Cognitive Psychology 142 (2023) 101562

Contents lists available at ScienceDirect

Cognitive Psychology

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

Information acquisition and decision strategies in intertemporal choice

Lisheng He [a,*], Daniel Wall [b], Crystal Reeck [c], Sudeep Bhatia [b]

[a] SILC Business School, Shanghai University, China
[b] Department of Psychology, University of Pennsylvania, United States
[c] Fox School of Business, Temple University, United States

ARTICLE INFO

Keywords:
Decision making
Intertemporal choice
Attention
Process tracing
Computational modeling

ABSTRACT

Intertemporal decision models describe choices between outcomes with different delays. While these models mainly focus on predicting choices, they make implicit assumptions about how people acquire and process information. A link between information processing and choice model predictions is necessary for a complete mechanistic account of decision making. We establish this link by fitting 18 intertemporal choice models to experimental datasets with both choice and information acquisition data. First, we show that choice models have highly correlated fits: people that behave according to one model also behave according to other models that make similar information processing assumptions. Second, we develop and fit an attention model to information acquisition data. Critically, the attention model parameters predict which type of intertemporal choice models best describes a participant's choices. Overall, our results relate attentional processes to models of intertemporal choice, providing a stepping stone towards a complete mechanistic account of intertemporal decision making.

1. Introduction

Choices often involve options received at different points in time (Frederick et al., 2002). For example, a consumer might finance an appliance purchase and delay the payment for the appliance, even if it means paying more overall. In the lab, intertemporal choices are generally operationalized as tradeoffs between two amounts of money received at different points in time. While often different in format to real-world choices, these in-lab choices correlate with many real-world outcomes, such as educational achievement, smoking behavior, and income (Golsteyn et al., 2014; Reimers et al., 2009). Given the importance and ubiquity of intertemporal choices, psychologists, economists, and neuroscientists have proposed several mathematical models to describe them (see Doyle, 2013; He et al., 2022; Bhatia et al., 2021 for an overview).

Recent work has shown that these models broadly fall into two classes: *discounting* and *time-as-attribute* models (He et al., 2022). Discounting models assume that people calculate the utility of each option separately by integrating its monetary amount and the delay until it is received. For example, in the exponential discounting model (Samuelson, 1937), the subjective value of a monetary amount is multiplied by a discount factor, which is an exponential transformation of the option's delay into the future. On the other hand, time-as-attribute models assume that people directly compare the two available options' time delays and monetary amounts. For example, the DRIFT model (Read et al., 2013) uses the differences between monetary amounts and delays of the two options to determine which

* Corresponding author at: SILC Business School, Shanghai University, 20 Chengzhong Rd, Shanghai, China.
E-mail address: hlisheng@shu.edu.cn (L. He).

https://doi.org/10.1016/j.cogpsych.2023.101562
Received 20 January 2022; Received in revised form 16 March 2023; Accepted 21 March 2023
Available online 28 March 2023

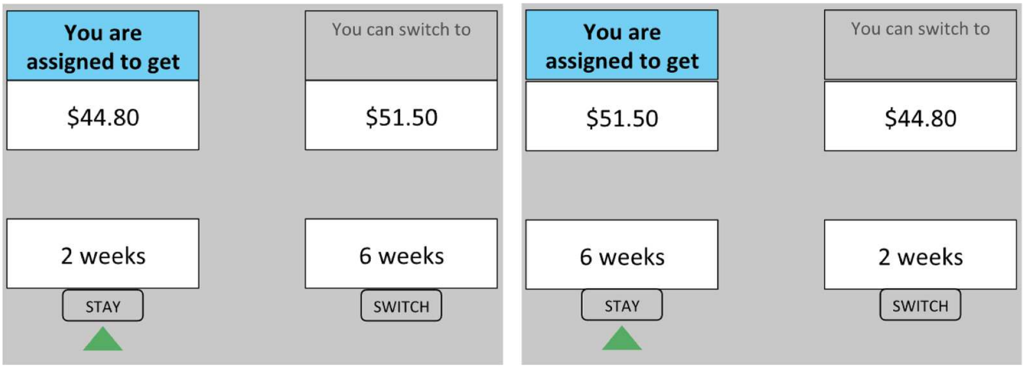

Fig. 1. (Taken from Reeck et al., 2017) Sample trial from the intertemporal choice task. Participants chose between a smaller monetary amount delivered sooner and a larger monetary amount delivered later. The trial on the left features a delay frame, as the smaller sooner option is the default, and participants can choose to delay consumption and switch to the larger later option. Conversely, the trial on the right features an acceleration frame, as the larger later option is the default, and participants can choose to accelerate consumption and switch to the smaller sooner option. During the task, information was occluded, and participants needed to move the mouse over the information they wanted to acquire.

one is better (though note that some time-as-attribute models also allow for within-option calculations, which can be interpreted as interaction effects between the time attribute and the monetary amount).

Much recent work has evaluated how well discounting and time-as-attribute models describe participants' choices (Dai & Busemeyer, 2014; Ericson et al., 2015; Scholten et al., 2014, 2016; Scholten & Read, 2013; Wulff & van den Bos, 2018). Although time-as-attribute models have been found to describe participants' choices better than discounting models, and also provide a superior account of response times (Amasino et al., 2019; Dai & Busemeyer, 2014), there is substantial heterogeneity in model fit. Some individuals have choices that are better fit by time-as-attribute models, while others have choices that are better fit by discounting models (Wulff & van den Bos, 2018).

Individual heterogeneity could be due to differences in underlying choice processes. Indeed, time-as-attribute and discounting models make different implicit assumptions about how people attend to information. Prior work has attempted to reduce choice models to their elementary information processes (Johnson et al., 2008; Johnson & Payne, 1985; Shah & Oppenheimer, 2008; Simon & Newell, 1971; Willemsen et al., 2011). For instance, calculating the expected value of a gamble (e.g., a 10% chance of $50) has been argued to involve: 1. reading the option's monetary amount, 2. looking at its probability, and 3. multiplying the monetary amount by the probability (Johnson & Payne, 1985). Similarly, calculating the discounted utility of a delayed monetary amount likely involves 1. reading the amount and calculating its value, 2. reading the delay and calculating the discount factor for that delay, and 3. multiplying the value by the discount factor. To calculate the discounted utility of an option, a person would thus have to hold both the option's monetary amount and delay in mind, implying that they would sequentially sample information *within options*. By contrast, to calculate the difference between monetary amounts, as in time-as-attribute models, a person would have to hold the amounts of both options in mind, implying they should sequentially sample information *within attributes* (Johnson et al., 2008; Johnson & Payne, 1985).

Unsurprisingly there is also substantial individual-level heterogeneity in attentional processes. Previous work has found that some participants sequentially sample information within options, whereas others sequentially sample information within attributes (Reeck et al., 2017). Additionally, participants who sequentially sample information within options are less patient. They are also less likely to show the accelerate-delay asymmetry, a common intertemporal framing effect according to which people become less impatient when the sooner option is framed as the acceleration of the delayed option than when the delayed option is described as the deferral of the sooner option (see Loewenstein, 1988; Weber et al., 2007), than those who sequentially attend to information within attributes. The accelerate-delay asymmetry is better described by a time-as-attribute model than a discounting model (Scholten & Read, 2013), suggesting that individual differences in framing effects may be, in part, due to differences in information processing. Relatedly, more attention to amounts, compared to delays, is associated with more patience (Amasino et al., 2019). Taken together, these results suggest that the dynamics of a participant's attention may map onto the type of model that best fits their choices.

While this suggestion is intriguing, it requires a formal link between information acquisition dynamics and choice model performance to be adequately tested. Such a link has two main requirements: 1. a quantitative model of information acquisition that allows for several different variables to jointly influence attentional dynamics, and 2. a comprehensive model fitting analysis that tests and compares several existing intertemporal choice models on participants' data. We attempt to create such a link. First, we fit 13 discounting models and five time-as-attribute models to two intertemporal choice datasets. Second, we develop an attention model that provides a rigorous description of the dynamics of participants' information acquisition strategies. We fit this model on participants' information acquisition data and subsequently use the parameters of the attention model to determine how well different types of models describe participants' choices. These tests provide a theoretically motivated and quantitatively rigorous understanding of the relationship between information acquisition and choice model predictions, and subsequently, shed light on the complex and diverse set of decision processes that underlie intertemporal choice.

Table 1

Models' aggregate marginal log-likelihoods of data using the Luce or Logit choice rule. The number in front of the parenthesis in each cell is the aggregate marginal log-likelihood. The larger the value, the better the model performance. The number in the parentheses indicates the proportion of participants whose choice data were best fit by the corresponding model. The boldface cell denotes the highest aggregate marginal log-likelihood in each column. Models #1–13 are discounting models and Models #14–18 are time-as-attribute models.

Models	Luce				Logit			
	E1-Accel ($N = 193$)	E1-Delay ($N = 193$)	E2-Days ($N = 133$)	E2-Weeks ($N = 134$)	E1-Accel ($N = 193$)	E1-Delay ($N = 193$)	E2-Days ($N = 133$)	E2-Weeks ($N = 134$)
1.Exponential	−2089 (1%)	−2054 (2%)	−1963 (22%)	−1917 (23%)	−1952 (0%)	−1961 (0%)	−1590 (2%)	−1553 (0%)
2.Quasi-hyperbolic	−2107 (7%)	−2073 (5%)	−2019 (3%)	−1978 (3%)	−2169 (2%)	−2183 (1%)	−1764 (0%)	−1756 (0%)
3.Constant-sensitivity	−2085 (13%)	−2048 (16%)	−1966 (21%)	−1907 (19%)	−1989 (1%)	−1994 (1%)	−1616 (1%)	−1586 (0%)
4.Double-exponential	−2131 (0%)	−2103 (0%)	−1999 (0%)	−1962 (0%)	−2051 (0%)	−2054 (0%)	−1641 (0%)	−1617 (0%)
5.Hyperbolic	−2288 (0%)	−2281 (0%)	−2344 (0%)	−2351 (0%)	−1950 (0%)	−1962 (0%)	−1569 (2%)	−1551 (1%)
6.Hyperbolic (Power)	−2188 (0%)	−2176 (0%)	−2196 (0%)	−2159 (0%)	−1944 (0%)	−1967 (0%)	−1595 (0%)	−1585 (0%)
7.Generalized hyperbolic	−2124 (0%)	−2109 (0%)	−2135 (0%)	−2099 (0%)	−1913 (0%)	−1944 (0%)	−1560 (0%)	−1555 (0%)
8.Generalized hyperbolic (IE)	−2118 (0%)	−2103 (0%)	−2114 (0%)	−2078 (0%)	−1807 (4%)	−1850 (5%)	−1489 (2%)	−1472 (1%)
9.Common aspect attenuation	−2267 (0%)	−2255 (0%)	−2312 (0%)	−2309 (0%)	−1907 (16%)	−1910 (13%)	−1556 (3%)	−1522 (3%)
10.ASAP	−2233 (0%)	−2214 (0%)	−2277 (2%)	−2264 (0%)	−1924 (2%)	−1930 (0%)	−1576 (0%)	−1541 (0%)
11.Interval	−2108 (0%)	−2083 (0%)	−2102 (0%)	−2060 (1%)	−1900 (0%)	−1912 (0%)	−1604 (0%)	−1574 (0%)
12.Generalized interval	−2119 (0%)	−2108 (0%)	−2117 (0%)	−2081 (0%)	−1928 (0%)	−1964 (1%)	−1591 (0%)	−1592 (0%)
13.Generalized interval (IE)	−2188 (1%)	−2186 (0%)	−2077 (0%)	−2048 (1%)	−1716 (2%)	−1766 (0%)	−1407 (8%)	−1401 (13%)
14.DRIFT	−1814 (20%)	−1809 (31%)	**−1828 (21%)**	**−1822 (22%)**	−1596 (17%)	−1570 (22%)	**−1223 (46%)**	**−1219 (44%)**
15.ITCH	−1890 (30%)	−1912 (23%)	−1983 (6%)	−1993 (1%)	**−1548 (57%)**	**−1546 (58%)**	−1256 (35%)	−1276 (31%)
16.Attribute-based (Power)	**−1768 (4%)**	**−1783 (3%)**	−1952 (0%)	−1935 (0%)	−1724 (0%)	−1741 (0%)	−1390 (0%)	−1397 (0%)
17.Tradeoff	−1832 (10%)	−1832 (10%)	−1979 (2%)	−1961 (0%)	−1892 (0%)	−1904 (0%)	−1538 (3%)	−1526 (4%)
18.Generalized tradeoff	−1824 (14%)	−1852 (10%)	−1896 (24%)	−1873 (29%)	−2004 (1%)	−2031 (0%)	−1641 (0%)	−1624 (2%)

2. Methods

2.1. Reeck et al. (2017) dataset

We firstly reanalyzed the data from Experiment 1 of Reeck et al. (2017). In this experiment, 200 participants made 36 intertemporal choices between a smaller sooner (SS) and a larger later (LL) option. The SS amount, SS_{amount}, ranged from \$15.10 to \$68.30; the LL amount, LL_{amount}, ranged from \$17.40 to \$85.40. The SS delay, SS_{delay}, was either zero weeks or two weeks; the LL delay, LL_{delay}, was either two weeks or four weeks greater than the corresponding SS delay. Further, with respect to the accelerate-delay asymmetry, each participant completed one block of 18 *delay* trials, where the SS option was presented as the default, and one block of 18 *accelerate* trials, where the LL option was presented as the default. (See Fig. 1 for more information). Participants were paid \$1 to complete the experiment.

Participants' attention was measured via MouselabWEB, which tracked mouse movements in the online experiment. In order to acquire information about an attribute, participants had to move their cursor over a box, which revealed the attribute (amount or delay). After the cursor was removed from the box, the attribute was occluded once again. This method allowed for the measurement of the order in which participants acquired information. The seven participants who acquired no information about the options for more than 20% of the trials were excluded from all analyses, leaving 193 participants in the full dataset (mean age = 34.00 years, 41% female). Boxes that were kept open for less than 200 ms were also removed. Unlike Reeck et al. (2017), we did not remove reacquisitions of information (e.g., viewing information about the SS_{amount} then re-acquiring the same information), as our attention model was able to accommodate them. For complete details of the study, see Reeck et al. (2017).

2.2. Lee et al. Dataset

We also analyzed the MouselabWEB data from a related but unpublished intertemporal choice study (Lee et al., in preparation). The monetary amounts for the SS option ranged from \$1.20 to \$80.30; the amounts for the LL option ranged from \$2.10 to \$85.90. The amounts for both options were sampled from one of four magnitudes: extra small (\$1.20 to \$4.60), small (\$11.20 to \$35.60), medium (\$20.10 to \$60.90), or large (\$31.60-\$85.90). Additionally, the delay format was manipulated between participants: delays were presented as days into the future (e.g., \$85.90 in 7 days) or weeks into the future (e.g., \$85.90 in 1 week). The SS delays ranged from 0 weeks (0 days) to 2 weeks (14 days); the LL delays ranged from 1 week (7 days) to 13 weeks (91 days).

The study recruited 301 participants (mean age = 34.3, 43.7% female) on Amazon Mechanical Turk. Participants completed one practice trial and 30 test trials and were paid $1.25 to complete the study. Additionally, participants were informed that 1% of participants would be paid according to their choices—e.g., they would receive the monetary amount at the delay specified in the choice. Based on the exclusion rules in Reeck et al. (2017), 24 participants were removed because they acquired no information about the choices on 20% or more of the trials. Additionally, 10 participants were removed because they failed a catch trial that had participants choose between a larger monetary amount received sooner than a smaller monetary amount. As with prior research, we also removed box-opening that was less than 200 ms. Participants' attention was once again tracked via MouselabWEB. For expositional convenience, the Reeck et al. (2017) dataset is referred to as Experiment 1 (E1) and the Lee et al. dataset as Experiment 2 (E2) in the paper (Table 1).

