence streams for each explicit attribute dimension as well as the aggregate overall value estimates. The process is otherwise identical to that in Models 1 and 2, and it unfolds as follows:

$$x_t = x_{t-1} + \mu_1^o - \mu_2^o + \mu_1^a - \mu_2^a + \mu_1^b - \mu_2^b + \varepsilon$$

$$\varepsilon \sim N(0, \sigma^2)$$

$$x_0 = 0$$

where μ_i^o is the reported overall value for option $i \in \{1, 2\}$, μ_i^j is the reported value of option i ∈ {1,2} along attribute dimension j ∈ {a,b}, and σ^2 is white noise common to the overall evidence accumulation process. (Note that in cases where the individual attributes are highly correlated and/or the attribute ratings jointly explain a large portion of the variance in overall values, it may be necessary to employ orthogonalization or dimensionality reduction techniques, if the goal is to make inferences about the relative weights or importance of attributes in determining the drift rate.) Choice probability and mean RT are derived as:

$$DV^o = \mu_1^o - \mu_2^o$$

$$DV^a = \mu_1^a - \mu_2^a$$

$$DV^b = \mu_1^b - \mu_2^b$$

$$p = \frac{1}{1 + e^{\left[\frac{-2(d^o DV^o + d^a DV^a + d^b DV^b)}{\sigma^2}\right]}}$$

$$RT = \frac{2p - 1}{\mathrm{d}^o DV^o + d^a DV^a + d^b DV^b}$$

where d^j are independent free parameters that allow for different rates of evidence accumulation for overall value estimates and within each attribute dimension j∈ {a,b}.

Drift diffusion model fitting procedure

We fit each of the three candidate models to the experimental data from each of our models under consideration. We then performed Bayesian model comparison to determine which of the models (if any) performed better than the others across the population of participants. For this model fitting and comparison exercise, we relied on the Variational Bayesian Analysis toolbox (VBA, available freely at https://mbb-team.github.io/VBA-toolbox/; (Daunizeau et al., 2014)) with Matlab R2020a. Within participant and across trials, we entered the experimental variables {ratings of overall, pleasure, and nutrition for each option} as input and {choice = 1 for left option, 0 for right option; RT} as output. We also provided the model-specific mappings from input to output as outlined in the analytical formulas above. As we fixed the threshold parameter θ to 1, the parameters to be fitted were thus the drift rate d and diffusion noise σ^2 terms described above in the model formulations. VBA requires prior estimates for the free parameters, for which we set the mean equal to 0 and the variance equal to e (to allow an arbitrarily large step size during the gradient descent search algorithm, yet constrain the algorithm to a reasonable search space) for each parameter. The theoretical drift rate and noise parameters are always positive; we thus constrained the search space of our model fitting algorithm to the positive domain. VBA then recovers an approximation to both the posterior density on unknown variables and the model evidence (which is used for model comparison). We used the VBA_NLStateSpaceModel function to fit the data for each participant individually, followed by the VBA_groupBMC function to compare the results of the model fitting across models for the full group of participants.

One benefit of using VBA to fit the data to our models is that it is computationally efficient, as it relies on Variational Bayesian analysis under the Laplace approximation. This iterative algorithm provides a free-energy approximation for the model evidence, which represents a natural trade-off between model accuracy (goodness of fit, or log likelihood) and complexity (degrees of freedom, or KL divergence between priors and fitted parameter estimates; see (Friston et al., 2007; Penny, 2012)). Additionally, the algorithm provides an estimate of the posterior density over the model's free parameters, starting with Gaussian priors. Individual log model evidence scores are then provided as input to the group-level random-effect Bayesian model selection (BMS) procedure. BMS provides an exceedance probability that measures how likely it is that a given model is more frequently implemented, relative to all other models under consideration, in the population from which participants were drawn (Rigoux et al., 2014; Stephan et al., 2009). This approach to fitting and comparing variants of DDM has already been successfully demonstrated in previous studies (Feltgen & Daunizeau, 2021; Lee & Usher, 2021; Lopez-Persem et al., 2016).

Logistic regression for changes in attribute weights

We fit Bayesian hierarchical regressions to the binary choice outcomes and the implied choice outcomes from the overall value rating sessions using the R package, brms, which relies on STAN for Markov Chain Monte Carlo sampling of the posterior distributions (Bürkner, 2017a, 2017b; R Core Team, 2020; Stan Development Team, 2021). Both regressions included varying intercepts and slopes for each participant, however we omit these terms in the notation below for conciseness and clarity.

$$\text{choice} \sim \beta_0 + \beta_1 * P_{dif} + \beta_2 * N_{dif} + \sigma$$

Here, choice indicates whether the participant selected the left item in the binary choice task or if the left item was rated higher in the overall value rating session. The differences in pleasure (P_{dif}) and nutrition (N_{dif}) ratings are computed as the rating for the left option minus the rating for the right option for the respective attributes. We used Gaussian priors with mean = 0 and SD = 1 for $\beta_0 : \beta_3$ and half Cauchy priors with location = 0 and scale = 5 for the standard deviations of all participant-specific varying effects. We estimated the posterior distributions for all parameters based on 2500 samples from 4 independent chains after 2500 warm-up samples.

Linear regressions for confidence ratings

We fit linear regressions to the choice confidence ratings, assuming that those ratings were right-censored given the high proportion of trials in which participants reported maximum confidence. In other words, we assumed that the true underlying confidence could extend beyond the upper limit of the confidence scale. We fit three right-censored linear regressions as Bayesian hierarchical models using the R package, brms, which relies on STAN for Markov Chain Monte Carlo sampling of the posterior distributions (Bürkner, 2017a, 2017b; R Core Team, 2020; Stan Development Team, 2021). All three regressions included varying intercepts and slopes for each participant, however we omit these terms in the notation below for conciseness and clarity.

$$(LR1) \quad confidence \sim \beta_0 + \beta_1 * OV_{dif} + \sigma$$

$$(LR2) \quad confidence \sim \beta_0 + \beta_1 * P_{dif} + \beta_2 * N_{dif} + \sigma$$

$$(LR3) \quad confidence \sim \beta_0 + \beta_1 * P_{dif} + \beta_2 * N_{dif} + \beta_3 * OV_{dif} + \sigma$$

In all regressions, the differences in overall value (OV_{dif}), pleasure ratings (P_{dif}), or nutrition ratings (N_{dif}) were computed as the absolute value of the difference between ratings for the option displayed on the left and right sides of the screen. We used Gaussian priors with mean = 0 and SD = 1 for $\beta_0 : \beta_3$ and half Cauchy priors with location = 0 and scale = 5 for the standard deviations of all participant-specific varying effects. We estimated the posterior distributions for all parameters based on 2000 samples from 3 independent chains after 2000 warm-up samples.

