Estimating the dynamic role of attention via random utility

Stephanie M. Smith[1] · Ian Krajbich[1,2] · Ryan Webb[3]

Abstract

When making decisions, people tend to look back and forth between the alternatives until they eventually make a choice. Eye-tracking research has established that these shifts in attention are strongly linked to choice outcomes. A predominant framework for understanding the dynamics of the choice process, and thus the effects of attention, is sequential sampling of information. However, existing methods for estimating the attention parameters in these models are computationally costly and overly flexible, and yield estimates with unknown precision and bias. Here we propose an estimation method that relies on a link between sequential sampling models and random utility models (RUM). This method uses familiar econometric tools (i.e., logistic regression) and yields estimates that appear to be unbiased and relatively precise compared to existing methods, in a small fraction of the computation time. The RUM thus appears to be a useful tool for estimating the effects of attention on choice.

1 Introduction

A central question for choice researchers is how decision makers collect information during the choice process, and how the information collected can be measured and modeled (Sims 2003). To understand the role of information accumulation, economists have recently turned to sequential sampling models (Fehr and Rangel 2011;

✉ Ian Krajbich
krajbich.1@osu.edu

[1] Department of Psychology, The Ohio State University, Columbus, USA

[2] Department of Economics, The Ohio State University, Columbus, USA

[3] Rotman School of Management, University of Toronto, Toronto, Canada

Krajbich et al. 2014; Woodford 2014; Gossner et al. 2018; Fudenberg et al. 2018). These models conceptualize the decision process as one of accumulation and comparison of noisy information up to a pre-specified confidence threshold. For example, the most well-known model of this class, the drift diffusion model (DDM), uses data on choice and the distribution of response time (RT) to identify latent decision parameters such as the strength of preference (quality of the sampled information) and the confidence required to decide.

Attention to a given option—and the simultaneous inattention to the other option(s)—constrains the sampling process: attention favors the focal alternative at the expense of other(s). While attention can in some cases be inferred from people's choices (Gabaix 2017), a more direct approach is to use eye tracking (Wang et al. 2010; Knoepfle et al. 2009; Reutskaja et al. 2011; Arieli et al. 2011; Lahey and Oxley 2016; Devetag et al. 2016; Polonio et al. 2015; Shi et al. 2013; Chen and Krajbich 2017). Eye-tracking data—either collected in the laboratory, or increasingly, field settings[1]—provides a dynamic, rich set of data that is useful for predicting choice above and beyond a subject's stated preferences. Recent work has established a stable individual difference in the degree to which visual attention and choice are linked (Smith and Krajbich 2018).

Given that the link between attention and choice is empirically well established and theoretically accounted for in the DDM framework, it is now important to develop statistical tools to study the link between attention and choice. Doing so will allow the researchers to better assess and predict the role of attention across contexts and decision makers. In this paper, we propose an effective and efficient method for estimating the role of attention in choice. Our approach is based on recent results which demonstrate that the basic DDM can be represented by a random utility model (RUM) in which the distribution of utilities depends on time (Echenique and Saito 2017; Webb 2019). Here, we extend this result to incorporate the role of attention. This leads to approximation methods for estimating the role of attention using familiar methods in the econometrician's toolkit, such as the standard logit model.

---

[1] Advances in mobile eye-tracking have enabled field research in the area. In a typical study, participants wear glasses with embedded eye-tracking technology while they shop in stores. The glasses record both eye movements and the environment, allowing researchers to investigate the role of attention in a field setting (e.g., Harwood and Jones 2014; Bagdziunaite et al. 2014). While academic research in this area is nascent, a number of commercial enterprises are already collecting and using this data. Three examples include: (1) market research firms like Nielsen are currently using eye tracking to measure consumer responses to advertisements; (2) Consumer-grade eye tracking is being tested and implemented by a number of firms, for both general use (Google Glass) as well as specific uses (VR headsets); and (3) Retailers like Amazon are implementing high-definition, high-density, camera technology in brick-and-mortar locations that can track a shopper's attention to specific products on the shelf. We expect further technological advances to be made in this area. For instance, researchers can take advantage of front-facing cameras on phones, tablets, and computers to better understand how consumers interact with and attend to the information on their screens.