2.3. Choice models

For this analysis, we compiled 18 models that predict choices between a smaller amount of money received sooner—the SS option—and a larger amount of money received later—the LL option. Tables A1, A2, and A3 in the Appendix A provide the complete list of models, their functional forms and parameters. These models take choice information—the amounts of money and the delays for both options— as inputs and predict participant's choices. We did not include other intertemporal choice models, such as query theory (Weber et al., 2007), because they typically do not have a concise mathematical form.

The 18 models can be roughly classified into three categories: 8 delay discounting models, e.g., exponential (Samuelson, 1937) and hyperbolic discounting (Mazur, 1987); 5 time-as-attribute models, e.g., the tradeoff model (Scholten & Read, 2010) and the ITCH model (Ericson et al., 2015); and 5 interval discounting models, e.g., the generalized-interval model (Scholten & Read, 2006) and the as-soon-as-possible (ASAP) model (Kable & Glimcher, 2010). This taxonomy represents the main approaches to modeling intertemporal decisions.

Delay discounting models assume a process where people calculate the value of the options independently and pick the higher value option. Consider an option $X = (x, t)$ offering $\$x$ with a delay t. A delay discounting model will assign a discounted utility in the form of $U(X) = u(x)d(t)$, where $u(x)$ is the subjective value of × and $d(t)$ is a discount function that is typically monotonically decreasing in t. Given their deep roots in economic analysis, most of these models do not explicitly predict how people gather information. However, as outlined in the introduction above, we can infer, based on their mathematical properties, that an efficient acquisition of information would entail within-option search.

Unlike delay discounting models, time-as-attribute models assume that decision-makers treat delays as an attribute in the same way they treat amounts, thus making intertemporal choice a special case of multi-attribute choice (Rubinstein, 2003; Scholten & Read, 2010). Correspondingly, decision-makers evaluate each option's (dis)advantage along amounts and delays. For a choice between $X = (x, t)$ and $Y = (y, s)$, where the amounts $0 < x < y$ and delays $0 \le t < s$, decision-makers evaluate X's advantage in temporal proximity as $U(X) = f(t, s)$, and Y's advantage in monetary amount as $U(Y) = g(y, x)$. Different time-as-attribute models have different forms of $f(\cdot)$ and $g(\cdot)$. Although most of these models do not make explicit claims about information acquisition processes, efficient information acquisition in these models would entail within-attribute search.

Finally, interval discounting models are like delay discounting models because each option is assigned a utility value that is discounted by some function of its delay. However, in assigning utilities to options, these models allow interactions between options' delays. Thus, for example, in a binary choice between $X = (x, t)$ and $Y = (y, s)$, Y's utility depends not only on y and s but also on t. Formally this can be written as $U(Y) = u(y)D(s, t)$. Even though the mechanistic assumptions of interval discounting models involve some type of within-attribute comparison, we find (both in the current paper and in prior work: He et al. 2022) that these models' quantitative predictions are typically closer to those of delay discounting models. For this reason, we combine delay discounting models and interval discounting models into a broad category of *discounting models*.

To obtain estimates of choice model performance, we fit each of the 18 choice models to individual-level data. We considered two different stochastic specifications, the Logit and Luce choice rules, to facilitate these fits and replicated all our tests with each of these stochastic specifications. Logit defines the probability of choosing option X in a binary choice between X and Y as $\Pr[X; Y] = \frac{1}{1+\exp(-\varepsilon(U(X)-U(Y)))}$, where ε represents the amount of determinism in the decision process. The larger the ε, the more significant the effect of $U(X) - U(Y)$ on $\Pr[X; Y]$. The Luce choice rule defines the probability of choosing X as $\Pr[X; Y] = \frac{U(X)^{\varepsilon}}{U(X)^{\varepsilon}+U(Y)^{\varepsilon}}$. According to the Luce choice rule, $\Pr[X; Y]$ is a monotonic function of the ratio between the two utilities, rather than a monotonic function of the difference between the two utilities. Thus, the Luce choice rule is also called the ratio rule.

We considered both the Logit and Luce rules in this paper as previous findings suggest that stochastic specifications can strongly influence intertemporal choice models' quantitative predictions and thus model fits (Dai, 2017; He et al., 2019; Regenwetter et al., 2018; Scholten et al., 2014). Note that the Probit rule is also a popular candidate choice rule. He et al. (2022) have found that the Logit and Probit rules give largely similar results, that is, the similarities between models measured using the Logit rule are nearly identical to those measured using the Probit rule (see Table 1 of He et al., 2022). By contrast Luce differs quite substantially from Logit and Probit. Our focus on Logit and Luce (and omission of Probit), in the current paper, is motivated by this result.

2.4. Choice model fits and evaluation

We evaluated each choice model's fit with a Bayesian approach that automatically penalizes model complexity for model comparisons. Specifically, we employed a simple Monte Carlo method to estimate the marginal log-likelihood (MLL) for each model for each participant's choice data (see He et al., 2019; Scholten et al., 2016 for similar applications), $MLL_i = \log p(D_n|i) = \log \int p(D_n|i,\Theta)p(\Theta|i)d\Theta$, where D_n is participant n's choice data and Θ is the set of free parameters for model i (Table A3 in the Appendix A presents the prior specification of the choice models' free parameters). For every estimation, we first drew one million samples of parameter values from the prior distribution. With each sample of parameter values, we calculated the likelihood of the choice data, obtaining one million likelihood values. The marginal likelihood was estimated as the average of the one million likelihood values. The marginal likelihood is commonly used in Bayesian model selection and underlies other key indexes of model fit, e.g., Bayes Factors (Andraszewicz et al., 2015; Fong & Holmes, 2020). Note that the calculation of a model's marginal likelihood of data is dependent on the prior distribution, rather than the posterior distribution, of the model parameters. This dependence is why the marginal likelihood is closely related to a model's prior predictive distribution.

To ensure that the estimation was stable and accurate, we ran the simulation for each model and each participant twice and the results reported below are based on the average of the two runs. We used three metrics to evaluate the convergence of simulation across runs. First, we found that in 99.6% of the cases, the two runs identified the same best fitting model (i.e., the model with the highest marginal likelihood). Second, the models' marginal likelihoods were highly correlated across runs for every participant, with all of the Pearson's coefficients larger than 0.998. Third, based on an equal prior model probability (i.e., 1/18), we calculated the models' posterior probabilities using the models' marginal likelihoods, and found that the between-run deviation in model posteriors were rather small, ranging from 0.00001 to 0.00197 (median = 0.00018; mean = 0.00023). Overall, the convergence checks suggest that our simple Monte Carlo method obtained stable estimates of model performance. The estimated individual-level marginal likelihoods can be found in the Excel file in supplemental materials.

Our main interest is in the relative performance between models. Thus, for every participant's choice data, we calculated a log Bayes factor (logBF) between pairs of models (Kass & Raftery, 1995). This metric measures the relative performance of models i and j as: $\text{logBF}_{ij} = \log\frac{p(D_n|i)}{p(D_n|j)} = \log p(D_n|i) - \log p(D_n|j) = MLL_i - MLL_j$, where MLL_i stands for the MLL for model i in accounting for choice data D_n.

Unlike non-Bayesian methods (e.g., Akaike Information Criterion) that rely on the best-fit parameters, our Bayesian method considered different parameter values so that it automatically struck a balance between the goodness of fit and model complexity. As in any other Bayesian modeling, the resulting Bayes factors may be sensitive to the prior specification of the model parameters. For that reason, we specified prior distributions for the model parameters (see Table A3), following previous literature on Bayesian fitting of intertemporal choice models (e.g., He et al., 2019; He et al., 2022; Scholten et al., 2014). For an additional robustness check, we used an alternative set of uniform prior specifications for a sensitivity analysis as detailed in the **Sensitivity Analysis** section below.

2.5. Attention model

There are numerous ways to describe information acquisition dynamics in intertemporal choice. Often this process is collapsed into a single descriptive number. For instance, the commonly used Payne index (Reeck et al., 2017) indexes the relative proportion of within-option—e.g., from the SS amount to the SS delay—and within-attribute—e.g., from the SS amount to the LL amount—-transitions. There are other single number indexes, such as the option index (describing option attention bias towards the SS or LL option) and the attribute index (describing attribute attention bias towards amounts and delays) as adopted by Amasino et al. (2019). While these single number summaries have been useful at predicting choices, they remove much information about the dynamics of information acquisition.

We resolve this issue by building a computational attention model to capture the potentially complex attention dynamics in intertemporal choice. In the datasets analysed, there are a total of four attentional states ($\mathscr{S} = \{SS_{amount}, SS_{delay}, LL_{amount}, LL_{delay}\}$). The attention model predicts switches between these states using transition probabilities $\Pr[s_t|s_{t-1}]$, where $s_{t-1}, s_t \in \mathscr{S}$ and $t = 1,2,3,\ldots,T$ are the time steps. We further write $\Pr[s_t|s_{t-1}]$ as a function of various predictors. Formally, the transition probability is:

$$\Pr[s_t|s_{t-1}] = \sigma\begin{pmatrix} \beta_1 x_t^{Optn} + \beta_2 x_t^{Attr} + \beta_3 x_t^{Optn} * x_t^{Attr} + \\ \tau_1 I^{Optn} + \tau_2 I^{Attr} + \tau_3 I^{Stat} + \\ (\delta_1 I^{Optn} + \delta_2 I^{Attr} + \delta_3 I^{Stat}) z_{t-1} + \\ \nu_1 I_t^{Novelty} \end{pmatrix} \tag{1}$$

$\sigma(\cdot)$ is the softmax function, a generalization of the logistic function that converts a set of real numbers indicating search propensity into probability distribution of different states in $\mathscr{S}$, such that $\sum_{s_t \in \mathscr{S}} \Pr[s_t|s_{t-1}] = 1$. x_t^{Optn} specifies whether s_t belongs to the SS option or the LL option (1 if SS; 0 if LL), x_t^{Attr} specifies whether s_t belongs to the monetary amount attribute or the time delay attribute (1 if amount; 0 if delay), and $x_t^{Optn} * x_t^{Attr}$ describes an interaction between amount and time to characterize whether s_t belongs to a particular state (1 if SS_{amount}; 0 otherwise). I^{Optn} indicates whether s_t and s_{t-1} belong to the same option (1 if same; 0 if different), I^{Attr} indicates whether s_t and s_{t-1} belong to the same attribute (1 if same; 0 if different), and I^{Stat} indicates whether s_t and s_{t-1} belong to the same state (1 if the same state; 0 if different). For expositional brevity, we drop the subscripts t and t-1 in these variables. z_{t-1} represents the value uncovered during the most recently attended to state, s_{t-1} (e.g., the specific monetary amount or the specific delay). Finally, the novelty variable , $I_t^{Novelty}$, indicates whether the attentional state, s_t, was previously attended to (i.e. the box had been opened) in the

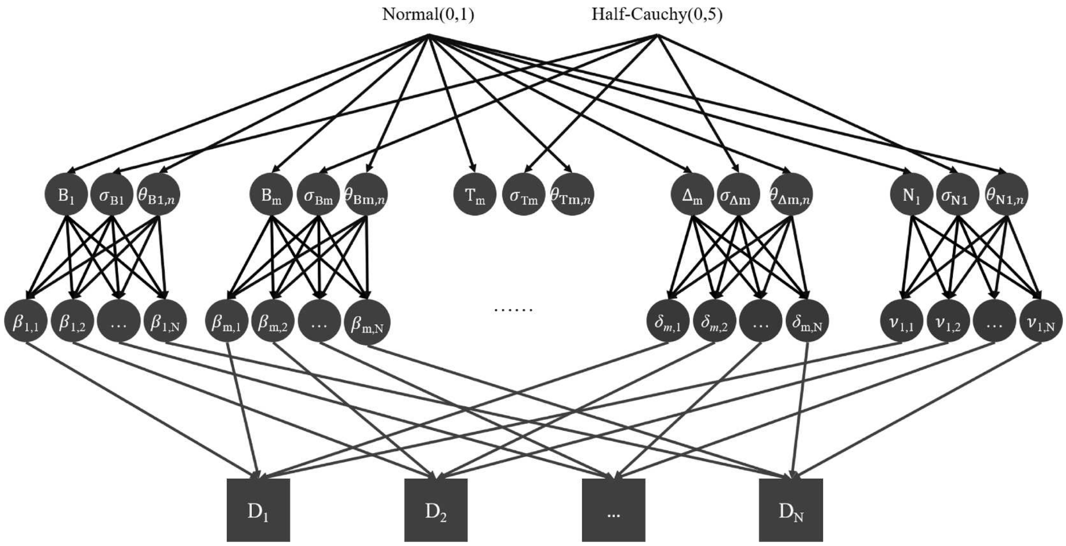

Fig. 2. Hierarchical Bayesian estimation of the attention model. B_m (m = 1, 2, and 3) represents the group-level attention bias parameters, T_m represents the group-level transition parameters, Δ_m represents the group-level dynamic transition parameters and N_1 represents the group-level novelty parameter. σ_μ represents the amount of variation across participants for each $\mu \in \{B_m, T_m, \Delta_m, N_1\}$. Correspondingly, $\beta_{m,n}$, $\tau_{m,n}$, $\delta_{m,n}$ and $\nu_{1,n}$ represent participant n's individual-level bias, transition, and dynamic transition parameters, respectively. Both group- and individual-level parameters are estimated by simultaneously fitting all individual-level process data $\{D_1, \ldots, D_n, \ldots, D_N\}$.

trial (1 = unsearched, 0 = searched). The variable, ν_1, models the tendency to sample novel pieces of information.

To make the values of amounts and delays comparable, we standardized the values along each attribute using z-scoring. The z-scores of delays were reversed as increasing the delay of an option makes it less attractive. Thus if the decision-maker attended to SS_{amount} at time *t*-1, revealing a value of SS_{amount} = $52.5 (corresponding to a z-score of 0.38 in the Reeck et al., 2017 dataset) and SS_{delay} had not been attended to previously, we would obtain a measure of the overall tendency of attending to SS_{delay} at *t* using $\beta_1*1 + \beta_2*0 + \beta_3*0 + \tau_1*1 + \tau_2*0 + \tau_3*0 + (\delta_1*1 + \delta_2*0 + \delta_3*0)*0.38 + \nu_1*1$. Similar calculations would reveal the overall tendency of attending to SS_{amount}, LL_{amount}, and LL_{delay} at *t*. We turn these tendencies into predicted probabilities for subsequent attention via the softmax function, a link function widely used in multinomial regression and machine learning. By fitting these parameters to the observed attentional search data, we can characterize the effects of various variables on attention probability.

The model parameters fall into four groups corresponding to the rows of equation (1). Row 1 includes the option and attribute *bias* parameters ($\beta' s$), which capture the overall (time-independent) attentional biases toward different options and attributes. Specifically, β_1 describes an overall attentional bias to SS versus LL, β_2 describes an overall attentional bias to amount versus delay, and β_3 describes an overall attentional bias towards SS_{amount} (i.e., the interaction term between options and attributes) in addition to that described by β_1 and β_2 (the overall attentional biases to each of the other three states can be obtained using a combination of these three coefficients).