Results

Lee and Holyoak (Lee & Holyoak, 2021) introduced the term *disparity* as a secondary characteristic of a decision, orthogonal to the primary characteristic most often used in the field:

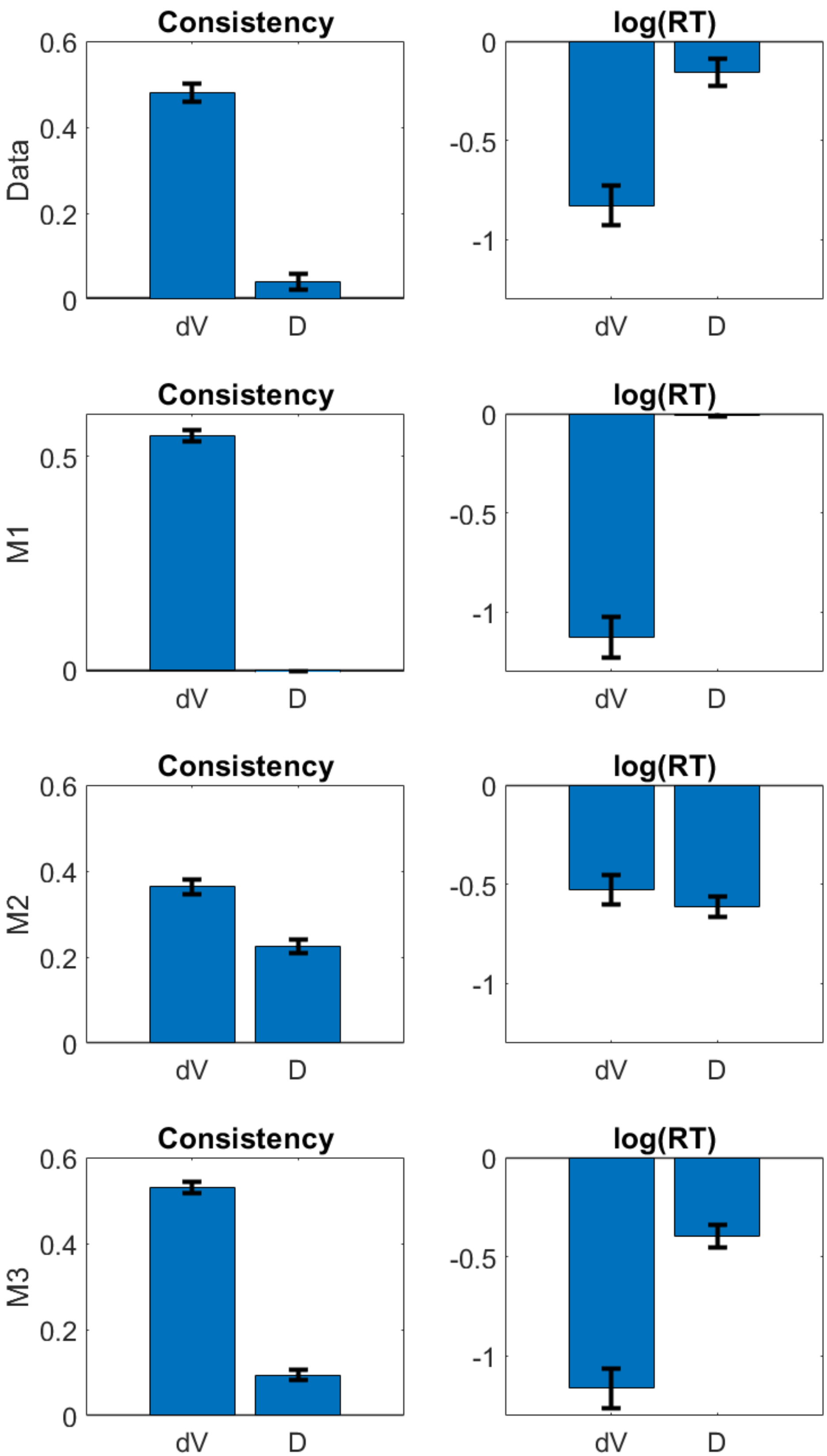
overall value difference. The disparity between two options {i,j} with respect to two attribute dimensions (here, P for pleasure and N for nutrition) is calculated as:

$$disparity_{i,j} \triangleq \left| \frac{[P_i\ N_i] * \begin{bmatrix} -b_P \\ b_N \end{bmatrix} - [P_j\ N_j] * \begin{bmatrix} -b_P \\ b_N \end{bmatrix}}{\left\| \begin{matrix} -b_P \\ b_N \end{matrix} \right\|} \right|, \text{for options } i, j. \quad (1)$$

where b_P and b_N are the importance weights for the P and N attributes, respectively. Note that it would be straightforward to extend this formula to incorporate more than two options or more than two attributes. Disparity calculated in this way effectively transforms the variable space from attribute dimensions to decision dimensions (see Figure 1).

Qualitative Model Predictions

We first show the qualitative predictions that each model (simulated under its participant-specific best fitting parameters) makes with respect to the effects of value difference (dV) and disparity (D) on choice consistency, RT, and how this compares to the empirical data (Figure 2). For the synthetic data, we used the real input data for each individual participant (value of option 1, or v1; value of option 2, or v2; pleasure of option 1, or p1; pleasure of option 2, or p2; nutrition of option 1, or n1; nutrition of option 2, or n2). The output data we used (choice probability and RT) was the output from the VBA model fitting procedure (described in the Methods section), where these variables were predicted using the best fitting parameter estimates for each participant under each model. Next, we performed mixed model regressions of choice (binomial) and of RT (linear) on dV and D, pooling all simulations together and including study and participant as random effect regressors. (Note: we coded the data such that option 1 always had the higher overall value.) Model 3 is the only model from the set we examined that is able to account for the qualitative benchmarks with respect to the positive impact of dV and D on choice consistency (whether or not the higher-rated option was chosen) and the negative impact of dV and D on RT (see Figure 2). Model 1 accounts for the dV relationships, but obviously cannot account for any relationship involving D (as this model contains no information about the attribute composition of the options). Model 2 does account for all of the qualitative benchmarks, but it overemphasizes the impact of D (relative to dV) on both consistency and RT (in comparison to the empirical data). It thus seemed probable that Model 3 would perform best in a formal model comparison of these three models.