2 Background

In the standard DDM, the agent is faced with two options of unknown utility. At each instant in time, the agent samples a stochastic signal about the utility of each option. The agent calculates the difference between the two signals and adds this new piece of evidence to the sum of all previous evidence. This process continues until the accumulated evidence reaches a pre-determined stopping criterion. The optimal stopping criterion depends on various assumptions, but in the standard DDM, it is constant over time (Fudenberg et al. 2018).

A central assumption in the standard DDM is that the stochastic signals from each option are i.i.d. over time. However, this assumption has been challenged based on two basic facts from cognitive science. First, people move their eyes around in discrete steps; eye movements are characterized by near-instantaneous transitions ("saccades") between periods of fixed gaze ("fixations") that typically last about 0.3 s. Second, decision makers extract more information from items that are being fixated than from those that are not. Together, these two facts imply unaccounted-for autocorrelation in the time series of sampled signals. In a binary decision, the agent spends some blocks of time looking at option i and others at option j. These blocks of time (or "dwells") correspond to periods during which signals about either option i or option j are amplified. Consequently, if both options are positively valued and option i receives more attention than expected, it should be more preferred than expected.

Currently, the most comprehensive method for estimating the relationship between attention and choice involves extending the DDM to include an attentional discounting parameter (θ), such that the rate of evidence accumulation ("drift rate") is modulated by attention. This extension is known as the attentional DDM, or aDDM (Krajbich et al. 2010). Its usefulness stems from its ability to integrate multiple inputs (stated or inferred preferences and eye movements) to determine the joint probability distributions of choice and RT.

The aDDM has been used to study choice behavior in a number of domains, including food (Fisher 2017; Krajbich et al. 2010; Krajbich and Rangel 2011; Towal et al. 2013), durable goods (Krajbich et al. 2012), artwork (Vaidya and Fellows 2015), social preferences (Ashby et al. 2016; Smith and Krajbich 2018), conditioned stimuli (Cavanagh et al. 2014), roaming bandits (Konovalov and Krajbich 2016), and explicit lotteries (Smith and Krajbich 2018; Stewart et al. 2015). Additional studies have used the aDDM for inspiration, but without explicit model fitting, in the domains of morality (Pärnamets et al. 2015) and intertemporal choice (Amasino et al. 2019). Needless to say, the link between attention and choice is under intense study in a wide variety of choice environments.

Smith and Krajbich (2018) describe how the aDDM accounts for several prominent features of typical choice data. As alluded to above, it captures the robust relationship between relative dwell-time advantage and choice probability: in a 2–3 s decision, a dwell-time advantage of 0.5 s results in a 30% increase in choice probability. It also accounts for the utility-dependent probability of choosing the last-seen item: at indifference a typical subject chooses the last-seen item 70% of

the time, but this probability increases (decreases) as the utility of the last-seen item increases (decreases). Finally, the model accounts for the finding that the effect of attention on choice increases with the utility of the option; in the extreme case of aversive items subjects are less likely to choose the more-attended option (Armel et al. 2008; Smith and Krajbich 2019).

Despite surprising stability across experiments in the link between attention and choice, there is of course interesting variation across individuals and contexts. Estimating this variation, and identifying which individuals or choice contexts are particularly prone to attentional manipulations is of particular importance. However, a major obstacle to studying this variation is the difficulty in fitting the aDDM to data. Fitting the model at the individual level requires many trials and even then, the fitting process is prone to issues. Though methods exist for simulating the RT distribution of the basic DDM, applying them to the aDDM is complicated by the fact that the simulations must incorporate the eye-movement data. As a result, the aDDM has typically been fit to data using very large Monte Carlo simulations coupled with grid search parameter selection to maximize likelihood functions (detailed below).