Row 2 includes the option and attribute *transition* parameters ($\tau' s$), which capture sequential transition tendencies within options, within attributes, and within states, respectively. Specifically, τ_1 describes how likely the decision-maker is to transition within the attributes of a single option (e.g., $SS_{amount} \rightarrow SS_{delay}$), τ_2 describes how likely the decision-maker is to transition across options within a single attribute (e.g., $SS_{amount} \rightarrow LL_{amount}$) and τ_3 describes how likely the decision-maker is to successively resample the same state (e.g., $SS_{amount} \rightarrow SS_{amount}$).

Row 3 includes the *dynamic value transition* parameters ($\delta' s$), which capture the effects of the value of the most recently attended information on within-option, within-attribute, or self-resampling tendencies respectively (i.e., interactions between the *I* variables and the value of z_{t-1}). Specifically, δ_1 describes whether attention to the other attribute within the same option depends on the value of the most recently attended information. If observing high values of an attribute makes decision-makers more likely to transition to other attributes of the same option, e.g., high values of SS_{amount} increase the probability of transitioning to SS_{delay}, then we should observe a positive δ_1. Similarly, δ_2 describes whether attention to the same attribute in a different option depends on the value of the most recently searched information. If high values of an attribute make decision-makers more likely to transition to the same attribute in other options (e.g., high values of SS_{amount} increase the probability of transitioning to LL_{amount}), we should observe a positive δ_2 . Finally, δ_3 describes whether attention to the same state depends on the value of the state. If observing high values of an attribute makes decision-makers more likely to resample the same attribute in the same option, e.g., high values of SS_{amount} increase the probability of transitioning back to SS_{amount}, then we should observe a positive δ_3. This specification was inspired by emerging evidence suggesting that the perceived value of an option influences subsequent information acquisition. Typically, high-value options attract more attention in information acquisition (Callaway et al., 2021; Fiedler & Glöckner, 2012; Gluth et al., 2020; Jekel et al., 2018; Scharf et al., 2019; Sepulveda et al., 2020).

Lastly, ν_1 in Row 4 describes the extent to which decision makers search for novel or unsearched information in the choice problem. If people tend to search for novel information, we should observe a positive ν_1.

In the attention model, τ_1 and τ_2 are closely related to the classic Payne index (Payne, 1976). The standard Payne index is:

$$P = \frac{T_O - T_A}{T_O + T_A},$$

where T_O is the number of within-option searches and T_A is the number of within-attribute searches. In binary choice involving two attributes, a random search process would generate a Payne index of 0, indicating that a positive value corresponds to largely within-option search, whereas a negative value corresponds to largely within-attribute search. τ_1 and τ_2 can be seen as the standardized T_O and $T_{A,}$ that control for a large set of other variables at play in attentional dynamics. Thus, we can define a model-based Payne index as:

$$M_P = \tau_1 - \tau_2.$$

Beyond that, the process model can also characterize participants' overall focus on different pieces of information with β_1, β_2 and β_3, how the observed value guides subsequent information acquisition with δ_1, δ_2 and δ_3, and how likely participants search for unsearched vs. searched information with ν_1. For example, β_1 can be seen as a model-based option index that describes the extent to which decision-makers attend to SS (relative to LL) and β_2 can be seen as a model-based attribute index that describes the extent to which decision-makers attend to monetary amounts relative to delays (see also Amasino et al. 2019 for a similar use of option index and attribute index).

2.6. Attention model fits and evaluation

We used hierarchical Bayesian analysis to fit the attention model, allowing for individual differences in each of the nine free parameters described in Equation (1) (Fig. 2). For the group-level hyperparameters B_m, T_m, Δ_m and N_m, where m represents the different parameters within each parameter category, we set the prior as the standard normal distribution. For each of the hyper-parameters, we needed to specify the dispersion of individuals' deviations from the group-level grand mean σ_μ ($\mu \in \{\mathrm{B}_m, \mathrm{T}_m, \Delta_m, \mathrm{N}_m\}$) and used a half-Cauchy distribution (with location = 0 and scale = 5) as the prior specification such that $\sigma_\mu \geq 0$.

At the individual level, we allowed participant n to deviate from the grand mean differently, with the amount of $\theta_{\mu,n}$ deviation (unit: σ_μ, where $\mu \in \{\mathrm{B}_m, \mathrm{T}_m, \Delta_m, \mathrm{N}_m\}$). The resulting individual-level parameters are the sum of the group-level mean and the individual-specific deviation from the group-level mean, such that:

$$\begin{aligned}\beta_{m,n} &= \mathrm{B}_m + \sigma_{\mathrm{B}_m}\theta_{\mathrm{B}_m,n},\\ \tau_{m,n} &= \mathrm{T}_m + \sigma_{\mathrm{T}_m}\theta_{\mathrm{T}_m,n},\\ \delta_{m,n} &= \Delta_m + \sigma_{\Delta_m}\theta_{\Delta_m,n},\\ \nu_{m,n} &= \mathrm{N}_m + \sigma_{\mathrm{N}_m}\theta_{\mathrm{N}_m,n},\end{aligned}$$

where m = 1,2 or 3 indexes the different parameters and n = 1,2,..., N indexes participant ID and thus $\beta_{m,n}$, $\tau_{m,n}$, $\delta_{m,n}$ and $\nu_{m,n}$ are the combined individual-specific and parameter-specific effects. The prior for $\theta_{\mu,n}$ ($\mu \in \{\mathrm{B}_m, \mathrm{T}_m, \Delta_m, \mathrm{N}_m\}$) is set at the standard normal distribution. $\beta_{m,n}$, $\tau_{m,n}$, $\delta_{m,n}$ and $\nu_{m,n}$ are used to fit the individual-level information acquisition data D_n.

We used the Hamiltonian Monte Carlo method in the *rstan* package for the posterior approximation (Stan Development Team, 2020). For each condition, we ran four independent chains, with each consisting of 2,500 formal iterations after 1,000 burn-in iterations, totaling 10,000 formal iterations for the posterior approximation. All Rubin-Gelman $\widehat{R}$ were below 1.05, indicating excellent convergence of the sampling method (Gelman & Rubin, 1992).

To validate our main attention model in Equation (1), we fit two nested models, and compared their performance to the main model. In one model, we used the bias and transition parameters (i.e., only the first and second lines of Equation (1). In the other, we added the dynamic value parameters (e.g., the δ parameters of Equation (1) with the omission of ν_1 in Row 4. According to the Deviance Information Criterion (DIC, Spiegelhalter et al., 2002), a popular metric that measures a model's goodness of fit to the data with model complexity penalized under the hierarchical Bayesian framework, the main model outperformed the two nested models in all four conditions. This result reinforces the need for all parameters in our main attention model.

2.7. Sensitivity analysis

To make sure that our findings were not sensitive to the prior specifications in the Bayesian estimation, we used uniform distributions as alternative prior specifications, thereby conducting a robustness check. For choice models, we used uniform distributions between the lower bound and upper bound as the prior for all model parameters. For model parameters that have no upper bound (see Table A3), their prior distributions in the Bayesian estimation were set to the uniform distribution between the corresponding lower bound (i.e., 0 or 1) and 10. For the hierarchical Bayesian estimation of the search model, we specified the prior of group-level hyperparameters B_m, T_m, Δ_m and N_m as a uniform distribution between −5 and 5 (i.e., B_m, T_m, Δ_m and N_m $\mathrm{Uniform}[-5,5]$). Other specifications of the model were unchanged.

With the alternative flat prior specifications, we found that all the key results reported in the paper using the main prior specification held. Therefore, we only report the results using the main prior specification in the main text. The detailed results based on the

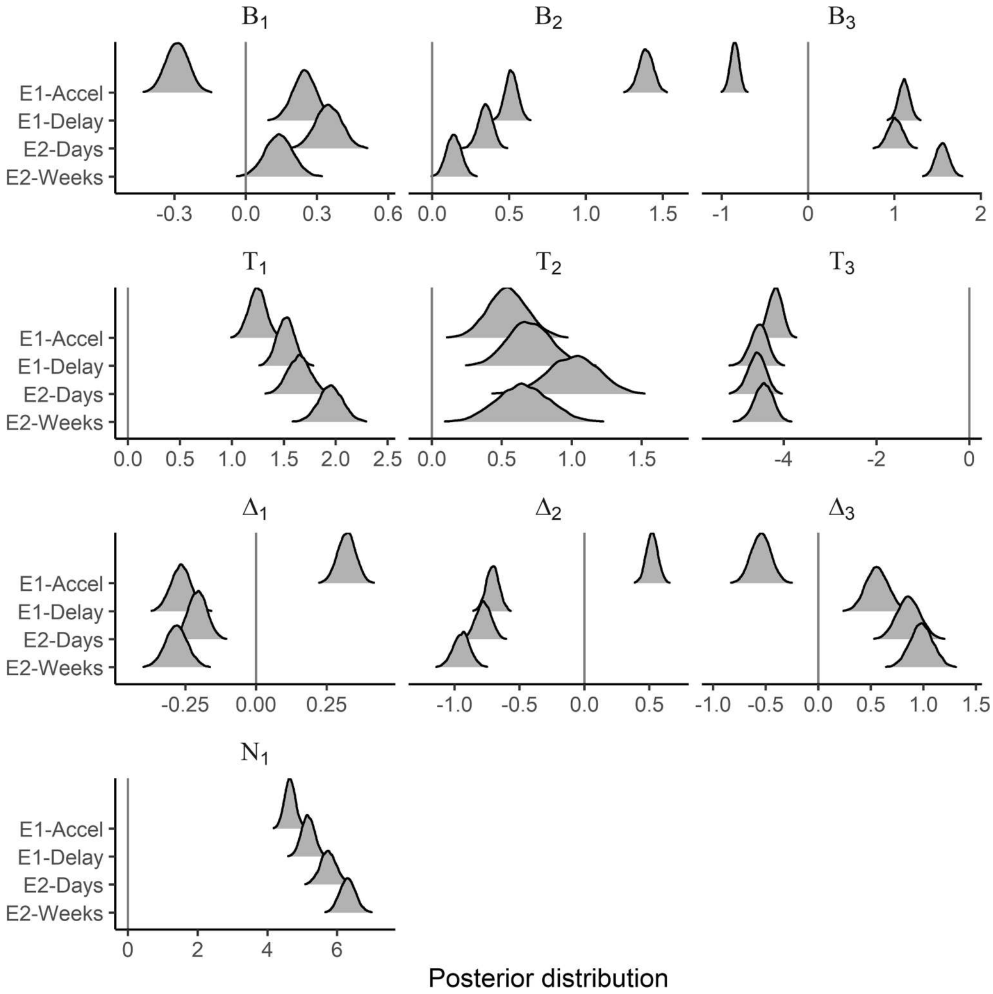

Fig. 3. Posterior distributions of group-level bias (B), transition (T), dynamic transition (Δ) and novelty (N) parameters in the attention model. Each of the ten panels displays one group-level parameter in the four experimental conditions, respectively. The vertical lines indicate zeros (i.e., the prior mean). The parameters were estimated independently for different conditions, but we plot each parameter in different conditions in the same panel for expositional convenience.

alternative prior specifications can be found in online supplemental materials (Table S1 and Figs. S1-S11).

2.8. Interim summary

Different choice models make competing information processing assumptions that can be related to parameters from our attention model. For instance, time-as-attribute models propose that people compare the attributes between the choice options, whereas the discounting models propose that people evaluate the attributes within choice options. We formally test these predictions using our attention model, which parameterizes within- versus between-option search tendencies, controlling for a very large number of other drivers of information search. Our model fits, as detailed below, indicate that there was substantial heterogeneity in participant search tendencies. Moreover, participants who searched in a within-attribute manner (i.e. $\tau_1 < \tau_2$ in Equation (1) were more likely than those who searched in a within-option manner (i.e. $\tau_1 > \tau_2$) to follow an attribute-based choice strategy (as opposed to a discounting-style choice strategy). In this way, we related our parameter estimates to the implicit information processing assumptions (and subsequently model fits) of our choice models.

3. Results

3.1. Data overview

There were a total of 460 participants in the final data for the two studies, with 193 participants in the Accelerate condition (E1-Accel) and the Delay condition (E1-Delay) of Experiment 1 (within-participant design), 133 participants in the Days condition of

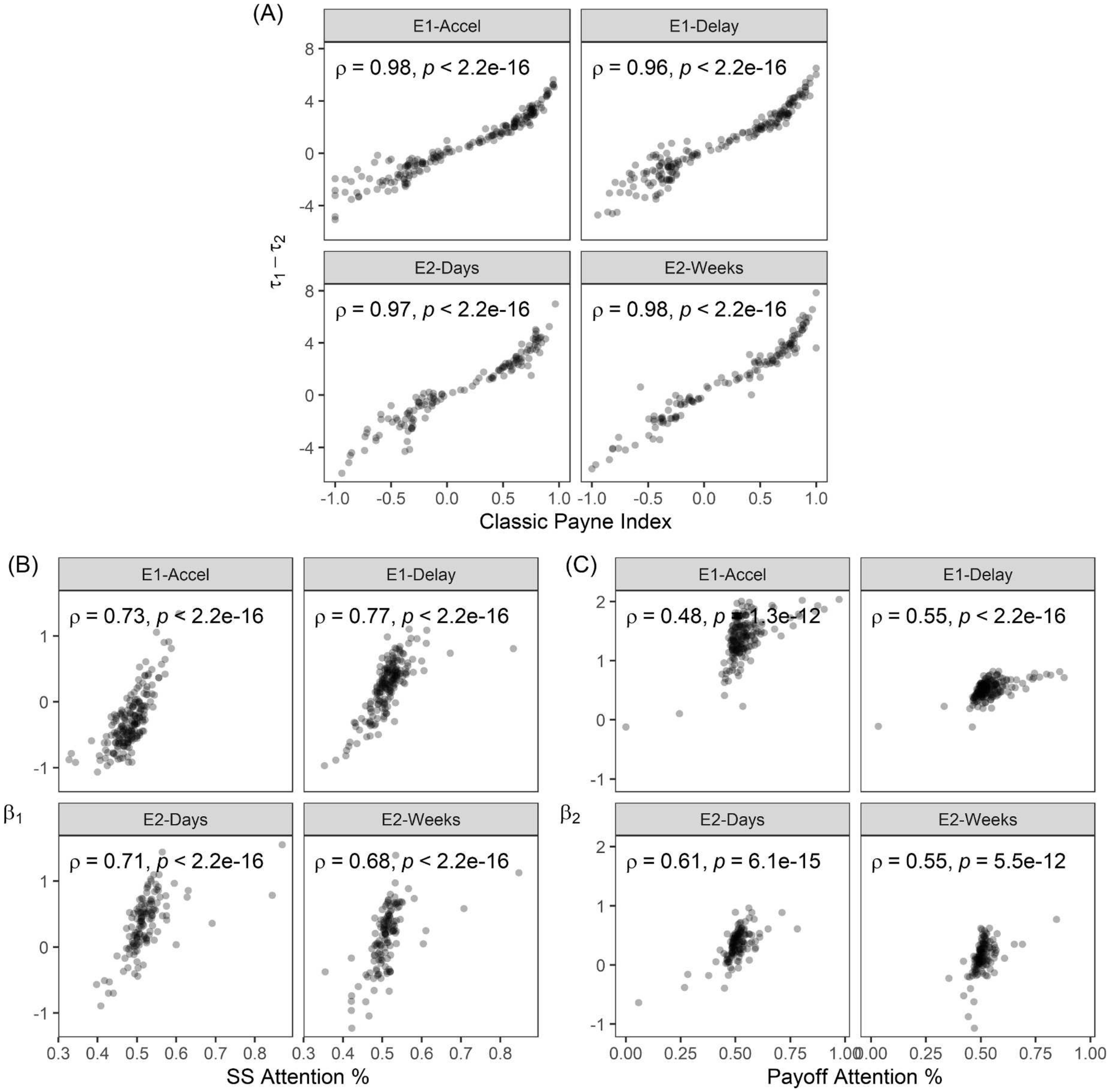

Fig. 4. Individual-level attention properties and attentional parameters in attention model in different conditions. (A) Correlations between the classic Payne index P and the model-based Payne index $M_P = \tau_1 - \tau_2$. (B) Correlations between the option index (SS attention proportions) and β_1. (C) Correlations between the attribute index (amount attention proportions) and β_2. Spearman's ρ's are reported in the figure.