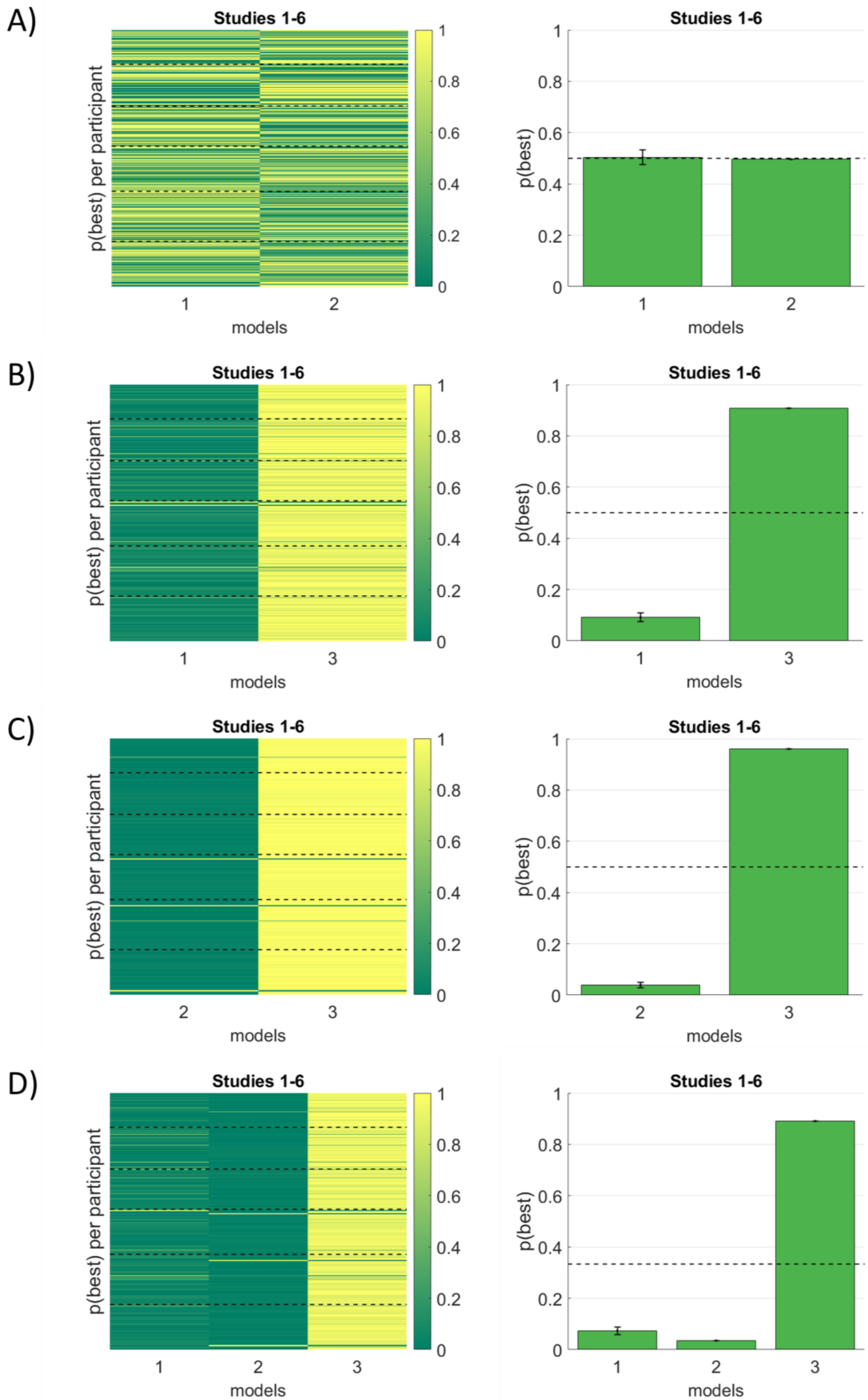


326

Figure 2: *Qualitative predictions of the effects of value difference (dV = value of option 1 – value of option 2) and disparity (D; see equation 1) on choice consistency and log(RT) in the empirical (top row) and simulated data (rows 2-4; shown for responses simulated using the best fitting parameters for each model; bar heights represent mean mixed model regression coefficients across participants; error barts represent s.e.m.).*

Quantitative model comparisons

To evaluate the models, we computed specific pairwise model comparisons as well as a simultaneous comparison of all three diffusion decision model specifications. First, we compared the fit of the model based on overall value (Model 1) to a model based on individual attribute values (Model 2), to test our hypothesis that decision makers place different importance or decision weights on attributes during choices than when rating the overall value of each option in isolation. If the attributes are weighted and aggregated the same way during ratings and choices, then Model 1, which uses the comprehensive overall value of the foods to estimate choices should outperform Model 2, which only uses a subset of the foods' attributes (pleasure and nutrition) to estimate choice outcomes and response times. If instead attributes are weighted differently during choices compared to ratings, then the flexibility to estimate attribute-specific decision weights may give Model 2 the advantage even though it is based on only two attributes out of a larger set. Across all six datasets (studies 0-5 from (Lee & Holyoak, 2021)), the participant population was split evenly between supporting Models 1 and 2 (estimated model frequency of 0.50 for each; see Figure 3A).


A)
Studies 1-6
p(best) per participant
models
1
2
p(best)
B)
Studies 1-6
models
1
3
C)
Studies 1-6
models
2
3
D)
Studies 1-6
models
1
2
3

347

Figure 3: *Pairwise model comparison results: A) ovDDM (model 1) versus maDDM (model 2). B) ovDDM (model 1) versus maDDM+ (model 3). C) maDDM (model 2) versus maDDM+ (model 3). D) Simultaneous comparison of ovDDM (model 1), maDDM (model 2), and maDDM+ (model 3). We show here the probability that each model best accounted for the data at the participant level (left panel), across the six studies we examined; each cell represents the probability that the model (column) best represents the behavior of the participant (row). The black dashed lines serve to indicate which participants belonged to each of the six datasets. We also show the probability that each model best explains the data across the participant population (right panel), across all studies. Here, the black dashed line indicates chance level if all models were equally probable a priori.*

Next, we tested a model in which we tried to combine the advantages of the comprehensive overall value ratings together with the flexibility to estimate choice-specific weights for a subset of the individual attributes. We fit a third DDM model (Model 3) that used the pleasure and nutrition ratings as well as information about the foods' values beyond those two attributes, in the form of the reported overall value. This provides a measure to quantify the benefit of including individual attribute ratings in the model in addition to overall value ratings. Model 3 performed much better than Model 1. Across all six datasets, Model 3 had an exceedance probability of 1 and an estimated model frequency of 0.91 (see Figure 3B).