This simulation/grid search approach has several important limitations. First, the properties of the parameter estimates are unknown. If parameter estimates are biased, this has important implications for inference and predicting choices out of sample. Second, a grid search method does not directly yield standard errors for inference. This is particularly an issue for studies that attempt to demonstrate a difference in parameter estimates across subjects or conditions. While, in principle, bootstrap methods can be used to compute standard errors, they are limited in this case by the severe computational burden of the fitting process. Even with parallelization, it can take days to fit the model to a group of subjects, which translates to weeks or even months of computation time to generate bootstrap confidence intervals. Moreover, the traditional fitting process can be somewhat arbitrary. There are several researcher degrees of freedom: how many simulations to run, how many parameter combinations to test, how to adjust the RT bins, how to use eye-tracking data in the fitting process, etc. Additionally, the reliance on RT data in the traditional likelihood calculations could be a problem if the RT data are contaminated by other non-decision-related processes.

Here, we address these issues by demonstrating the equivalence of the aDDM to models within the class of RUM. This leads to approximation methods for estimating parameters of the aDDM directly from choice and eye-movement data using the RUM toolkit (e.g., logit). As such, the properties of these estimates are well described and standard errors can be computed directly. In contrast to the aDDM, the RUM approach to estimating attentional effects does not require any arbitrary binning decisions; it uses raw continuous measures of dwell proportion. While the aDDM might provide slightly different estimates of θ depending on the exact binning procedure, the RUM method provides a unique estimate. We find significant improvement in parameter recovery for the attentional discounting parameter θ with more accurate Type 1 error rates compared to existing methods for estimating the aDDM.

3 Binary choice model

In the standard DDM, the agent is faced with two options i and j. Each option has a subjective value v_i which could be a function of various observable attributes X_i.[2] At each instant in time, the agent receives a stochastic Gaussian signal $z_{i,t} \sim N\left(v_i, s_i^2\right)$ and similarly for j. Each signal $z_{i,t}$ is added to the previous signals for that option. Formally, for each option i and j, there is an accumulator[3]:

$$Z_{i,t} = \sum_{k=1}^{t} z_{i,k} \tag{1}$$

In the aDDM, there is one small, but important difference from the above: the mean of the signal distribution depends on the fixation location. Let $e_{i,t}$ capture the role of attention and let $e_{i,t} = 1$ if the agent is attending to option i, 0 otherwise. Then, the signal $z_{i,t}$ is distributed with a mean that depends on decision-specific inputs v_i and $e_{i,t}$, as well as an agent-specific parameter θ:

$$z_{i,t} \sim N\left(v_i\left(\theta + e_{i,t}(1-\theta)\right), s_i^2\right), \tag{2}$$

where $\theta \in [0, 1]$. If $\theta = 1$, then attention plays no role, as in the standard DDM. If $\theta = 0$ then evidence is only accumulated for the attended option.

To close the model, a stopping rule must be specified. In the DDM and aDDM, this rule takes the form of the difference between the accumulators crossing a constant threshold b:

$$\left|Z_{i,t} - Z_{j,t}\right| \geq b, \tag{3}$$

with the largest accumulator, $\arg\max\left\{Z_{i,t*}\right\}$, yielding the choice, where $t*$ is the first t for which Eq. (3) is satisfied. The drift rate at time t is therefore the net rate of evidence accumulation towards option i: $(v_i - v_j)\theta + (e_{i,t}v_i - e_{j,t}v_j)(1 - \theta)$.

There are three technical issues worth noting about the model. First, since at each moment in time, the stopping rule calculates the difference between two Gaussian random variables with variance s_i^2, only the summed variance can be estimated:

$$\sigma^2 = s_i^2 + s_j^2. \tag{4}$$

Second, in addition to the decision time $t*$, there is additional time required for the agent to attend to the first option and to physically indicate their choice once the threshold b is crossed. This extra time is known as the non-decision time (T_{er}). Thus, the recorded RT is:

---

[2] For an example of how components of value can be estimated from observable attributes in the DDM framework, see Chiong et al. (2018).