Experiment 2 (E2-Days), and 134 participants in the Weeks condition of Experiment 2 (E2-Weeks). Participants made 18, 18, 30, and 30 intertemporal choices in these four conditions. Although participants in Experiment 1 took part in both the accelerate and delay conditions, we modeled process and choice data for each of the two conditions separately. Prior research has demonstrated that acceleration versus delay framing alters decision making and that this effect depends on how people process choice information (Li et al., 2021; Reeck et al., 2017; Weber et al., 2007). Therefore, we decided to model these trials separately. Because the framing of delay affects both choices and decision process and because the manipulation in E2 was conducted between participants, we elected to model these as separate experiments in the current analysis.

Participants chose the smaller sooner (SS) option over the larger later (LL) option in 56.8%, 60.7%, 60.8%, and 61.3% of trials in the E1-Accel, E1-Delay, E2-Days, and E2-Weeks conditions, respectively. Additionally, in the four conditions, 54.9%, 65.3%, 69.9%, and 67.9% of participants chose SS in more than half of the trials. The differences between E1-Accel and E1-Delay are due to the experimental manipulation in Reeck et al. (2017), which varied whether SS or LL was the default option.

On the process side, we calculated the classic Payne index using the relative difference between within-option and within-attribute transitions across trials. The average Payne index across all participants was 0.14, 0.15, 0.15, and 0.25 in the E1-Accel, E1-Delay, E2-Days, and E2-Weeks conditions, respectively. Additionally, 53.9%, 52.8%, 51.9% and 61.2% of participants had a positive Payne index

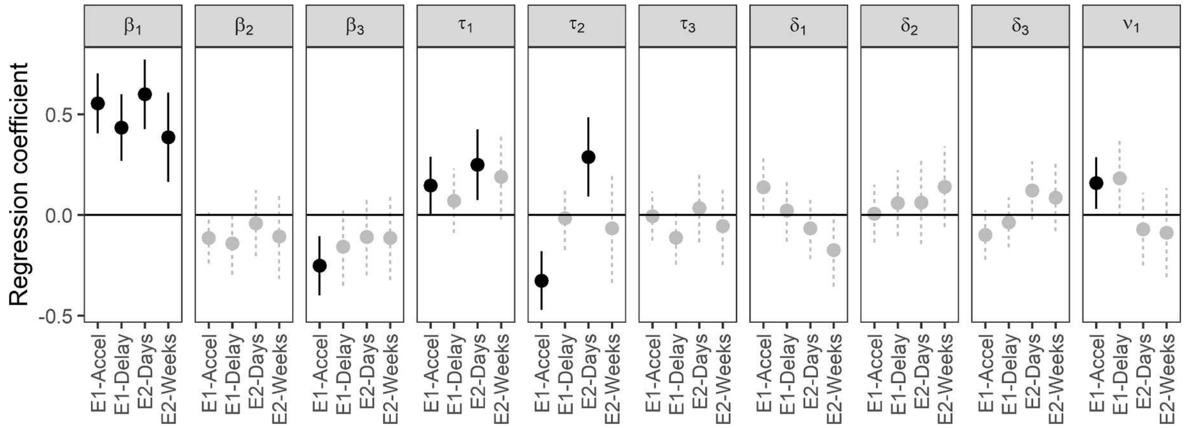

Fig. 5. Regression coefficients from using the bias (β), transition (τ), and dynamic transition (δ) parameters and the novelty parameter (ν) in the attention model to predict SS choice proportions in different experiments and conditions. All independent and dependent variables were standardized in the regression models. Error bars represent 95% confidence intervals. Black dots and bars indicate significance and grey dots and bars indicate insignificance respectively under threshold alpha = 5%.

(corresponding to mostly within-option search), whereas 45.1%, 47.2%, 48.1% and 38.8% of participants had a negative Payne index (corresponding to mostly within-attribute search). Two of the 193 participants in E1-Accel had a Payne index of zero. In short, there was considerable heterogeneity in search tendencies with a nearly even split between within-option versus within-attribute search.

Moreover, consistent with prior work (Reeck et al. 2017), we found a correlation between the Payne index and patience: Participants with higher Payne indices were more likely to choose the SS option, resulting in correlations of 0.37 ($p < .001$; $R^2 = 0.14$), 0.16 ($p = .030$; $R^2 = 0.03$), 0.15 ($p = .077$; $R^2 = 0.02$) and 0.22 ($p = .009$; $R^2 = 0.05$) between participant-level Payne indices and participant-level average SS choice proportions, in the E1-Accel, E1-Delay, E2-Days, and E2-Weeks conditions respectively. Thus, within-option search was associated with less patience.

3.2. Attention Model: Fit summary

3.2.1. Group-level Estimation

The hierarchical Bayesian fitting of the attention model to the MouselabWEB data revealed the rich attention biases, transition properties and dynamics in the decision process. As shown in top panels of Fig. 3, on the group level, participants in the E1-Delay, E2-Days and E2-Weeks conditions tended to search SS more often than LL ($B_1 > 0$). The E1-Accel condition displayed the reverse, searching LL more often than SS. All conditions showed an attribute attention bias towards amounts relative to delays ($B_2 > 0$). Interestingly, there was an attention bias *against* SS amounts in the E1-Accel ($B_3 < 0$) while other conditions showed attention biases *towards* SS amounts ($B_3 > 0$). This result was likely due to the acceleration treatment that highlighted the LL amounts as the default, leading more attention towards the LL amounts in the E1-Accel condition.

Parameters T_1 through T_3 captured the transitions between different attentional states. On the group level, participants tended to make within-option transitions ($T_1 > 0$), as well as within-attribute transitions ($T_2 > 0$), relative to random search. But they were unlikely to resample the same state in two consecutive steps ($T_3 < 0$).

Parameters Δ_1 through Δ_3 further captured how these transition tendencies varied as functions of the currently sampled attribute value. Here, we again observed opposite patterns in the E1-Accel condition and the other three conditions. In the E1-Accel condition, high values increased within-option ($\Delta_1 > 0$) and within-attribute transitions ($\Delta_2 > 0$) but decreased within-state resampling ($\Delta_3 < 0$). Conversely, in the other three conditions, high values decreased within-option ($\Delta_1 < 0$) and within-attribute transitions ($\Delta_2 < 0$) but increased within-state resampling ($\Delta_3 > 0$). This divergence across conditions may be due to a sensitivity to the initially sampled piece of information: the first box-opening for the E1-Accel condition was usually the Larger Later amount; the first opening for the other conditions was usually the Smaller Sooner amount. The value of the initially sampled piece of information may have downstream consequences on attention and choice (indeed, participants were more likely to choose the Larger Later amount in this condition).

The group-level novelty parameter (N_1) was positive in all conditions, suggesting that participants displayed strong tendencies to search for information that they had not seen previously. This result is consistent with the relative performance of the main attention model in compared to nested models without the novelty predictor.

3.2.2. Individual-level Estimation

Hierarchical Bayesian model fitting also captured search tendencies on the individual level. We established the links between attentional parameters and the various process summary statistics. Primarily we expected τ_1 and τ_2 to be strongly related to the Payne index, P. To test this, we calculated a model-based Payne index, $M_P = \tau_1 - \tau_2$. As in Fig. 4A, the correlations between the participant-level model-based Payne Index, M_P, and classic Payne Index, P, were extremely high for all four conditions. There is not an identity

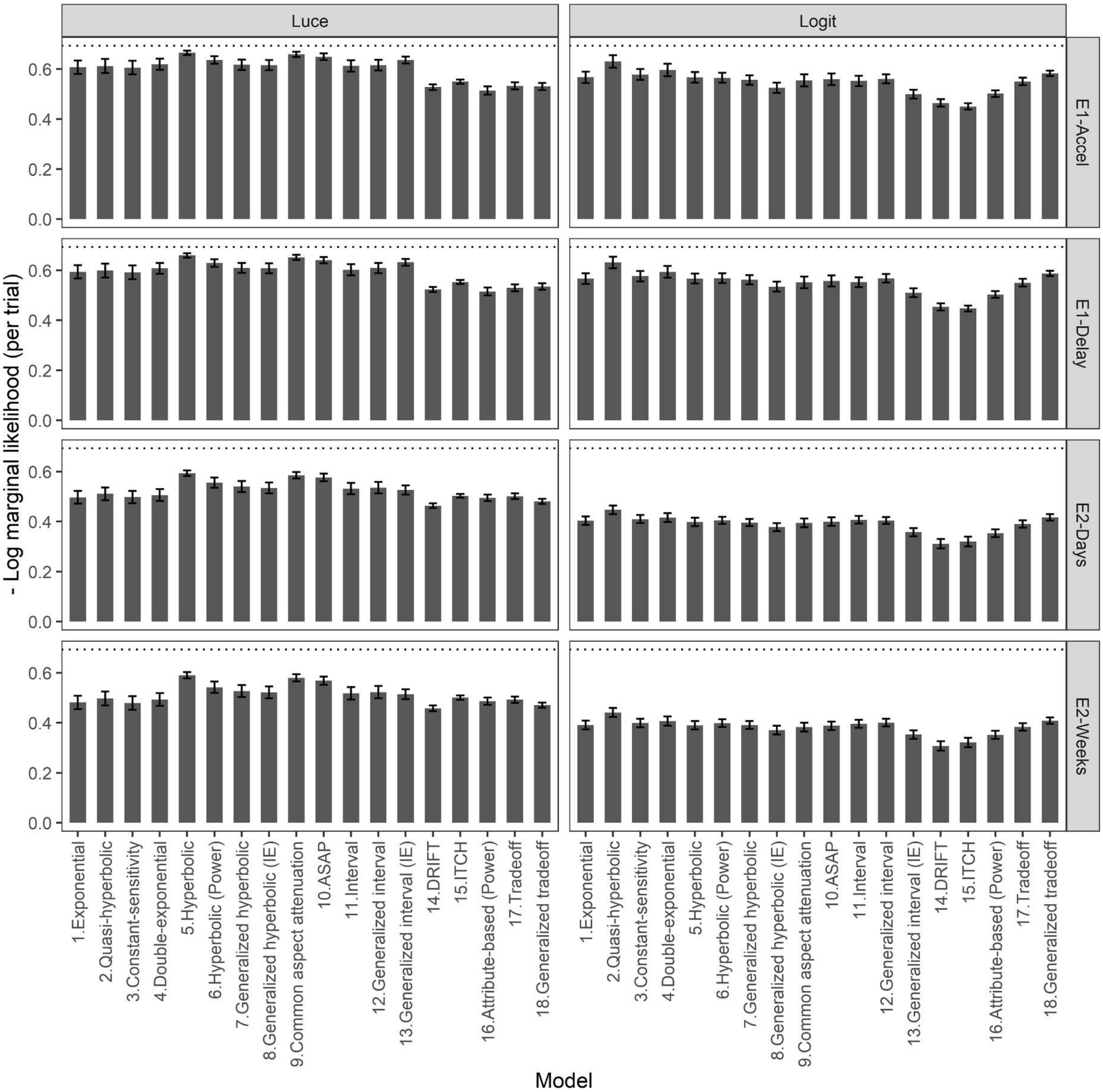

Fig. 6. Log marginal likelihood per trial as predicted by each choice model. All original log marginal likelihood values are reversed for display. Error bars indicate 95% confidence intervals across individuals. The dotted lines indicate the baseline random choice model, with the value at -log(0.5).

relationship between M_P and P because M_P controls for various other search tendencies.

As shown in Fig. 4B, the estimated β_1 in the attention model captured individual-level differences in the option index, with all Spearman $\rho > 0.73$. The correlations between the estimated β_2 and the attribute index were also highly significant, although the Spearman ρ's were not extremely high (Fig. 4C).

3.3. Attention Model: Parameters predict patience

Considering Reeck et al.'s (2017) findings that search tendencies such as the classic Payne index were correlated with patience level, we sought to predict individual-level SS choice proportions using the complete sets of parameters in the attention model in the four conditions, respectively. To account for overfitting, we used a ridge regression with leave-one-out cross-validation to predict participant's patience: the set of attentional parameters predicted participants' patience level successfully, with out-of-sample R^2 equal to 0.37, 0.31, 0.39, and 0.14 in the E1-Accel, E1-Delay, E2-Days, and E2-Weeks conditions respectively. Recall that the Payne index predicted participants' patience with an R^2 of only 0.14, 0.03, 0.02, and 0.05 in the four conditions, indicating that the attention model parameters provided a much better account of choice tendencies than the simple measurement of the Payne index.

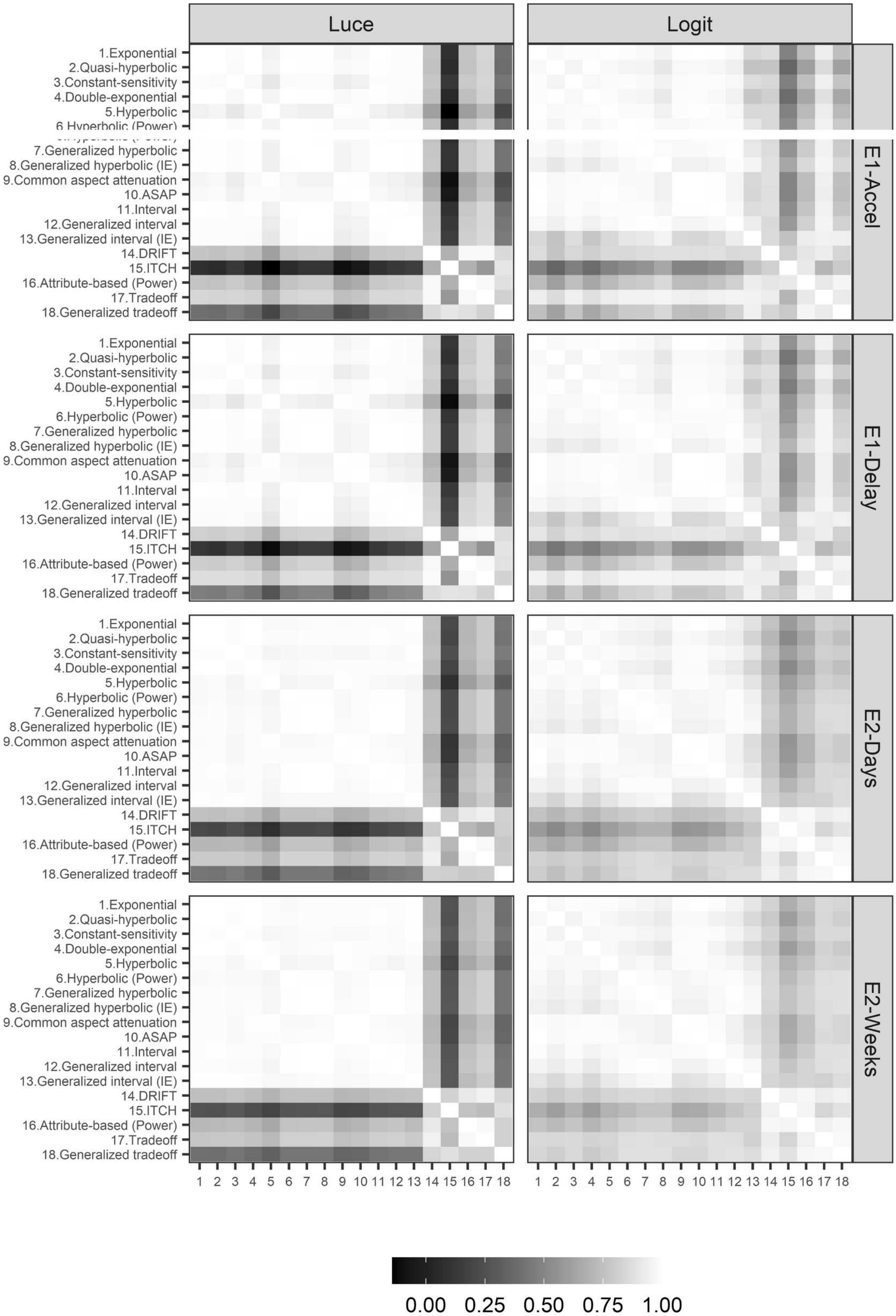

Fig. 7. Pairwise Pearson's R between choice models' fits to individual-level choice data. Models #1–13 are discounting models and Models #14–18 are time-as-attribute models.