We also tested Model 3 against Model 2, to quantify the benefit of including overall value ratings in the model in addition to individual attribute ratings (pleasure and nutrition). Model 3 performed much better than Model 2. Across all six datasets, Model 3 had an exceedance probability of 1 and an estimated model frequency of 0.96 (see Figure 3C).

When we formally compared all three models simultaneously, the results were consistent with the pairwise comparisons reported in the preceding paragraphs. Across all datasets, the winning model was Model 3, with an exceedance probability of 1 and an estimated model frequency of 0.89. Model 1 had an estimated frequency of 0.07, and Model 2 had an estimated frequency of 0.04 (see Figure 3D).

The last model comparison analysis that we performed examined the dataset for which we had overall value ratings as well as four individual attribute ratings (taste, health, texture, appearance). Here, the model in which the attributes each individually contributed (Model 2) outperformed the standard DDM based on overall value ratings alone (Model 1; but note the small sample size, n=17). In the other model comparisons, the results were the same as with our main datasets. Even though this additional dataset had attribute ratings covering a larger range of potentially relevant attributes (4 versus 2), Model 3 still dominated the model comparison (exceedance probability = 1; estimated model frequency for Model 3 = 0.96, for Models 1 and 2 = 0.02 each; see Figure 4).

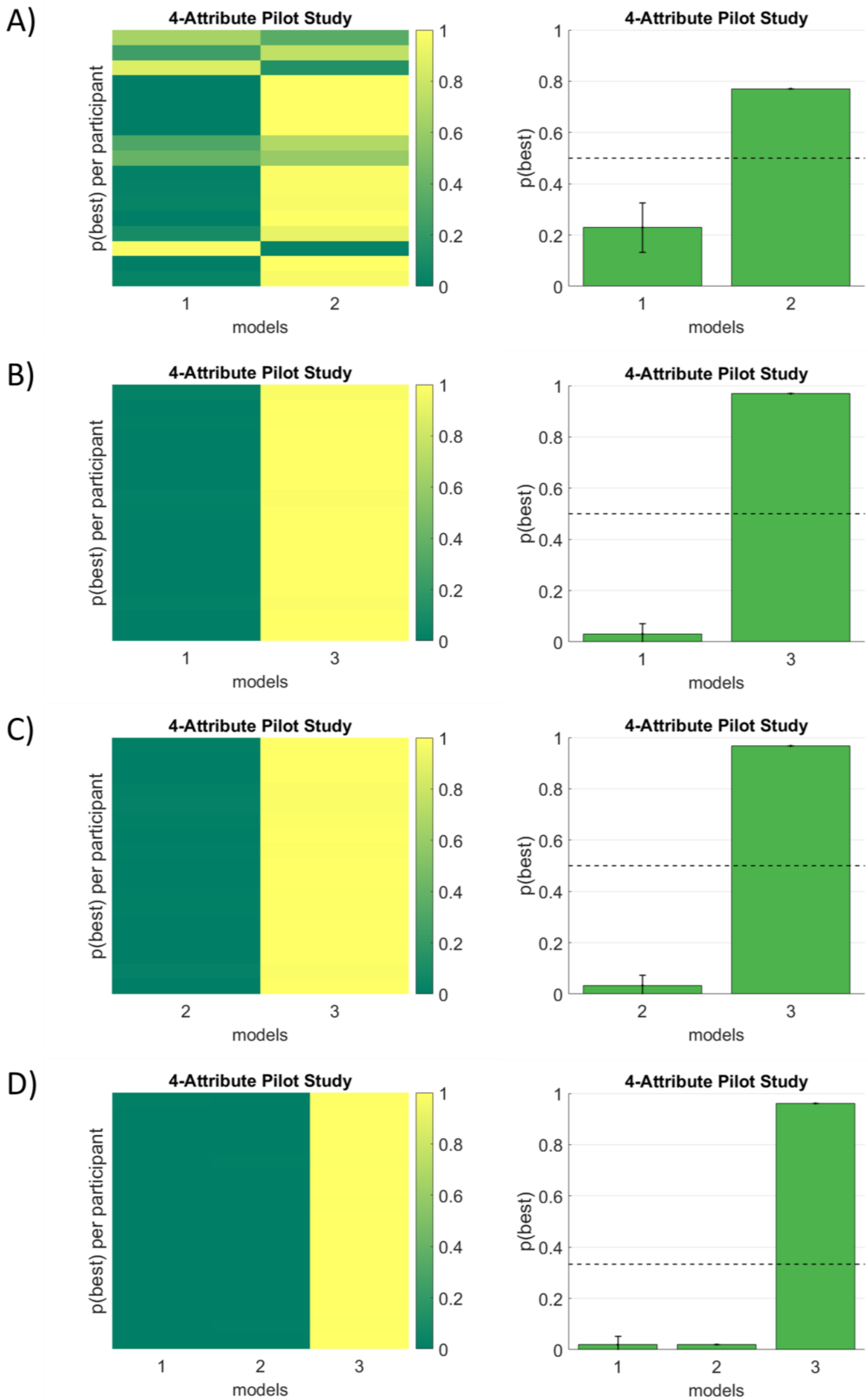

A)
4-Attribute Pilot Study
p(best) per participant
models
4-Attribute Pilot Study
p(best)
models
B)
4-Attribute Pilot Study
p(best) per participant
models
4-Attribute Pilot Study
p(best)
models
C)
4-Attribute Pilot Study
p(best) per participant
models
4-Attribute Pilot Study
p(best)
models
D)
4-Attribute Pilot Study
p(best) per participant
models
4-Attribute Pilot Study
p(best)
models

Figure 4: *Model comparison results for four-attribute choice data: A) ovDDM (model 1) versus maDDM (model 2). B) ovDDM (model 1) versus maDDM+ (model 3). C) maDDM (model 2) versus maDDM+ (model 3). D) Simultaneous comparison of ovDDM (model 1), maDDM (model 2), and maDDM+ (model 3). We show here the probability that each model best accounted for the data at the participant level (left panel), across the six studies we examined; each cell represents the probability that the model (column) best represents the behavior of the participant (row). We also show the probability that each model best explains the data across the participant population (right panel), across all studies. The black dashed line indicates chance level if all models were equally probable a priori.*