[3] In many applications, an "initial point" $Z_{i,0}$ can be specified to allow for any priors that decision makers might have.

$$RT = t^* + T_{er}. \tag{5}$$

While T_{er} is estimated during the full model-fitting procedure, it does not factor into the estimate of θ.

Third, the units in the model are arbitrary, and so when estimating the best-fitting aDDM parameters, we include a scaling parameter d that converts any observables X_i to subjective valuations, $v_i = X_i d$. Because of the arbitrary units in the model, the parameter set d, b, and σ are not uniquely identified, thus one must be normalized to identify the others. Here we let $b = \pm 1$, as is standard with the aDDM.

Our first result demonstrates that the aDDM can be mapped into the space of RUMs. The result follows the mapping from the general sequential sampling model class into the RUM demonstrated by Webb (2019), depicted in Fig. 1. Since an accumulation model picks the alternative with the largest accumulator, and this is true even when the levels are divided by the RT (a positive number), the determinant of choice in classic DDM is a maximization over v_i with an error term suitably normalized by the RT distribution. For example, a longer RT leads to a smaller error variance, typically referred to as the "speed-accuracy tradeoff."

The reason this result can be extended to include the aDDM lies in the exchangeability property of the accumulation. Since the accumulation in the aDDM is a Markov process, the order in which attention is allocated does not change the level of the accumulator. This means that only the proportion of attention allocated to an alternative interacted with the underlying value (v_i) determines the level of the accumulator. In combination with the threshold b, this determines the choice probability.[4]

If we let u_i be the utility of option i, E_i be the fraction of t^* spent attending to option i (with $E_j = 1 - E_i$), and η_i be a random component of the utility, then the following result states this equivalence.

Proposition 1 *The aDDM can be represented by a RUM of the form:*

$$u_i = (E_i + \theta E_j)v_i + \eta_i, \tag{6}$$

where η_i has mean zero and the choice probabilities are given by,

$$P(i, t) = \text{Prob}\left[E_i v_i - E_j v_j + \theta\left(E_j v_i - E_i v_j\right) > \eta_i - \eta_j\right]. \tag{7}$$

Proof Following Webb (2019),

$$Z_{i,t^*} = v_i \sum_1^{t^*} \left(\theta + e_{i,t}(1 - \theta)\right) + \sum_1^{t^*} \epsilon_{i,t}$$

---

[4] It should be noted that the distribution of choice probabilities is not completely independent of the order of attention in this model. For example, a process with high autocorrelation in the gaze pattern will lead to an earlier boundary crossing than a process with low autocorrelation, even with equal ex ante division of attention in both cases. This will be captured by the distribution of η_i implied by Proposition 1 and how it depends on time.

(a)