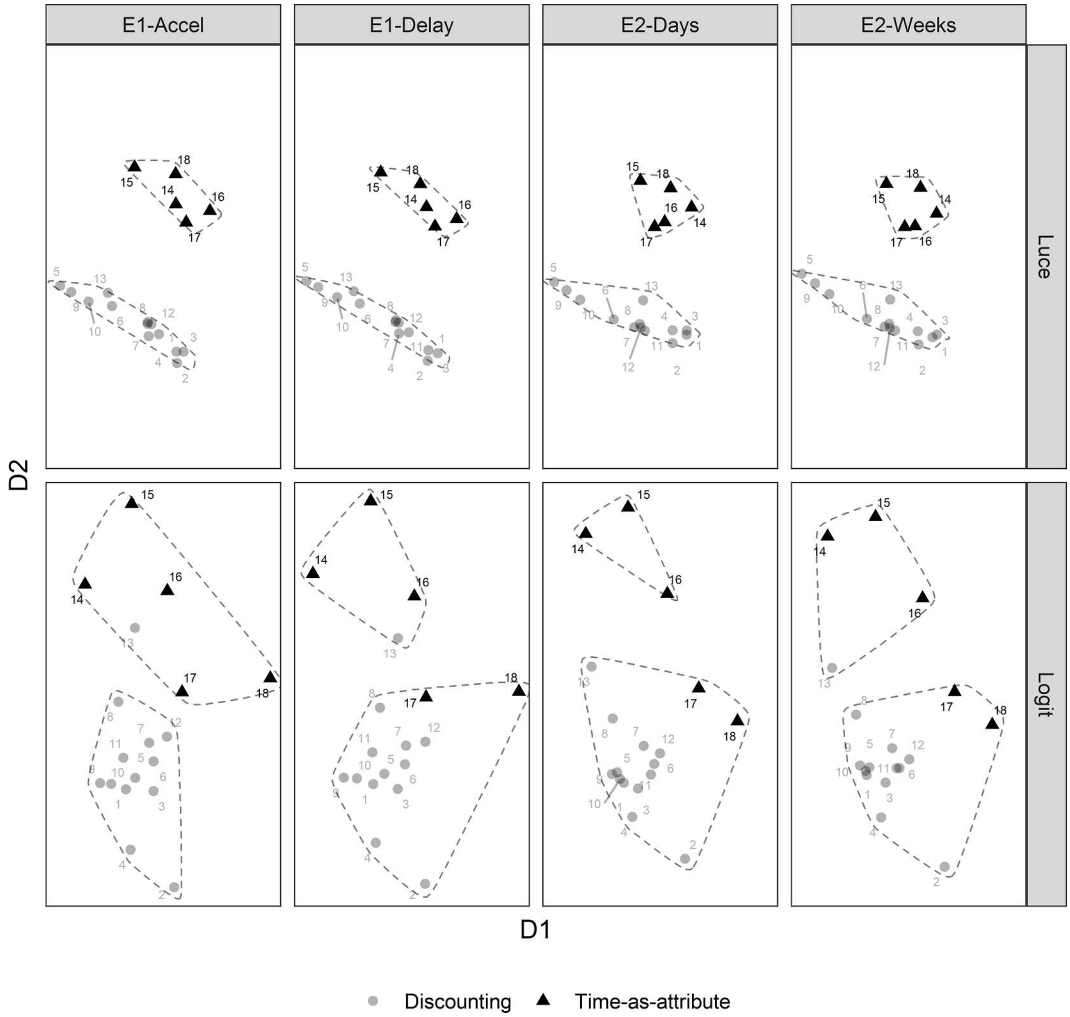

Fig. 8. Data-driven model clusters using k-means clustering (k = 2). The models are arranged per the classic multidimensional scaling (MDS) solution on the N-dimensional representation of models using participant-level marginal log-likelihoods. The two axes (D1 and D2) are the MDS outputs. The dashed enclosing circles indicate the cluster affiliation from k-means clustering. Model names corresponding to the model IDs in the figure can be found in Table 1 and Fig. 7.

The sets of regression coefficients and corresponding 95% confidence intervals are shown in Fig. 5. In all conditions, β_1, the model-based option index, reliably predicted SS choice proportions. The more attention paid to the SS option, the more likely SS was to be chosen. The model-based attribute index, β_2, also predicted SS choice proportions, although it did not reach the conventional significance threshold in two conditions. Participants with more attention towards the amount attribute were less likely to choose the SS options. Consistent with Reeck et al. (2017), within-option transitions (τ_1) led to more SS choices while within-attribute transitions (τ_2) led to more LL choices.

3.4. Choice model types and model fits

After establishing the relationship between search and the revealed patience in the choice data, we compared the performance of different choice models, recovering decision strategies simply from the choice data. We fit 18 choice models to participant-level choice data using a Bayesian method, using both Logit and Luce specifications, respectively. As shown in Table 1, time-as-attribute models typically outperformed delay and interval discounting models. With Logit as the stochastic specification, 74.6%, 79.3%, 83.4%, and 81.3% of participants were best-fit by a time-as-attribute model in the E1-Accel, E1-Delay, E2-Days, and E2-Weeks conditions, respectively. The best performing model (in terms of the aggregate marginal log-likelihood or aggregate MLL) in Experiment 1 (both E1-Accel and E1-Delay) was the Intertemporal Choice Heuristics (ITCH) model (Ericson et al., 2015), which had an MLL of −1548 and best described 57.0% of participants in E1-Accel and had an MLL of −1546 and best described 57.5% of participants in E1-Delay. Likewise, the best performing model in Experiment 2 (both E2-Days and E2-Weeks) was Difference-Ratio-Interest-Finance-Time (DRIFT) model (Read et al., 2013), which provided an MLL of −1223 and the best-fit for 45.9% of participants in E2-Days and

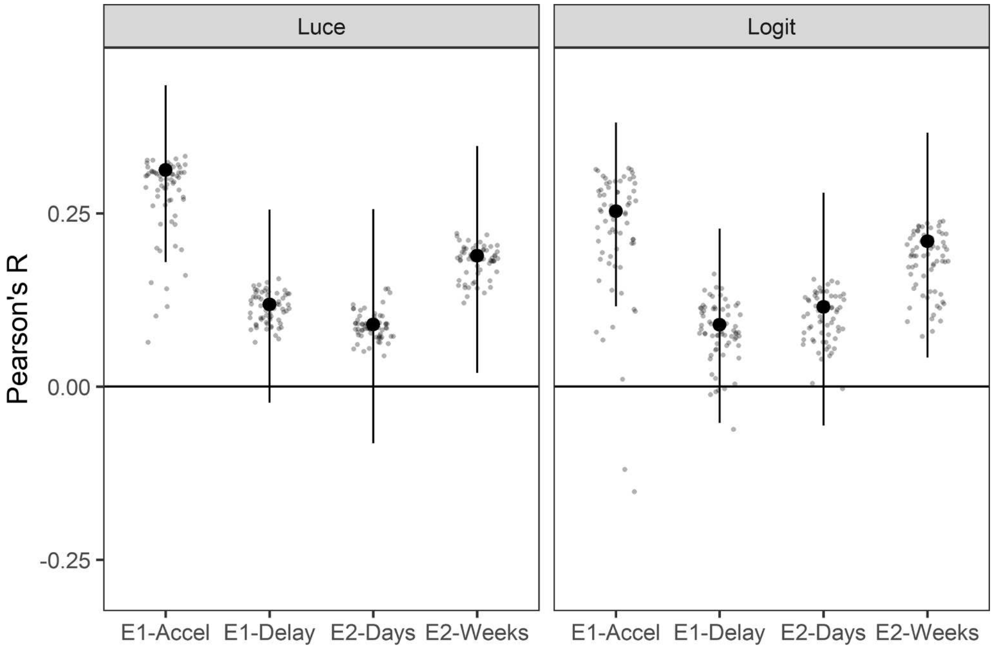

Fig. 9. Pearson's R between model-based Payne index M_P and log Bayes factor (logBF) between a discounting model and a time-as-attribute model. The prominent dots (and ranges) represent the Pearson's R (and 95% confidence intervals) between M_P and $logBF_{DA}$, the logBF between the average discounting model and the average time-as-attribute model. The 65 smaller jitter points within each panel represent the Pearson's Rs between M_P and $logBF_{ij}$, the latter of which is the log Bayes factor between a discounting model i and a time-as-attribute model j.

provided an MLL of −1219 and the best-fit for 44.0% of participants in E2-Weeks. In aggregate, the best performing model was DRIFT, which best described 29.7% of participants and had an MLL of −5607.

Similarly, with the Luce choice rule, 77.7%, 76.7%, 52.6%, and 52.2% of participants were best fit by a time-as-attribute model in the E1-Accel, E1-Delay, E2-Days, and E2-Weeks conditions, respectively. The best performing model (in terms of marginal log-likelihood) was the attribute-based model with a power function (Dai & Busemeyer, 2014) for Experiment 1 (both conditions) and the DRIFT model for Experiment 2 (both conditions). The advantage of the Attribute-based (Power) model in Experiment 1 was mostly driven by particularly good fits to a small proportion of participants (as it was the best fitting model only by 3% to 4% of the participants). In aggregate, the best performing model was DRIFT, which had an MLL of −7273 and provided the best fit to 25.6% of participants. Note that 79.1% of individual-level model fits were better with Logit than with Luce in pairwise comparisons between the two. Still, the two stochastic specifications produced highly correlated model fits. If one participant was better fit by model A than model B using Logit, she was likely better fit by A than B using Luce as the stochastic specification (mean Pearson's $R = 0.77$, ranging from 0.39 to 0.98; mean Spearman's $rho = 0.67$, ranging from 0.26 to 0.97).

Note that all the 18 intertemporal choice models, either discounting or time-as-attribute, provided reasonably good fit to the choice data. In Fig. 6, we compared the models' marginal likelihood per trial with that of a baseline model that predicts purely random choice (i.e., choice probability of 0.5). and found that all the substantive models outperformed the baseline model with confidence in all experiments and conditions, with either the Luce or Logit stochastic specification. On the individual level, the models provided a better fit to a majority of the participants (with Logit specification: 94% on average, ranging from 66% to 100%; with Luce specification: 80% on average, ranging from 66% to 99%).

We tested for systematic relationships between model fits for the 18 models in two ways. First, we used pairwise correlations between the participant-level marginal log-likelihood (MLL) values of each pair of models in the four conditions. Correlations between pairs of models are always positive, reflecting the fact that some participants' choices are easier to model than others. Intuitively participants behaving in a very noisy manner would have low MLL values for all models fit to their data. More importantly, discounting models were highly correlated with each other (mean Pearson's $R = 0.98$ with Logit, $R = 0.96$ with Luce) and time-as-attribute models were highly correlated with each other (mean Pearson's $R = 0.91$ with Logit, $R = 0.83$ with Luce). In contrast, discounting models were only moderately correlated with time-as-attribute models (mean Pearson's $R = 0.79$ with Logit, $R = 0.58$ with Luce) (see Fig. 7). These correlations suggest that the models tested in this paper can be clustered into two main groups depending on whether they involve discounting or time-as-attribute comparisons, and it does not matter much whether the discounting model is delay- or interval-based. In addition, the correlations between discounting models and time-as-attribute models were lower with the Luce rule than with the Logit rule, consistent with the argument that the Luce rule distinguishes the two categories of models more sharply than the Logit rule does (Scholten et al., 2014). Similar patterns emerged when Spearman's rho was used for calculating the correlations between the models' MLL to individual-level choice data (see Fig. A1 in Appendix A for visualization).

We tested the above claims more rigorously using a clustering analysis. For this purpose, we placed each model in an N-dimensional

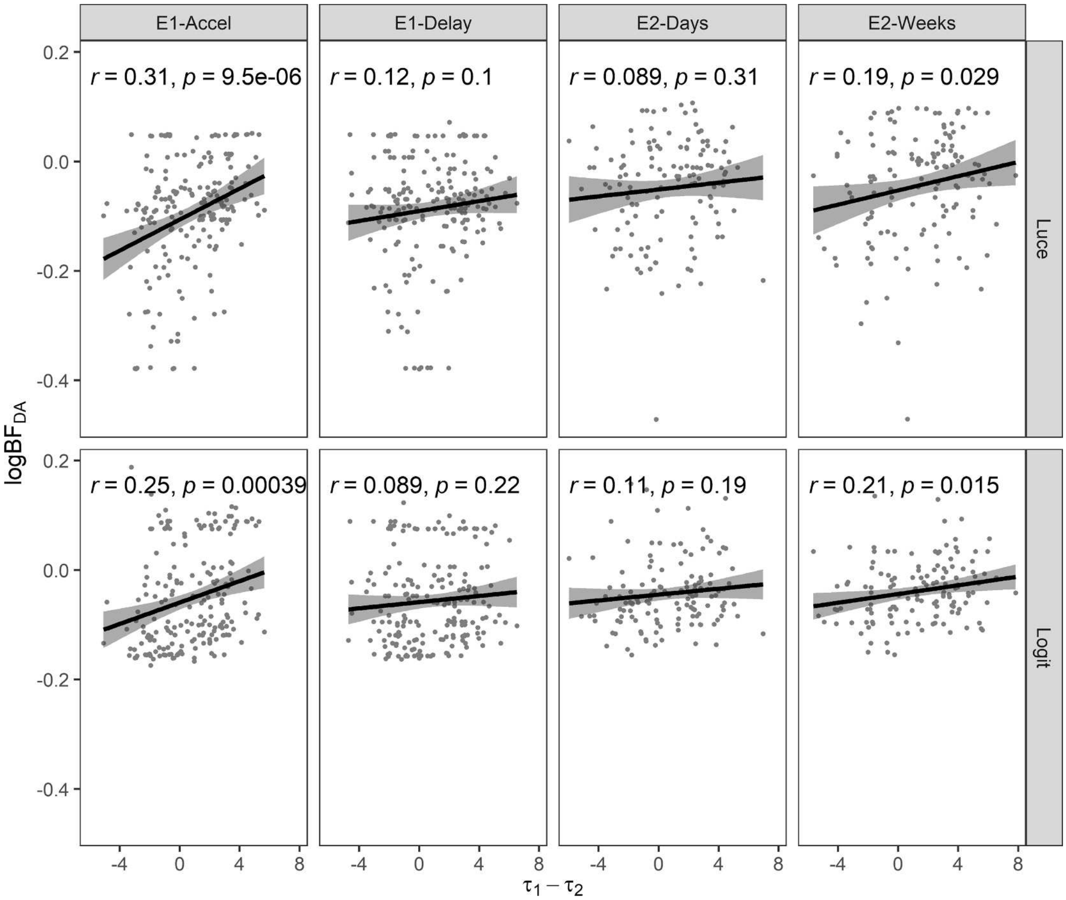

Fig. 10. Scatterplots of $logBF_{DA}$ and model-based Payne index (i.e., $M_P = \tau_1 - \tau_2$). Each dot represents one participant. To control for the potentially different number of trials per participant (due to the experimental design and missing data), the normalized $logBF_{DA}$ per trial is plotted on the vertical axis.

space, where *N* was the number of participants, and correspondingly, model coordinates were the MLL values of the participants' choice data fit to the model, in each condition and for each stochastic specification. We then used the k-means clustering to the original *N*-dimensional space, setting $k = 2$, to generate two data-driven clusters of models. The proportions of discounting and time-as-attribute models in each cluster are shown in a two-dimensional model space (generated by classic multidimensional scaling on the *N*-dimensional space) in Fig. 8. As with prior work on simulated participants, most models in the larger cluster were discounting models, whereas most of the models in the smaller cluster were time-as-attribute models (He et al., 2022).