Model recoverability

To verify that our model-fitting procedure is suitable for this specific analysis, we performed a test of model recoverability. Specifically, we took as the model input the actual data for each participant in Studies 1-6 (ratings of overall value, pleasure, and nutrition for each option). We then simulated the set of choice probabilities and mean RTs for each participant, separately according to each of our models, using the actual participant-specific fitted parameters for each model. Finally, we fit all simulated data (per simulated participant) to each of our models and performed the same formal model comparison as with our real experimental data. The results of this procedure can be seen in Figure 5 as a model confusion matrix. This matrix shows, for each true generative model, the percentage of simulated participants (under that model) that were attributed to each of the best fit models by our model-fitting procedure. As shown in the matrix, model confusion was low and the procedure attributed the true model as the best fitting model for the vast majority of the simulated participants (recovery accuracy: *87% for Model 1, 66% for Model 2, 69% for Model 3*). There was a non-trivial amount of confusion between Model 3 and Models 1 and 2, which is expected because Model 3 is essentially a combination of Models 1 and 2. Nevertheless, this confusion seems to have hurt Model 3 as much as it helped (i.e., the matrix is relatively symmetrical), so the results of our quantitative model comparison should be valid.

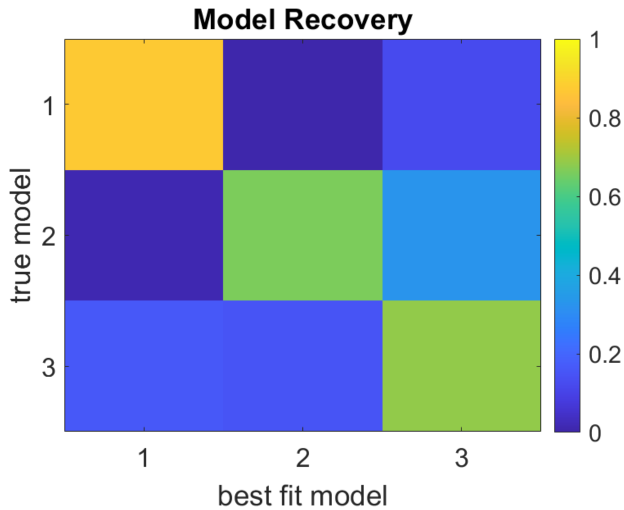


Figure 5: *Model recovery analysis. Each cell in the "confusion matrix" summarizes the percentage of simulated participants (under each true model) for which our model-fitting procedure attributed each of the (best fit) models. Note that our fitting procedure attributed the true model as the best fitting model for 87%, 66%, 69% of the time for Models 1-3, respectively.*

Changes in attribute weights

The results above show that the relative importance or weight of the pleasure and nutrition attributes change between the overall value rating and choice tasks even though, in theory, the same valuation processes should be employed in both cases. Other attribute weights may change as well, but we focus on pleasure and nutrition because we have the most data on those two attributes. Participants in these studies typically chose the more pleasurable of the two food options in the choice task. Therefore, we used hierarchical Bayesian logistic regression analyses to determine if pleasure weights reliably increased relative to nutrition weights across participants, or if instead participants were equally likely to show either an increase or a decrease in the weighting of pleasure relative to nutrition. We determined implied choices from the overall value ratings, classifying the food item with the higher rating as the chosen option. Pooling the data across all six experiments, we found that in choices compared to ratings, there was a consistent increase in the influence of pleasure (mean change in logistic regression coefficient = 0.57, 95%

highest density interval (HDI) = [-0.15, 1.31]) and a consistent decrease in the influence of nutrition (mean change in logistic regression coefficient = -0.25, 95% HDI = [-0.64, 0.14]). The mean difference in differences (i.e., interaction) between pleasure and nutrition weights across choice and rating sessions was 0.82 (95% HDI = [-0.10, 1.70]), and the posterior probability of this change being greater than zero was 0.96. Thus, on average, participants' choices were more influenced by pleasure than nutrition, relative to their overall value ratings. The difference in the influence of the attributes between the different valuation tasks is one reason why the DDMs using attribute-level information explain choices better than the DDM based on overall value ratings alone.

Choice Confidence

The datasets we examined (and many others in the literature) show clear and robust relationships between the differences in the overall value ratings of options within each pair and choice consistency, RT, and confidence. We also found the expected negative relationship between RT and confidence in our data sets (cross-participant mean correlation = -0.35, s.e.m. = 0.01). The DDM analyses demonstrated that models including separate ratings of individual attributes explain choice consistency and RT better than models based on an aggregated overall value ratings alone. Standard DDM formulations do not directly predict choice confidence. However, any valid model of choice confidence should be able to account for the robust empirical finding that value difference has a positive impact on confidence. It would be interesting to know whether the differences in individual attribute ratings also have such an impact on confidence. To that end, we tested whether the overall value ratings or the independent ratings for each attribute, or their combination (i.e., the independent variable inputs to the three DDM variants we examined) better explained the reported levels of choice confidence. We used a right-censored linear regression to fit the confidence data in all cases, because the participants' confidence ratings were clearly bounded by the upper limit of the reporting scale. In line with our DDM results on RTs and choice outcomes, the regression model for confidence based on both overall value and individual attributes was best (Table 1).

Table 1. Model comparison of linear regressions for confidence ratings

	Overall value	Pleasure + Nutrition	Pleasure + Nutrition + Overall value
LOOIC	2066.2 ± 190.2	2750.8 ± 192.6	1675.4 ± 192.9

This table reports the mean ± SD of the LOO information criteria computed using Pareto smoothed importance sampling (Vehtari et al., 2017) for the three linear regressions we compared. Lower LOOIC values indicate better expected out-of-sample predictions.

Discussion

The construction of a choice option's subjective value is an active, malleable process. We have shown that when options are composed of multiple distinct attributes, the manner in which these attributes are evaluated and potentially combined to determine the overall value of each option relative to the other depends on the valuation context. Specifically, the contributions of inherent attributes such as pleasure and nutrition to the overall value of food rewards differs when the foods are evaluated in isolation compared to when choices are made between pairs of foods, even though the goal of the valuation process should be the same in both cases. These findings indicate that preferences over naturalistic multi-attribute goods are sensitive to the details of the methods used to elicit them.