DDM

$$v_i t^* + \sum_1^{t^*} \epsilon_{i,t}$$

RUM

$$v_i + \frac{1}{t^*} \sum_1^{t^*} \epsilon_{i,t}$$
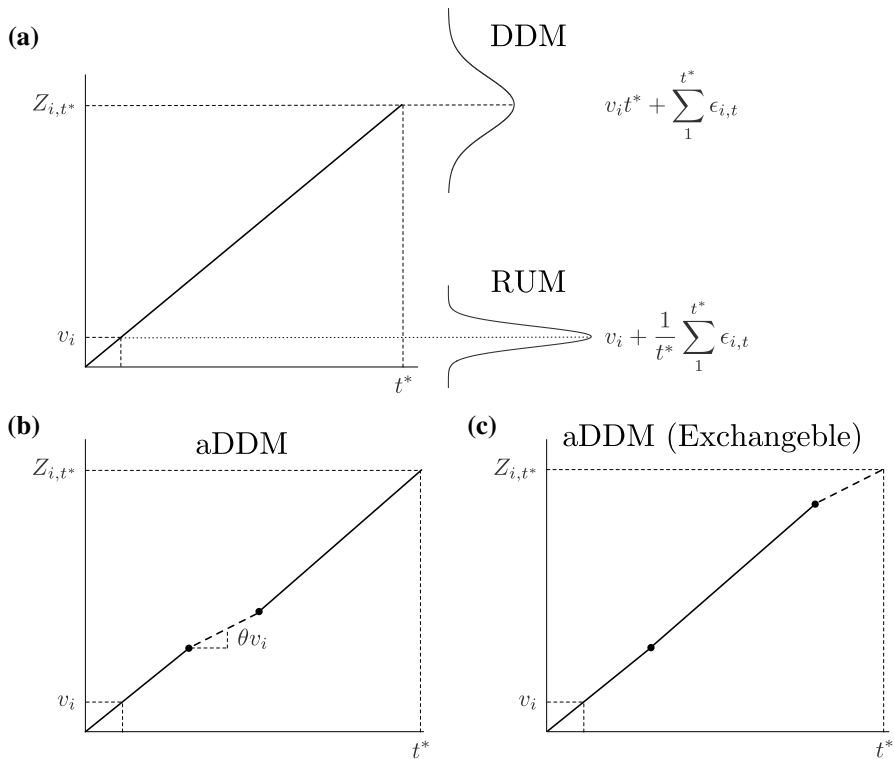
(b) aDDM

(c) aDDM (Exchangeble)

Fig. 1 **a** The equivalence between the DDM and the RUM. Here, the *x*-axis represents time and the *y*-axis represents accumulated evidence for option *i*. In the DDM, the distribution of accumulated evidence has a mean that is a linear function of time with slope given by the value of option *i*, plus some error. The distribution of the RUM, on the other hand, is equal to the DDM distribution, scaled down linearly by time. Thus, the function for the RUM distribution is simply the DDM distribution function, divided by t^*. **b** In the aDDM, attention toward the other alternative leads to a discounting of the valuation by θ (dashed line), and therefore, a slower rate of evidence accumulation. The solid line represents accumulated evidence over time while attention is devoted to option *i*, and the dashed line represents accumulated evidence over time while attention is devoted to option *j*. In this example, attention is devoted to option *i*, then option *j*, and then option *i* again. **c** The final level of the accumulator does not depend on the order of attentional shifts (compare to b). Here, attention is devoted to option *i* and then option *j*. The total amount of time spent on each option is the same as in **b**, but the order of attentional shifts is irrelevant, so the total accumulated evidence is equivalent to **b**

where $\epsilon_{i,t}$ is the deviation from the drift rate due to the stochastic sampling. Moreover, $\arg\max\left\{ Z_{i,t*} \right\} = \arg\max\left\{ \frac{Z_{i,t*}}{t*} \right\}$, therefore this model can be represented via the RUM

$$u_i = D(e_{i,t})v_i + \eta_i$$

where $D(e_{i,t}) \equiv \frac{1}{t^*} \sum_1^{t^*} \left(\theta + e_{i,t}(1 - \theta) \right)$, and $\eta_i \equiv \frac{1}{t^*} \sum_1^{t^*} \epsilon_{i,t}$.

The function $D(e_{i,t})$ has the following interpretation.

$$D(e_{i,t}) = \frac{1}{t^*} \sum_1^{t^*} \left(\theta + e_{i,t}(1-\theta) \right)$$

$$= \frac{1}{t^*} \sum_1^{t^*} e_{i,t} + \theta \frac{1}{t^*} \sum_1^{t^*} \left(1 - e_{i,t} \right)$$

$$= E_i + \theta(1 - E_i) = E_i + \theta E_j$$

∎

The relation of the aDDM to the RUM has important consequences for parameter estimation which we now discuss.