3.5. Attention Model: Process explanations for choice model fits

The above results suggest that, for our data, the models are clustered into two main groups: Discounting models (either delay or interval discounting) and time-as-attribute models. Thus, the natural question is whether there is a relationship between the search processes adopted by the participants and the choice models (discounting or time-as-attribute) that best describe those participants.

We tested this relationship using two different techniques. First, we examined pairwise differences in model fits between each discounting model and each time-as-attribute model using the log Bayes factor (logBF). We write the log Bayes factor between a given discounting model *i* and a given time-as-attribute model *j* as $logBF_{ij}$. With a total of 13 discounting models and five time-as-attribute models, this results in a total of $13 \times 5 = 65$ separate participant-level $logBF_{ij}$ lists. We did this for the four conditions separately, with Logit and Luce as the stochastic specification, respectively, totaling $65 \times 4 \times 2 = 520$ $logBF_{ij}$ lists.

We correlated each of these $logBF_{ij}$ lists with participant-level model-based Payne indices, M_P, for each of the four conditions and each of the stochastic specifications, resulting in a total of 520 Pearson's *R*. Each of these correlations describes the extent to which our model-based Payne index predicts participant-level model fit differences for a given discounting model and a given time-as-attribute model. The jitter points in Fig. 9 show the distribution of these correlations. This figure indicates that the correlations between our model-based Payne index and discounting versus time-as-attribute model fits were almost always positive, regardless of the time-as-attribute or discounting model considered. This pattern also emerges reliably in all conditions, with both stochastic specifications. Thus, participants who employed an option-based search strategy were more likely to exhibit a better fit with discounting models, as

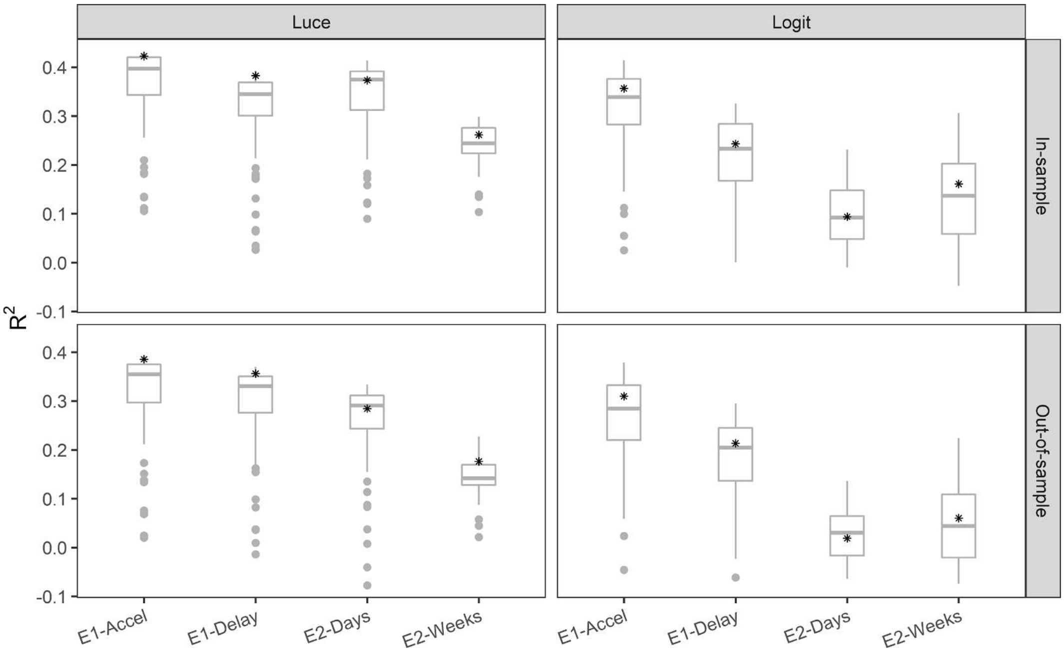

Fig. 11. Distribution of in-sample adjusted R^2 and out-of-sample R^2 in regressions using the attention model parameters to predict the logBF between a discounting model and a time-as-attribute model. The stars represent the R^2 of predicting $logBF_{DA}$, the logBF between the average discounting model and the average time-as-attribute model.

opposed to time-as-attribute models, compared to those who employed an attribute-based search strategy.

Our second test involved averaging the marginal log-likelihoods (MLL) for the discounting and time-as-attribute models to get an average model fit statistic for the two classes of models, respectively. We can write these aggregate fit statistics as MLL_D and MLL_A. Here MLL_D is the list of average participant-level MLLs for the discounting models, and MLL_A is the list of average participant-level MLLs for the set of time-as-attribute models. These lists can be subtracted to get a single list of aggregate participant-level MLL differences, $logBF_{DA} = MLL_D - MLL_A$. $logBF_{DA}$ describes the degree to which participants are better described by the average discounting model vs. the average time-as-attribute model, factoring out each model's idiosyncratic fit. As in the analysis outlined above, we correlated $logBF_{DA}$ with the model-based Payne index, M_P, for each of our four conditions and the two stochastic specifications. These correlations, along with 95% confidence intervals, are shown as the solid points and lines in Fig. 9. Once again, correlations were positive. As with the first test shown above, this indicates that participant-level heterogeneity in within-option vs. within-attribute search was closely related to participant-level heterogeneity in discounting versus time-as-attribute model fits.

In Fig. 10, we further detail the individual-level correspondence between $logBF_{DA}$ and the model-based Payne index (i.e., $M_P = \tau_1 - \tau_2$) in the scatterplots. Again, Fig. 10 shows that the correlations between $logBF_{DA}$ and model-based Payne index were positive in all conditions, regardless of the Luce or Logit choice rule being used for choice modeling. Of course, these two measures were not perfectly matched. On an absolute scale, it appeared that the model-based Payne index categorized an individual as an option-based decision maker (i.e., $M_P > 0$) more often than the $logBF_{DA}$ did (i.e., $logBF_{DA} > 0$). However, their relative tendencies in $logBF_{DA}$ were reasonably well predicted by their search patterns as captured by M_P. That said, the model-based Payne index may not be the only search tendencies that underlie the relative model fit between discounting and time-as-attribute models. Our attention model parameters allowed a more extensive test of the relationships between information search dynamics and decision strategies.

3.6. Attention Model: Parameters predict relative choice model fit

In this section, we tested the overall predictive power of the attention model parameters. We did so by using all the ten attentional parameters and again attempted to predict both pairwise differences in model fits for discounting versus time-as-attribute models ($logBF_{ij}$) as well as differences in aggregate model fits for these models ($logBF_{DA}$). We ran these regressions separately for data from each of our four conditions. The top row of Fig. 11 displays the distribution of the regressions' adjusted R^2 values for the Logit and Luce specifications. As can be seen, the R^2 values from our analysis were mostly quite high. The average adjusted R^2 value across all regressions is 0.30, indicating that, on average, 30% of the variance in model fit differences between discounting and time-as-attribute models was explained purely by differences in process-level data. With the Luce rule, which heightens the distinction between the two model categories (Scholten et al., 2014), the average adjusted R^2 value across all regressions reached as high as 0.36. We also conducted a leave-one-out cross-validation to address potential overfitting, and the distributions of the out-of-sample R^2 showed almost the same patterns as in-samples R^2 (see bottom panels in Fig. 10). Furthermore, the average in-sample and out-of-sample R^2 in predicting $logBF_{DA}$ were 0.33 and 0.23 respectively (see Fig. 11 for a breakdown), indicating, once again, that the differences in model fit

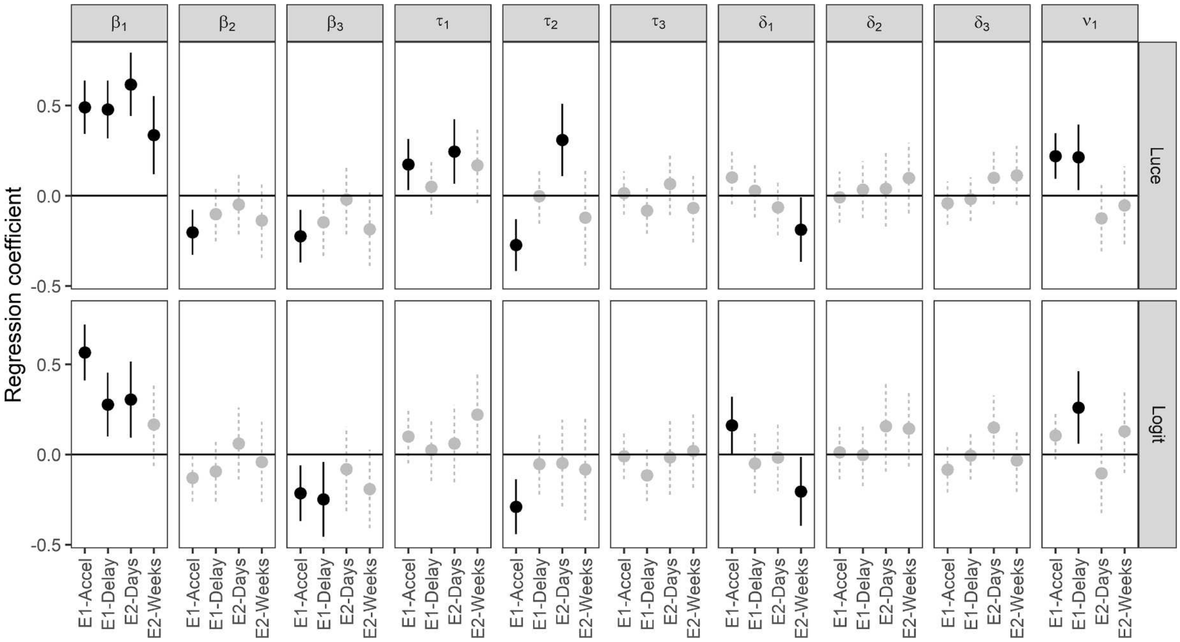

Fig. 12. Regression coefficients from using bias (β), transition (τ), dynamic transition (δ) and novelty (ν) parameters in the attention model to predict the logBF between the average discounting model and the average time-as-attribute model. All independent and dependent variables were standardized in the regression models. Error bars represent 95% confidence intervals. Black dots and bars indicate significance and grey dots and bars indicate insignificance respectively under threshold alpha = 5%.

between discounting and time-as-attribute models can be predicted using process-level data with a decent degree of accuracy.

We further tested which attentional parameters best predicted logBF_{DA} as shown in Fig. 12. In line with the Payne index's explanatory role in the relative model fits between discounting models and time-as-attribute models, τ_1 positively predicted logBF_{DA} while τ_2 negatively predicted logBF_{DA} in all conditions. The pattern was especially strong with the Luce rule. Additionally, β_1 (option bias) and β_2 (attribute bias) also predicted logBF_{DA} in different conditions. Across different stochastic specifications and experimental conditions, participants that were better fit by the discounting models (i.e., $\text{logBF}_{\text{DA}} > 0$) attended to SS on average 53.1% of the time, whereas participants that were better fit by attribute models (i.e., logBF_{DA} less than 0) attended to SS only 49.7% of the time. Likewise, participants that were better fit by time-as-attribute models attended to the monetary amount attribute on average 53.6% of the time, whereas participants that were better fit by discounting models attended to the monetary amount attribute only 49.5% of the time. These patterns emerged in all conditions and reached statistical significance in most of them, particularly when the Luce rule was used for choice modeling (see Fig. 13).

4. Discussion

This paper establishes a formal link between the dynamics of participants' information acquisition and the type of models that best describe their choices. In line with prior research on simulated participants (He et al., 2022), we found that when fit to actual participants' choices, intertemporal choice models clustered into two categories: discounting and time-as-attribute. These two categories were consistent with the information processing assumptions of the various models. Whereas discounting models predict that participants will sequentially search within an option (e.g., look at the smaller sooner amount and then look at the smaller sooner time), time-as-attribute models predict that participants will sequentially search within an attribute (e.g., look at the smaller sooner amount and then look at the larger later amount).

To formally model the dynamics of participants' information acquisition, we built and tested an attention model. Our model provides a comprehensive but tractable representation of the dynamics of participants' attention. We found that the attention model parameters for within-attribute and within-option search strongly predicted whether a participant was better described by a discounting or time-as-attribute model. Additionally, parameters that captured option and attribute biases also helped predict the class of model that a participant belonged to. Overall, these results corroborate our hypothesized connections between attentional processes and choice strategies.