Most options that humans need to evaluate and choose from in daily life have multiple attributes or dimensions, and we found that these attributes are combined in a context-dependent manner to determine an option's value. Previous work on naturalistic multi-attribute decisions has shown that people form option representations based on a large number of separate underlying attributes (Bhatia & Stewart, 2018). Our results on both choice outcomes and post-decision confidence ratings are consistent with these findings. In addition, our model simulation and comparison results show that the representations that determine an option's subjective value differ between rating and binary choice tasks. From the simulation tests, it is clear that decision models based on a unitary overall value do not generate the influence of disparity (Lee & Holyoak, 2021) on choices or response times seen in the empirical data. In terms of quantitative model comparisons, if the overall values of the foods in these six studies were invariant across the ratings and choice tasks, then DDMs using only the overall values to determine the drift rate (Model 1) would be preferred over DDMs using either a subset of the food attributes (i.e., Model 2 with pleasure and nutrition) or individual attributes plus the overall values (Model 3). This is because relative to Model 1, Model 2 adds complexity (one additional parameter) while at the same time reducing the completeness of the information about the food items – assuming overall value is determined by more than just pleasure and nutrition. Instead, the model comparisons showed a tie between Models 1 and 2 in our primary data sets. However, we also report preliminary evidence from a dataset with ratings of four separate attributes that multi-attribute DDM specifications such as Model 2 do far better when they include ratings for more than two individual attributes. Moreover, if overall value representations were constant, then Model 3 would add redundant complexity compared to Model 1 and be penalized for that complexity without benefiting from greater explanatory power in the comparisons. In fact, Model 3 (overall value + pleasure + nutrition) is the best in terms of generating the observed effects of attribute disparity and accounting for the pattern of choice outcomes and response times.

The superior ability of DDM and regression models including attribute-level information to explain the effects of disparity on choices and RTs, as well as in explaining choice confidence, indicates that the importance of one or more attributes reliably differs during binary choices relative to ratings. Specifications of the DDM that use reports of overall value as the input to the evidence accumulation process implicitly hold the relative importance of each attribute fixed, and thus cannot account for differences in value computation between ratings of single options and choices over two or more options. In contrast, a multi-attribute DDM specification will directly estimate the importance weights for each attribute within the choice context and is therefore better able to explain choice behavior. However, these models are agnostic about how or why the importance of specific attributes differs when individuals are computing the overall value of a single option compared to choosing between two options.

Many sequential sampling models of decision making posit that attention and salience play an important role in value computation and comparison. Examples of such models are Decision Field Theory (DFT; Townsend & Busemeyer, 1993), an extension of DFT known as the Multi-Attribute Dynamic Decision (MADD) model (Diederich, 1997), and the attentional DDM (aDDM; Smith & Krajbich, 2019; Krajbich, Armel, & Rangel, 2010). In the DFT model, the drift rate can vary across deliberation time if, for example, one option is more salient but the other is truly more valuable. The MADD model makes the multi-attribute nature explicit, and the drift rate fluctuates over time as the decision maker shifts focus across the set of relevant attributes. Although the aDDM has generally been applied to the overall values of options, or to distinct items within a bundle (Fisher, 2017, 2021), it would be conceptually similar to the MADD if applied at the attribute level for goods that are inherently multidimensional. This also has some similarities with query theory, which holds that the order in which a decision maker considers different aspects (or attributes) of an option alters its resultant valuation (E. Johnson et al., 2007; Weber et al., 2007). In all these models, it is assumed that options or attributes that receive more attention will be favored during the value comparison process. However, the relationship between value and attention is most likely bidirectional (Anderson et al., 2011; Callaway et al., 2021; Gluth et al., 2018; Jang et al., 2021; Towal et al., 2013).

Differences in the amount of attention directed to specific attributes during the evaluation and decision contexts could explain changes in the relative importance of the attributes across those contexts. Consistent with this idea, changes in the proportions of visual fixations to locations on a computer screen indicating the monetary amount versus the probability of winning when pricing versus choosing between lotteries are associated with inconsistencies between the two preference elicitation contexts (Alos-Ferrer et al., 2021; Kim et al., 2012). Fixation patterns towards monetary amount versus delay affect temporal discounting rates (Fisher, 2021a), although it has not yet been tested whether this might vary in pricing versus choice contexts. Internal attention processes may have effects similar to visual attention for naturalistic multi-attribute goods. The focus of internal attention is more difficult to measure than visual attention (i.e., fixation locations), but studies combining decision tasks with neuroimaging or electrophysiology and machine learning techniques may give us a window into these cognitive processes (Aoi et al., 2020; Peixoto et al., 2021). Experimental manipulation of focus on a specific attribute (Fisher, 2018; Hare et al., 2011) may also prove useful, if an appropriate method to dynamically shift attention within each trial is developed.

There may be other unknown mechanisms beyond differences in attention allocation that led to changes in the importance weights given to attributes during rating compared to choice tasks. Although both the rating and choice processes are noisy to some extent, noise is an unlikely explanation of our results. Unbiased noise in the two tasks could not account for the consistent increases in the weight on pleasure relative to nutrition during the choice versus rating tasks.

However, it is possible that changes in motivation, engagement, or perceptions of the food items and/or task goals may have differed in the overall value rating compared to binary choice sessions instead of or in addition to any change in attribute-level attention. The current data do not allow us to examine these alternative mechanisms in detail, and it will be important to address them in future studies.

In addition to providing further insight into the mechanistic nature of value-based decisions, our current work has practical implications for future studies of decision-making. We have shown that it is best to use as much attribute-level information as possible when modeling decisions over multi-attribute stimuli. Most, if not all, naturalistic stimuli are composed of multiple attributes, thus most studies of decision-making should incorporate attribute-level information. At the same time, it will often be impractical to collect information on a large number of attributes, especially if one needs subjective opinions about the attributes from each participant in an experiment. Our results indicate that combining attribute-specific and overall values may be a good compromise between attempting to include comprehensive attribute-level information and conforming to practical constraints. Naturally, which attribute-level information to obtain and how to best combine it with some type of overall value rating will depend on the hypotheses and experimental design. Given the clear evidence that the value-comparison process is based on context-dependent attribute weights, experiments that use a well-designed combination of attribute-specific and aggregate-level information should prove to be the most useful in advancing our understanding of many important decision mechanisms.