4 Finite sample performance of the estimator

Consider two alternatives i and j with ratings X_i and X_j, therefore the subjective values are given by $v_i = X_i d$. In this case, we can consider estimation of Eq. (7) from a standard logit regression

$$P_i = \text{Prob}\left[(E_i X_i - E_j X_j)d + (E_j X_i - E_i X_j)\gamma > \eta_i - \eta_j\right], \tag{8}$$

where $\hat{\theta}$ can be recovered via $\hat{\theta} = \frac{\hat{\gamma}}{\hat{d}}$.[5]

The valuation can also be alternative specific provided enough variation over alternatives in the dataset, so $v_i = X_i d_i$. In this case, the standard logit regression is

$$P(i,t) = \text{Prob}\left[E_i X_i d_i - E_j X_j d_j + E_j X_i \gamma_i - E_i X_j \gamma_j > \eta_i - \eta_j\right], \tag{9}$$

and $\hat{\theta}_i$ can be recovered via $\hat{\theta}_i = \frac{\hat{\gamma}_i}{\hat{\beta}_i}$. The attentional parameter can be further restricted to be symmetric across alternatives, so $\theta_i = \theta_j = \theta$.

Note that, strictly, the estimation of Eqs. (14) and (15) is misspecified, since the random utility error distribution implied by the DDM differs from that used by the logit (see Webb 2019). However, in the next section, we will demonstrate that the approximation of this distribution provided by the logit model does not appear to bias the resulting model estimates meaningfully.

4.1 Simulated data

To evaluate our proposed estimation method, we performed a Monte Carlo parameter recovery exercise. The data generating process was the aDDM with the following

---

[5] For a related derivation of θ, though using a DDM-based method rather than Logit, see Cavanagh et al. 2014.

parameters: $d \in \{0.00015, 0.0002, 0.00025\}$ per millisecond, $\sigma \in \{0.015, 0.02, 0.025\}$, $\theta \in \{0.1, 0.3, 0.5, 0.7, 0.9\}$, $T_{er} \in \{300, 400, 500\}$ milliseconds. For every combination of these parameter values, 20 datasets were simulated using the empirical subjective value ratings and gaze patterns from 100 randomly sampled trials from Krajbich et al. (2010). Therefore, for each value of θ, there were 540 datasets, each with 100 simulated trials. We then recovered estimates of θ for each dataset using a standard logit estimation of the RUM from Eq. (8).

To evaluate the grid search method, we followed standard practice and fit each of the datasets using maximum simulated likelihood estimation (MSLE). This method is quite involved and requires simulating a large number of sample datasets for each set of parameters in the grid. We bin the simulations and the datasets in the same way. We first separate the trials from the dataset according to the absolute difference in rating between the two options, the decision made, and the direction of the final fixation. Next, we bin the trials according to RTs, using a median split for decisions where the lower-rated option was chosen and with quantile bins of $\{0.1, 0.3, 0.5, 0.7, 0.9\}$ for decisions where the weakly higher-rated option was chosen.

To implement the grid search, we used three iterations of 5000 random grid parameter combinations each (Bergstra and Bengio 2012), with 10 simulations per trial. The range of parameters for the first iteration was $d \in [0.0001, 0.0003]$ per millisecond, $\sigma \in [0.01, 0.03]$, $\theta \in [0, 1]$, $T_{er} \in [200, 600]$ milliseconds. The second and third iterations' parameter ranges were based on the top 5% best-fitting model parameters of the previous iterations (according to log likelihood). We adjusted these ranges according to a dummy parameter, ξ, to correct for artificial contraction of the parameter range over time (e.g., if ξ's upper value decreased by 1% of its initial range, we increased the upper value on each of the parameters by 1% of their respective ranges, with the initial upper values as an upper bound).

Our results demonstrate that the RUM estimates of θ appear to be unbiased and estimated relatively precisely (Fig. 2). This is in contrast to estimates of the aDDM using the grid search method. In particular, there is a stark improvement both in the mean estimates and an improved standard deviation. Computation time of the RUM is negligible at 0.03 s per dataset, while the grid search method takes ~2.5 h per dataset (note: this can be reduced to 10 min per dataset if parallelized on 50 cores). This comparison, while impressive, is actually conservative since we chose relatively tight boundaries for the grid search that included the true parameter values.