These results have multiple implications for intertemporal choice research and decision making more broadly. First, they show that individual differences in choice model fit are due, in part, to individual differences in information processing. Different people search differently, which is reflected in their choices, as well as the model that best describes them. This in turn implies that some amount of heterogeneity in behavioral outcomes in the real-world may be due to heterogeneity in search. Thus, it may be possible to predict and improve real-world decisions using attentional data. This is an exciting topic for future work. This heterogeneity in search strategies

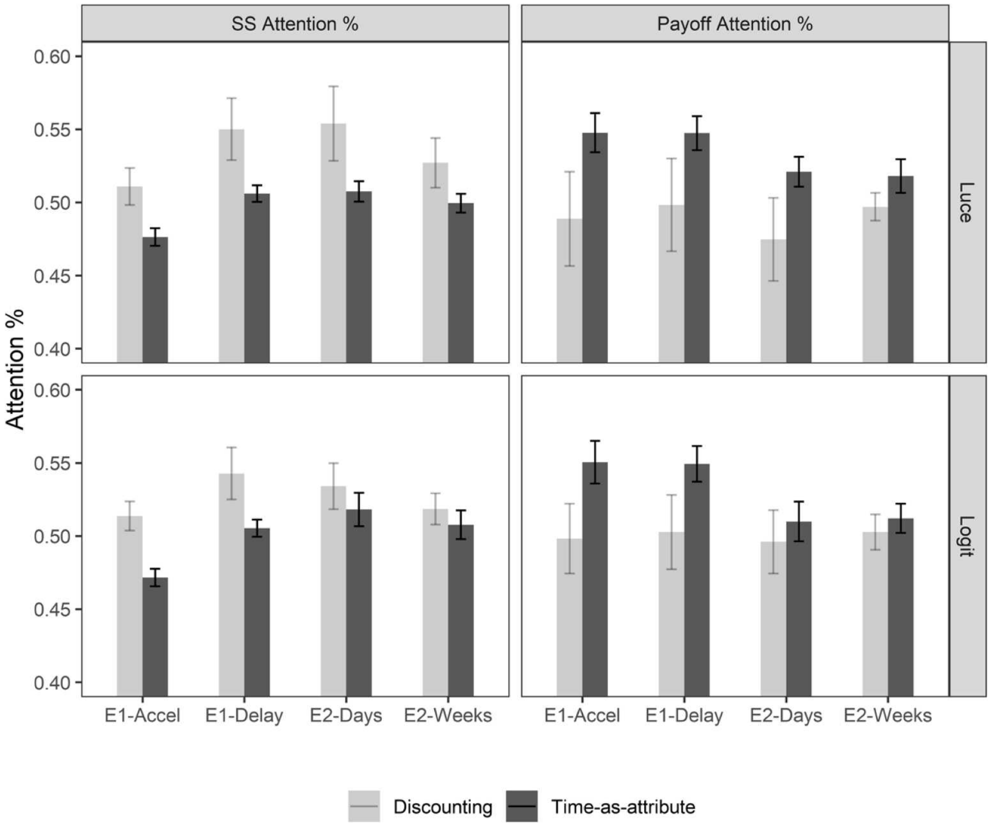

Fig. 13. Proportions of attention allocation by participants whose choices were either more consistent with discounting models or time-as-attribute models, respectively. Error bars represent 95% confidence intervals.

may also be driven by distinct neural mechanisms, suggesting that neuroscientific investigations of intertemporal choice should consider participants' attentional search processes.

Secondly, our results illustrate the importance of a more systematic treatment of attention during decision making. While prior work has related single number indexes of search to choice (Amasino et al., 2019; Reeck et al., 2017), we show that a formal mathematical model of attention significantly increases predictive power. This increase comes from the joint modeling of several determinants of attention, including tendencies for within-attribute and within-option search, as well overall attentional biases towards SS/LL options or payoff/delay attributes.

Our results relate to recent research on the role of attention in a closely related domain, risky choice (Pachur et al. 2018; Zilker & Pachur, 2021). This research finds that the weighting of probabilities—as estimated by Prospect Theory–can arise from biases in how people search for information. We expand on this work in two main ways. First, instead of relating attention to a single theory of choice, we relate attention to multiple theories of choice. That is, instead of examining the effect of attention on model parameters, we examine the link between the various attention biases and model comparisons. Second, Zilker and Pachur (2021) linked Prospect Theory to the attentional Drift Diffusion Model, an important theory of how attention influences decision making (Krajbich et al., 2010; Ratcliff et al., 2016). While this is undoubtedly an important theoretical connection, the attentional Drift Diffusion Model does not predict attention (it only models the consequences of attention on choice). By contrast, our attention model predicts where people attend to.

Of course, future work could combine the outputs of our attentional model with theories of cognitive decision processes (such as those built on the Drift Diffusion framework) to fully characterize information acquisition and decision making in intertemporal choice (see e.g. Zhao et al., 2022). It is currently not clear how a sequence of attentional operations (as in our attentional model) could lead to the emergence of the types of complex utility maximization theories (choice models) that have been proposed previously. We know that this happens in the human mind, but how it happens is a difficult question to solve, and has been a major research question for decades. Although we have not built a single joint model of attention and choice, we have still fleshed out a detailed set of attentional processes in a formal model, and examined the relationship of their parameters with 18 different choice models. This sheds lights on the elements of attentional processes that could give rise to different types of decision making behaviors. We hope that these results will be the basis of further work that identifies how the outputs of the attentional process are integrated in a cognitively plausible decision module to generate choice.

Beyond time and risk, our research on attention is also related to broader work on choice processes. For instance, prior work has

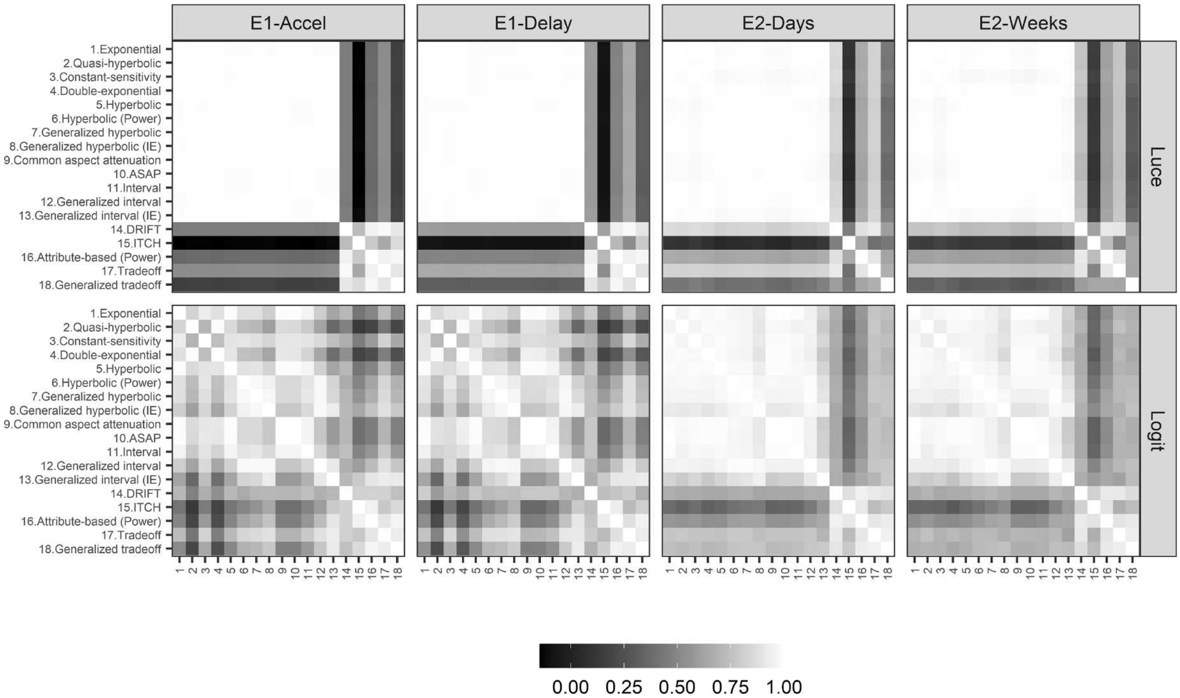

Fig. A1. Pairwise Spearman's rank correlation between choice models' fits to individual-level choice data. Models #1–13 are discounting models and Models #14–18 are time-as-attribute models.

Table A1
Summary of intertemporal choice models.

ID	Model	Category	Authors	Year	Source (Journal or Book)
1	Exponential	Discounting	Samuelson	1937	Review of Economic Studies
2	Quasi-hyperbolic	Discounting	Laibson	1997	Quarterly Journal of Economics
3	Constant-sensitivity	Discounting	Ebert and Prelec	2007	Management Science
4	Double-exponential	Discounting	McClure et al.	2007	Journal of Neuroscience
5	Hyperbolic	Discounting	Mazur	1987	The Effect of Delay and Intervening Events on Reinforcement Value
6	Hyperbolic (Power)	Discounting	Mazur	1987	The Effect of Delay and Intervening Events on Reinforcement Value
7	Generalized hyperbolic	Discounting	Loewenstein et al.	1992	Quarterly Journal of Economics
8	Generalized hyperbolic (IE)	Discounting	Scholten et al.	2014	Cognitive Science
9	Common aspect attenuation	Discounting	Green et al.	2005	Journal of Experimental Psychology: Learning, Memory & Cognition
10	ASAP	Discounting	Kable and Glimcher	2010	Journal of Neurophysiology
11	Interval	Discounting	Read	2001	Journal of Risk and Uncertainty
12	Generalized interval	Discounting	Scholten and Read	2006	Management Science
13	Generalized interval (IE)	Discounting	Scholten et al.	2014	Cognitive Science
14	DRIFT	Time-as-attribute	Read et al.	2013	Journal of Experimental Psychology: Learning, Memory & Cognition
15	ITCH	Time-as-attribute	Ericson et al.	2015	Psychological Science
16	Attribute-based (Power)	Time-as-attribute	Dai and Busemeyer	2014	Journal of Experimental Psychology: General
17	Tradeoff	Time-as-attribute	Scholten and Read	2010	Psychological Review
18	Generalized tradeoff	Time-as-attribute	Scholten et al.	2014	Cognitive Science

shown that attention is a fundamental driver of value construction (Smith & Krajbich, 2019). This work relies on drift–diffusion models and fits to response time data, and could be combined with our attentional model to better account for the underlying dynamics of information acquisition. In this way we could describe not only what gets attended to, but also how information sampling drives preferences and eventual decisions.

There exists other experimental work that has examined the relationships between information search and choice strategies by external manipulations of the participants' decision strategies (Arieli et al., 2011; Fisher, 2021; Liu et al., 2021; Zhang et al., 2022). For example, Zhang et al. (2022) present an experiment that explicitly instructed the participants to make intertemporal choices by

Table A2
Functional forms of intertemporal choice models. The notations are designed for choices between $X = (\$x, t)$ and $Y = (\$y, s)$, where $y > x > 0$, $s > t \geq 0$. $U(X)$ denotes the utility or choice propensity of X and $U(Y)$ denotes the utility or choice propensity of Y. For delay discounting models, $U(Y)$ is not presented but can be obtained by replacing x and t in $U(X)$ with y and s. Free parameters are denoted by Greek letters, with corresponding domains and prior distributions shown in Table A3.

ID	Model	Function
1	Exponential	$U(X) = \delta^t u(x)$
2	Quasi-hyperbolic	$U(X) = \begin{cases} \delta^t u(x) = u(x), when t = 0 \\ \mu\delta^t u(x), when t > 0 \end{cases}$
3	Constant-sensitivity	$U(X) = e^{-(\beta t)^{\alpha}} u(x)$
4	Double-exponential	$U(X) = (\omega\delta^t + (1-\omega)\tau^t)u(x)$
5	Hyperbolic	$U(X) = \frac{u(x)}{1+\alpha t}$
6	Hyperbolic (Power)	$U(X) = \frac{u(x)}{1+\alpha t^{\beta}}$
7	Generalized hyperbolic	$U(X) = \frac{u(x)}{(1+\alpha t)^{\beta/\alpha}}$
8	Generalized hyperbolic (IE)	$U(X) = \frac{u_{SRS}(x)}{(1+\alpha t)^{\beta/\alpha}}$
9	Common aspect attenuation	$U(X) = \frac{u(x)}{1+\alpha\mu t}$ $U(Y) = \frac{u(y)}{1+\alpha(\mu t + (s-t))}$
10	ASAP	$U(X) = \frac{u(x)}{1+\alpha t}$ $U(Y) = \frac{u(y)}{(1+\alpha t)\cdot(1+\alpha(s-t))}$
11	Interval	$U(X) = u(x)\delta^{t^{\tau}}$ $U(Y) = u(y)\delta^{t^{\tau}+(s-t)^{\tau}}$
12	Generalized interval	$U(X) = \frac{u(x)}{(1+\alpha t^{\tau\zeta})^{\beta/\alpha}}$ $U(Y) = \frac{u(y)}{((1+\alpha t^{\tau\zeta})(1+\alpha(s^{\tau}-t^{\tau})^{\zeta}))^{\beta/\alpha}}$
13	Generalized interval (IE)	$U(X) = \frac{u_{SRS}(x)}{(1+\alpha t^{\tau\zeta})^{\beta/\alpha}}$ $U(Y) = \frac{u_{SRS}(y)}{((1+\alpha t^{\tau\zeta})(1+\alpha(s^{\tau}-t^{\tau})^{\zeta}))^{\beta/\alpha}}$
14	DRIFT	$U(X) = \kappa(s-t)$ $U(Y) = \tau\left(\left(\frac{y}{x}\right)^{\frac{1}{s-t}} - 1\right) + (1-\tau)\gamma\frac{y-x}{x} + (1-\tau)(1-\gamma)(y-x)$
15	ITCH	$U(X) = \kappa\left(\tau(s-t) + (1-\tau)\frac{2(s-t)}{s+t}\right)$ $U(Y) = \gamma(y-x) + (1-\gamma)\frac{2(y-x)}{y+x}$
16	Attribute-based (Power)	$U(X) = \kappa(s^{\tau} - t^{\tau})$ $U(Y) = y^{\gamma} - x^{\gamma}$
17	Tradeoff	$U(X) = \frac{\kappa}{\beta}(log(1+\beta s) - log(1+\beta t))$ $U(Y) = \frac{1}{\alpha}(log(1+\alpha y) - log(1+\alpha x))$
18	Generalized tradeoff	$U(X) = \frac{\kappa}{\lambda}\log\left(1 + \lambda\left(\frac{\frac{1}{\beta}(\log(1+\beta s) - \log(1+\beta t))}{\zeta}\right)^{\zeta}\right)$ $U(Y) = \frac{1}{\alpha}(\log(1+\alpha y) - \log(1+\alpha x))$

Note. The following additional functions are used in Table A2:

- $I(\cdot)$ is an indicator function that returns 1 if the argument is true and 0 otherwise.
- Power value function: $u(x) = x^{\gamma}$.
- Increasingly elastic value function as in Scholten et al. (2014): $u_{SRS}(x) = (1-\omega)x^{1-\omega} + \lambda\omega x^{\omega}$.

calculating the subjective values of each option. They found that participants following such an "calculation" instruction produced more attribute-based saccades than those in a typical, no-instructed intertemporal choice task. This indicates that calculations of subjective value (as required by discounting models) are more likely to be generated by within-attribute calculations rather than within-option calculations. While this claim seemingly contradicts our results, as well as common intuitions about the information processing assumptions underlying time-as-attribute and discounting based choice strategies (Johnson et al., 2008; Johnson & Payne, 1985), it could be explained by differences between ours and Zhang et al.'s (2022) experimental setups. For example, while Zhang et al.

Table A3
Parameter bounds and prior distributions for intertemporal decision models.

Parameter	Domain	Prior distribution
$\alpha, \beta, \varepsilon, \kappa, \lambda, \rho, (\zeta - 1)$	$[0, +\infty)$	Exponential (rate = 1)
$\delta, \gamma, \mu, \tau$	[0, 1]	Uniform(0, 1)
ω	[0.5, 1]	Uniform(0.5, 1)

Note. $(\zeta - 1)$ has the domain of $[0, +\infty)$, meaning that the domain of ζ is $[1, +\infty)$.

(2022) examined the differences in cognitive processing due to explicit task instructions, we study individual differences in natural intertemporal choice tasks. The instructed condition in Zhang et al (2022) was likely qualitatively different than a typical intertemporal choice task without unnatural instructions. Specifically, this condition required participants to calculate the future value of an amount of money –either 1 month or 2 months into in the future. For instance, participants had to choose whether ¥100 today or ¥111 in 1 month had the higher value given a hyperbolic monthly interest rate of 10% (¥111 in 1 month is higher than the appreciated ¥100: ¥110). While this manipulation aimed to encourage within-option transitions, as the authors noted, it may not necessarily do so. Their eye-movement data suggested there might exist other ways a participant could solve this problem. Bridging these strands of work presents a promising direction for future research.