References

Alós-Ferrer, C. (2018). A Dual-Process Diffusion Model. *Journal of Behavioral Decision Making*, *31*(2), 203–218. https://doi.org/10.1002/bdm.1960

Alos-Ferrer, C., Buckenmaier, J., & Garagnani, M. (2020). Stochastic Choice and Preference Reversals. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.3748599

Alós-Ferrer, C., Granić, Đ.-G., Kern, J., & Wagner, A. K. (2016). Preference reversals: Time and again. *Journal of Risk and Uncertainty*, *52*(1), 65–97. https://doi.org/10.1007/s11166-016-9233-z

Alos-Ferrer, C., Jaudas, A., & Ritschel, A. (2021). Attentional Shifts and Preference Reversals: An Eye-Tracking Study. *Judgment and Decision Making*, *16*(1), 57–93. https://doi.org/10.5167/uzh-198757

Anderson, B. A., Laurent, P. A., & Yantis, S. (2011). Value-Driven Attentional Capture. *Proceedings of the National Academy of Sciences*, *108*(25), 10367–10371. https://doi.org/10.1073/pnas.1104047108

Aoi, M. C., Mante, V., & Pillow, J. W. (2020). Prefrontal cortex exhibits multidimensional dynamic encoding during decision-making. *Nature Neuroscience*, *23*(11), 1410–1420. https://doi.org/10.1038/s41593-020-0696-5

Bettman, J. R., Luce, M. F., & Payne, J. W. (1998). Constructive Consumer Choice Processes. *Journal of Consumer Research*, *25*(3), 187–217. https://doi.org/10.1086/209535

Bhatia, S., & Stewart, N. (2018). Naturalistic Multiattribute Choice. *Cognition*, *179*, 71–88. https://doi.org/10.1016/j.cognition.2018.05.025

Bürkner, P.-C. (2017a). brms: An R Package for Bayesian Multilevel Models Using Stan. *Journal of Statistical Software*, *80*(1), 1–28. https://doi.org/10.18637/jss.v080.i01

Bürkner, P.-C. (2017b). Advanced Bayesian Multilevel Modeling with the R Package brms. *ArXiv:1705.11123 [Stat]*. http://arxiv.org/abs/1705.11123

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. https://doi.org/10.1037/0033-295X.100.3.432

Callaway, F., Rangel, A., & Griffiths, T. L. (2021). Fixation Patterns in Simple Choice Reflect Optimal Information Sampling. *PLOS Computational Biology*, *17*(3), e1008863. https://doi.org/10.1371/journal.pcbi.1008863

Daunizeau, J., Adam, V., & Rigoux, L. (2014). VBA: A Probabilistic Treatment of Nonlinear Models for Neurobiological and Behavioural Data. *PLOS Computational Biology*, *10*(1), e1003441. https://doi.org/10.1371/journal.pcbi.1003441

Dawes, R. M. (1979). The robust beauty of improper linear models in decision making. *American Psychologist*, *34*(7), 571–582. https://doi.org/10.1037/0003-066X.34.7.571

Dawes, R. M., & Corrigan, B. (1974). Linear models in decision making. *Psychological Bulletin*, *81*(2), 95–106. https://doi.org/10.1037/h0037613

Diederich, A. (1997). Dynamic Stochastic Models for Decision Making under Time Constraints. *Journal of Mathematical Psychology*, *41*(3), 260–274. https://doi.org/10.1006/jmps.1997.1167

Feltgen, Q., & Daunizeau, J. (2021). An Overcomplete Approach to Fitting Drift-Diffusion Decision Models to Trial-By-Trial Data. *Frontiers in Artificial Intelligence*, *0*. https://doi.org/10.3389/frai.2021.531316

Fischer, G. W., Carmon, Z., Ariely, D., & Zauberman, G. (1999). Goal-Based Construction of Preferences: Task Goals and the Prominence Effect. *Management Science*, *45*(8), 1057–1075. https://doi.org/10.1287/mnsc.45.8.1057

Fishburn, P. C. (1974). Exceptional Paper—Lexicographic Orders, Utilities and Decision Rules: A Survey. *Management Science*, *20*(11), 1442–1471. https://doi.org/10.1287/mnsc.20.11.1442

Fisher, G. (2017). An Attentional Drift Diffusion Model over Binary-Attribute Choice. *Cognition*, *168*, 34–45. https://doi.org/10.1016/j.cognition.2017.06.007

Fisher, G. (2018). Nutrition labeling reduces valuations of food through multiple health and taste channels. *Appetite*, *120*, 500–504. https://doi.org/10.1016/j.appet.2017.09.013

Fisher, G. (2021a). Intertemporal Choices Are Causally Influenced by Fluctuations in Visual Attention. *Management Science*. https://doi.org/10.1287/mnsc.2020.3732

Fisher, G. (2021b). A Multiattribute Attentional Drift Diffusion Model. *Organizational Behavior and Human Decision Processes*, *165*, 167–182. https://doi.org/10.1016/j.obhdp.2021.04.004

Friston, K., Mattout, J., Trujillo-Barreto, N., Ashburner, J., & Penny, W. (2007). Variational free energy and the Laplace approximation. *NeuroImage*, *34*(1), 220–234. https://doi.org/10.1016/j.neuroimage.2006.08.035

Gigerenzer, G., & Gaissmaier, W. (2011). Heuristic Decision Making. *Annual Review of Psychology*, *62*(1), 451–482. https://doi.org/10.1146/annurev-psych-120709-145346

Gluth, S., Spektor, M. S., & Rieskamp, J. (2018). Value-Based Attentional Capture Affects Multi-Alternative Decision Making. *ELife*, *7*, e39659. https://doi.org/10.7554/eLife.39659

Grether, D., & Plott, C. (1979). Economic Theory of Choice and the Preference Reversal Effect. *American Economic Review*, *69*, 623–638.

Hare, T. A., Malmaud, J., & Rangel, A. (2011). Focusing Attention on the Health Aspects of Foods Changes Value Signals in vmPFC and Improves Dietary Choice. *Journal of Neuroscience*, *31*(30), 11077–11087. https://doi.org/10.1523/JNEUROSCI.6383-10.2011

Jang, A. I., Sharma, R., & Drugowitsch, J. (2021). Optimal Policy for Attention-Modulated Decisions Explains Human Fixation Behavior. *ELife*, *10*, e63436. https://doi.org/10.7554/eLife.63436

Johnson, E., Häubl, G., & Keinan, A. (2007). Aspects of Endowment: A Query Theory of Value Construction. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, *33*, 461–474. https://doi.org/10.1037/0278-7393.33.3.461

Johnson, E. J., & Payne, J. W. (1985). Effort and Accuracy in Choice. *Management Science*, *31*(4), 395–414. https://doi.org/10.1287/mnsc.31.4.395

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

Kim, B., Seligman, D., & Kable, J. W. (2012). Preference Reversals in Decision Making Under Risk are Accompanied by Changes in Attention to Different Attributes. *Frontiers in Neuroscience*, *0*. https://doi.org/10.3389/fnins.2012.00109

Krajbich, I., Armel, C., & Rangel, A. (2010). Visual Fixations and the Computation and Comparison of Value in Simple Choice. *Nature Neuroscience*, *13*(10), 1292–1298. https://doi.org/10.1038/nn.2635

Lee, D. G., & Daunizeau, J. (2021). Trading mental effort for confidence in the metacognitive control of value-based decision-making. *ELife*, *10*, e63282. https://doi.org/10.7554/eLife.63282

Lee, D. G., & Holyoak, K. J. (2021). Coherence Shifts in Attribute Evaluations. *Decision*, No Pagination Specified-No Pagination Specified. https://doi.org/10.1037/dec0000151

Lee, D. G., & Usher, M. (2021). Value Certainty in Drift-Diffusion Models of Preferential Choice. *BioRxiv*, 2020.08.22.262725. https://doi.org/10.1101/2020.08.22.262725

Levav, J., Heitmann, M., Herrmann, A., & Iyengar, S. S. (2010). Order in Product Customization Decisions: Evidence from Field Experiments. *Journal of Political Economy*, *118*(2), 274–299. https://doi.org/10.1086/652463

Lichtenstein, S., & Slovic, P. (1971). Reversals of Preference between Bids and Choices in Gambling Decisions. *Journal of Experimental Psychology*, *89*(1), 46–55. https://doi.org/10.1037/h0031207

Lopez-Persem, A., Domenech, P., & Pessiglione, M. (2016). How prior preferences determine decision-making frames and biases in the human brain. *ELife*, *5*, e20317. https://doi.org/10.7554/eLife.20317

Payne, J. W., Bettman, J. R., & Johnson, E. J. (1988). Adaptive strategy selection in decision making. *Journal of Experimental Psychology: Learning, Memory and Cognition*, *14*, 534–552.

Payne, J. W., Bettman, J. R., & Johnson, E. J. (1993). *The adaptive decision maker*. Cambridge University Press. https://doi.org/10.1017/CBO9781139173933

Peixoto, D., Verhein, J. R., Kiani, R., Kao, J. C., Nuyujukian, P., Chandrasekaran, C., Brown, J., Fong, S., Ryu, S. I., Shenoy, K. V., & Newsome, W. T. (2021). Decoding and perturbing decision states in real time. *Nature*, 1–6. https://doi.org/10.1038/s41586-020-03181-9

Penny, W. D. (2012). Comparing Dynamic Causal Models using AIC, BIC and Free Energy. *NeuroImage*, *59*(1), 319–330. https://doi.org/10.1016/j.neuroimage.2011.07.039

R Core Team. (2020). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing. https://www.R-project.org/

Ratcliff, R. (1978). A theory of memory retrieval. *Psychol. Rev*, *85*(2), 59–108.

Ratcliff, R., & McKoon, G. (2008). The Diffusion Decision Model: Theory and Data for Two-Choice Decision Tasks. *Neural Computation*, *20*(4), 873–922. https://doi.org/10.1162/neco.2008.12-06-420

Ratcliff, R., & Rouder, J. N. (1998). Modeling Response Times for Two-Choice Decisions. *Psychological Science*, *9*(5), 347–356. https://doi.org/10.1111/1467-9280.00067

Rigoux, L., Stephan, K. E., Friston, K. J., & Daunizeau, J. (2014). Bayesian Model Selection for Group Studies—Revisited. *NeuroImage*, *84*, 971–985. https://doi.org/10.1016/j.neuroimage.2013.08.065

Roe, R. M., Busemeyer, J. R., & Townsend, J. T. (2001). Multialternative Decision Field Theory: A Dynamic Connectionist Model of Decision Making. *Psychological Review*, *108*(2), 370–392. https://doi.org/10.1037/0033-295X.108.2.370

Russo, J. E., & Dosher, B. A. (1983). Strategies for multiattribute binary choice. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *9*(4), 676–696. https://doi.org/10.1037/0278-7393.9.4.676

Russo, J. E., Medvec, V. H., & Meloy, M. G. (1996). The Distortion of Information during Decisions. *Organizational Behavior and Human Decision Processes*, *66*(1), 102–110. https://doi.org/10.1006/obhd.1996.0041

Seidl, C. (2002). Preference Reversal. *Journal of Economic Surveys*, *16*(5), 621–655. https://doi.org/10.1111/1467-6419.00184

Shah, A. K., & Oppenheimer, D. M. (2008). Heuristics made easy: An effort-reduction framework. *Psychological Bulletin*, *134*(2), 207–222. https://doi.org/10.1037/0033-2909.134.2.207

Stan Development Team. (2021). *Stan Modeling Language Users Guide and Reference Manual, Version 2.27.0*. http://mc-stan.org/

Stephan, K. E., Penny, W. D., Daunizeau, J., Moran, R. J., & Friston, K. J. (2009). Bayesian model selection for group studies. *NeuroImage*, *46*(4), 1004–1017. https://doi.org/10.1016/j.neuroimage.2009.03.025

Towal, R. B., Mormann, M., & Koch, C. (2013). Simultaneous Modeling of Visual Saliency and Value Computation Improves Predictions of Economic Choice. *Proceedings of the National Academy of Sciences*, *110*(40), E3858–E3867. https://doi.org/10.1073/pnas.1304429110

Tversky, A. (1972). Elimination by aspects: A theory of choice. *Psychological Review*, *79*(4), 281–299. https://doi.org/10.1037/h0032955

Tversky, A., Sattath, S., & Slovic, P. (1988). Contingent Weighting in Judgment and Choice. *Psychological Review*, *95*(3), 371–384. https://doi.org/10.1037/0033-295X.95.3.371

Tversky, A., Slovic, P., & Kahneman, D. (1990). The Causes of Preference Reversal. *The American Economic Review*, *80*(1), 204–217.

Vehtari, A., Gelman, A., & Gabry, J. (2017). Practical Bayesian Model Evaluation Using Leave-One-out Cross-Validation and WAIC. *Statistics and Computing*, *27*(5), 1413–1432. https://doi.org/10.1007/s11222-016-9696-4

Weber, E. U., & Johnson, E. J. (2009). Mindful Judgment and Decision Making. *Annual Review of Psychology*, *60*(1), 53–85. https://doi.org/10.1146/annurev.psych.60.110707.163633

Weber, E. U., Johnson, E. J., Milch, K. F., Chang, H., Brodscholl, J. C., & Goldstein, D. G. (2007). Asymmetric Discounting in Intertemporal Choice: A Query-Theory Account. *Psychological Science*, *18*(6), 516–523. https://doi.org/10.1111/j.1467-9280.2007.01932.x