What distinguished our work from others was the link between the information search processes and the computational modeling of intertemporal choice at the *individual* level. Over the last few decades, researchers have developed formal quantitative discounting and time-as-attribute models and fit them to the choice data to identify the type of the intertemporal decision maker. This has made our formal link of natural information search processes and decision strategies described above possible. Of course, our modeling framework was not flawless. For example, previous work has suggested that the stochastic specification plays an important role in the explanatory and predictive scope of models (Dai & Busemeyer, 2014; He et al., 2019; He et al., 2022; Regenwetter et al., 2018), and is arguably integral to formal decision models (Loomes & Sugden, 1995). We were not able to exhaust all possible stochastic specifications, such as the random utility models with a Probit rule (Cheng & González-Vallejo, 2016) and the accumulators that specify choice probabilities as a function of the accumulation process (Dai & Busemeyer, 2014), for a complete evaluation of choice models. Still, we applied the two commonly used specifications to establish the robustness of the link between information search and decision strategies in intertemporal choice. Additionally, it is likely that our attention model is incomplete and more complex processes can be included. In particular, we suspect that there may be some type of non-Markovian property (beyond a novelty bias) to the attentional process, so that there is an effect of items sampled at t on attention at $t + k$ (for $k > 1$). This could be modeled as decay in memory (see Zhao et al., 2022) which would predict that this dynamic effect gets weaker as k gets larger. A comprehensive evaluation of these modeling choices warrants further investigation (see He et al., 2022 for an example).

Also, individual differences in attention and choice model fit may be due to other, unmeasured factors. For instance, people with high numeracy or cognitive ability may find it easier to perform the math required for some discounting models and, therefore, may be more likely to use discounting models than people with lower numeracy (Frederick, 2005; Peters et al., 2006). These people may also display unique attentional sampling patterns, resulting in systematically different parameters in our attention model. Related work in risky choice finds that how easy attributes are to compare affects search, further suggesting that ability may affect search. That said, exogenously manipulating attention has been shown to affect patience, suggesting a causal connection between attention and choice (Reeck et al., 2017). This causal connection implies that by manipulating attention, we may shift which model best describes a person's choices. Testing such manipulations, to better understand the causal role of attention, is an important topic for future work.

Overall, our work highlights the relationships between attentional dynamics and choice. It shows that choice models cluster into two main categories, and that the choice data's fits to these models depend on decision-makers' attentional dynamics. Understanding this relationship is necessary for psychologically informed choice models, which, in turn, are essential to interventions designed to help people make better decisions.

5. Author Note

L.H. was supported by the National Natural Science Foundation of China (72101156) and Shanghai Pujiang Program (2020PJC102). S.B. was supported by the National Science Foundation (SES-1847794). We appreciate the High Performance Computing Center of Shanghai University, and Shanghai Engineering Research Center of Intelligent Computing System (No.19DZ2252600) for providing the computing resources and technical support. The funding sources have no involvement in the study design, data analysis and manuscript preparation.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

Appendix A

See Fig. A1 and Tables A1, A2, and A3.

Appendix B. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.cogpsych.2023.101562.

References

Amasino, D. R., Sullivan, N. J., Kranton, R. E., & Huettel, S. A. (2019). Amount and time exert independent influences on intertemporal choice. *Nature Human Behaviour, 3*(4), 383–392. https://doi.org/10.1038/s41562-019-0537-2

Andraszewicz, S., Scheibehenne, B., Rieskamp, J., Grasman, R., Verhagen, J., & Wagenmakers, E.-J. (2015). An introduction to Bayesian hypothesis testing for management research. *Journal of Management, 41*(2), 521–543.

Arieli, A., Ben-Ami, Y., & Rubinstein, A. (2011). Tracking decision makers under uncertainty. *American Economic Journal: Microeconomics, 3*(4), 68–76. https://doi.org/10.1257/mic.3.4.68

Bhatia, S., Loomes, G., & Read, D. (2021). Establishing the laws of preferential choice behavior. *Judgment and Decision Making, 16*(6), 1324–1369.

Callaway, F., Rangel, A., & Griffiths, T. L. (2021). Fixation patterns in simple choice reflect optimal information sampling. *PLoS computational biology, 17*(3), e1008863.

Cheng, J., & González-Vallejo, C. (2016). Attribute-wise vs. alternative-wise mechanism in intertemporal choice: Testing the proportional difference, trade-off, and hyperbolic models. *Decision, 3*(3), 190.

Dai, J. (2017). Are intertemporal preferences transitive? A Bayesian analysis of repeated individual intertemporal choices. *Decision, 4*(1), 1–24. https://doi.org/10.1037/dec0000054

Dai, J., & Busemeyer, J. R. (2014). A probabilistic, dynamic, and attribute-wise model of intertemporal choice. *Journal of Experimental Psychology: General, 143*(4), 1489–1514. https://doi.org/10.1037/a0035976

Doyle, J. R. (2013). Survey of Time Preference, Delay Discounting Models. *Judgment & Decision Making, 8*(2), 116–135. https://doi.org/10.2139/ssrn.1685861

Ebert, J. E., & Prelec, D. (2007). The fragility of time: Time-insensitivity and valuation of the near and far future. *Management Science, 53*(9), 1423–1438.

Ericson, K., White, J., Laibson, D., & Choen, J. (2015). Money Earlier or Later? Simple Heuristics Explain Intertemporal Choices Better than Delay Discounting. *Psychological Science, 26*(6), 826–833.

Fiedler, S., & Glöckner, A. (2012). The Dynamics of Decision Making in Risky Choice: An Eye-Tracking Analysis. *Frontiers in Psychology, 3*. https://doi.org/10.3389/fpsyg.2012.00335

Fisher, G. (2021). Intertemporal choices are causally influenced by fluctuations in visual attention. *Management Science, 67*(8), 4961–4981.

Frederick, S., Loewenstein, G., & Donoghue, T. O. (2002). Time Discounting and Preference: A Critical Review. *Journal of Economic Literature, 40*(2), 351–401. https://doi.org/10.1126/science.151.3712.867-a

Frederick, S. (2005). Cognitive reflection and decision making. *Journal of Economic Perspectives, 19*(4), 25–42.

Fong, E., & Holmes, C. C. (2020). On the marginal likelihood and cross-validation. *Biometrika, 107*(2), 489–496.

Gelman, A., & Rubin, D. B. (1992). Inference from iterative simulation using multiple sequences. *Statistical Science, 7*(4), 457–472.

Gluth, S., Kern, N., Kortmann, M., & Vitali, C. L. (2020). Value-based attention but not divisive normalization influences decisions with multiple alternatives. *Nature human behaviour, 4*(6), 634–645.

Golsteyn, B. H. H., Grönqvist, H., & Lindahl, L. (2014). Adolescent Time Preferences Predict Lifetime Outcomes. *The Economic Journal, 124*(580), F739–F761. https://doi.org/10.1111/ecoj.12095

Green, L., Myerson, J., & Macaux, E. W. (2005). Temporal discounting when the choice is between two delayed rewards. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 31*(5), 1121–1133.

He, L., Golman, R., & Bhatia, S. (2019). Variable time preference. *Cognitive Psychology, 111*, 53–79.

He, L., Zhao, W. J., & Bhatia, S. (2022). An ontology of decision models. *Psychological review, 129*(1), 49–72.

Jekel, M., Glöckner, A., & Bröder, A. (2018). A new and unique prediction for cue-search in a parallel-constraint satisfaction network model: The attraction search effect. *Psychological review, 125*(5), 744–768.

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

Johnson, E. J., Schulte-Mecklenbeck, M., & Willemsen, M. C. (2008). Process models deserve process data: Comment on Brandstätter, Gigerenzer, and Hertwig (2006). *Psychological Review, 115*(1), 263–273. https://doi.org/10.1037/0033-295X.115.1.263

Kable, J. W., & Glimcher, P. W. (2010). An "As Soon As Possible" Effect in Human Intertemporal Decision Making: Behavioral Evidence and Neural Mechanisms. *Journal of Neurophysiology, 103*(5), 2513–2531. https://doi.org/10.1152/jn.00177.2009

Kass, R., & Raftery, A. (1995). Bayes Factors. *Journal of the American Statistical Association, 90*(430), 773–795.

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.

Laibson, D. (1997). Golden eggs and hyperbolic discounting. *The Quarterly Journal of Economics, 112*(2), 443–478.

Lee, B., Reeck, C., & Johnson, E. (In preparation). Two processes of patience: Information search shapes consumer patience.

Li, Y., Krefeld-Schwalb, A., Wall, D. G., Johnson, E. J., Toubia, O., & Bartels, D. M. (2021). EXPRESS: The More You Ask, the Less You Get: When Additional Questions Hurt External Validity. *Journal of Marketing Research*, 00222437211073581.

Liu, H. Z., Lyu, X. K., Wei, Z. H., Mo, W. L., Luo, J. R., & Su, X. Y. (2021). Exploiting the dynamics of eye gaze to bias intertemporal choice. *Journal of Behavioral Decision Making, 34*(3), 419–431.

Loewenstein, G., & Prelec, D. G. (1992). Anomalies in intertemporal choice: Evidence and an interpretation. *The Quarterly Journal of Economics, 107*(2), 573–597. https://doi.org/10.2307/2118482

Loewenstein, G. F. (1988). Frames of Mind in Intertemporal Choice. *Management Science, 34*(2), 200–214. https://doi.org/10.1287/mnsc.34.2.200

Loomes, G., & Sugden, R. (1995). Incorporating a stochastic element into decision theories. *European Economic Review, 39*(3–4), 641–648.

Mazur, J. E. (1987). An adjusting procedure for studying delayed reinforcement. In M. L. Commons, J. E. Mazur, J. A. Nevin, & H. Rachlin (Eds.), *The effect of delay and of intervening events on reinforcement value: quantitative analyses of behavior* (pp. 55–73). Hillsdale, NJ: Lawrence Erlbaum.

McClure, S. M., Ericson, K. M., Laibson, D. I., Loewenstein, G., & Cohen, J. D. (2007). Time discounting for primary rewards. *Journal of Neuroscience, 27*(21), 5796–5804.

Pachur, T., Schulte-Mecklenbeck, M., Murphy, R. O., & Hertwig, R. (2018). Prospect theory reflects selective allocation of attention. *Journal of Experimental Psychology: General, 147*(2), 147–169.

Payne, J. W. (1976). Task complexity and contingent processing in decision making: An information search and protocol analysis. *Organizational Behavior and Human Performance, 16*(2), 366–387. https://doi.org/10.1016/0030-5073(76)90022-2

Peters, E., Västfjäll, D., Slovic, P., Mertz, C. K., Mazzocco, K., & Dickert, S. (2006). Numeracy and decision making. *Psychological Science, 17*(5), 407–413.

Ratcliff, R., Smith, P. L., Brown, S. D., & McKoon, G. (2016). Diffusion decision model: Current issues and history. *Trends in Cognitive Sciences, 20*(4), 260–281.

Read, D. (2001). Is time-discounting hyperbolic or subadditive? *Journal of Risk and Uncertainty, 23*(1), 5–32.

Read, D., Frederick, S., & Scholten, M. (2013). DRIFT: An analysis of outcome framing in intertemporal choice. *Journal of Experimental Psychology. Learning, Memory, and Cognition, 39*(2), 573–588. https://doi.org/10.1037/a0029177

Reeck, C., Wall, D., & Johnson, E. J. (2017). Search predicts and changes patience in intertemporal choice, 201707040 *Proceedings of the National Academy of Sciences*. https://doi.org/10.1073/pnas.1707040114.

Regenwetter, M., Cavagnaro, D. R., Popova, A., Guo, Y., Zwilling, C., Lim, S. H., & Stevens, J. R. (2018). Heterogeneity and parsimony in intertemporal choice. *Decision, 5*(2), 63–94.

Reimers, S., Maylor, E. A., Stewart, N., & Chater, N. (2009). Associations between a one-shot delay discounting measure and age, income, education and real-world impulsive behavior. *Personality and Individual Differences, 47*(8), 973–978. https://doi.org/10.1016/j.paid.2009.07.026

Rubinstein, A. (2003). "Economics and Psychology"? The case of hyperbolic discounting. *International Economic Review, 44*(4), 1207–1216.

Samuelson, P. (1937). A note on the measurement of utility. *The Review of Economic Studies, 4*(2), 155–161. https://doi.org/10.1901/jeab.1980.33-285

Scharf, S. E., Wiegelmann, M., & Bröder, A. (2019). Information search in everyday decisions: The generalizability of the attraction search effect. *Judgment and Decision Making, 14*(4), 488–512.

Scholten, M., & Read, D. (2006). Discounting by Intervals: A Generalized Model of Intertemporal Choice. *Management Science, 52*(9), 1424–1436. https://doi.org/10.1287/mnsc.1060.0534

Scholten, M., & Read, D. (2010). The psychology of intertemporal tradeoffs. *Psychological Review, 117*(3), 925–944. https://doi.org/10.1037/a0019619

Scholten, M., & Read, D. (2013). Time and outcome framing in intertemporal tradeoffs. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 39*(4), 1192–1212. https://doi.org/10.1037/a0031171

Scholten, M., Read, D., & Sanborn, A. (2014). Weighing Outcomes by Time or Against Time? Evaluation Rules in Intertemporal Choice. *Cognitive Science, 38*, 399–438. https://doi.org/10.1111/cogs.12104

Scholten, M., Read, D., & Sanborn, A. (2016). Cumulative weighing of time in intertemporal tradeoffs. *Journal of Experimental Psychology: General, 145*(9), 1177–1205. https://doi.org/10.1037/xge0000198

Sepulveda, P., Usher, M., Davies, N., Benson, A. A., Ortoleva, P., & De Martino, B. (2020). Visual attention modulates the integration of goal-relevant evidence and not value. *Elife, 9*, e60705.

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

Simon, H. A., & Newell, A. (1971). Human problem solving: The state of the theory in 1970. *American Psychologist, 26*(2), 145–159. https://doi.org/10.1037/h0030806

Smith, S. M., & Krajbich, I. (2019). Gaze Amplifies Value in Decision Making. *Psychological Science, 30*(1), 116–128. https://doi.org/10.1177/0956797618810521

Spiegelhalter, D. J., Best, N. G., Carlin, B. P., & Van Der Linde, A. (2002). Bayesian measures of model complexity and fit. *Journal of the Royal Statistical Society: Series B (statistical methodology), 64*(4), 583–639.

Stan Development Team (2020). RStan: The R interface to Stan. R package version 2.21.2. http://mc-stan.org/.

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.

Willemsen, M. C., Böckenholt, U., & Johnson, E. J. (2011). Choice by value encoding and value construction: Processes of loss aversion. *Journal of Experimental Psychology. General, 140*(3), 303–324. https://doi.org/10.1037/a0023493

Wulff, D. U., & van den Bos, W. (2018). Modeling Choices in Delay Discounting. *Psychological Science, 29*(11), 1890–1894. https://doi.org/10.1177/0956797616664342

Zhang, Y. Y., Zhou, L., Li, S., & Liang, Z. Y. (2022). Computation of subjective value does not always elicit alternative-based information searching in intertemporal choice. *Journal of Behavioral Decision Making*.

Zhao, W. J., Coady, A., & Bhatia, S. (2022). Computational mechanisms for context-based behavioral interventions: A large-scale analysis. *Proceedings of the National Academy of Sciences*, *119*(15), e2114914119.

Zilker, V., & Pachur, T. (2021). Nonlinear probability weighting can reflect attentional biases in sequential sampling. *Psychological Review*.