Finally, it is difficult to assess the performance of the grid search method in parameter inference, because it does not directly yield standard errors of the estimates. Given the computational burden of the grid search, even when the grid search is parallelized over 50 cores, minimal 1000 bootstrap samples would take ~7 days to estimate. The RUM method, however, yields standard errors directly with Type 1 errors that are near the expected 0.05 level.

Clearly, the RUM method is effective and efficient at estimating the degree to which attention shapes the choice process. However, it is also beneficial when estimating the rest of the parameters, as well. After estimating the attention parameter θ with the RUM method, we used a grid search (as described above) to estimate the remaining parameters for each of the 2700 datasets. Essentially, we fixed θ to be the RUM estimate across all three iterations, and fit the remaining parameters.
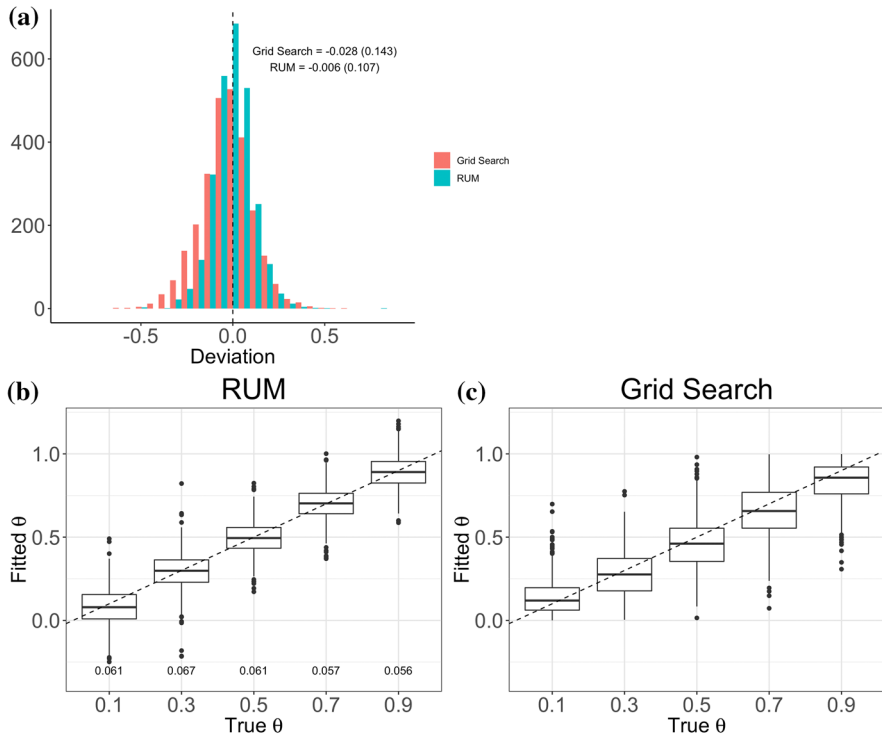
Fig. 2 **a** Histograms of the deviations with mean (sd) and **b**, **c** direct comparisons between the recovered θ values and their true values from each of the methods on the 2700 datasets, with Type 1 error rates in **b**. The RUM method is more precise and less biased than the traditional grid search method

Because we know the true parameters that generated each of the 2700 datasets, we can evaluate the success of each method in identifying these parameters. Indeed, the RUM method provides significantly better fits than the traditional grid search, total RUM log likelihood$=-1,483,267$ versus total grid search log likelihood$=-1,485,244$. A Wilcoxon signed rank test shows that the RUM fits are significantly better, as well, $V=2,648,900$, $p<10^{-16}$. At the level of the dataset, 68% of the 2700 paired log likelihoods are higher in the RUM method. Moreover, the root mean squared error of the RUM method is smaller than that of the grid search method for all parameters (d: 3.510×10^{-5} vs. 3.514×10^{-5}; σ: 2.282×10^{-3} vs. 2.319×10^{-3}; θ: 0.107 vs. 0.145; T_{er}: 99.54 vs. 101.2) and the correlations between true and fitted parameters are higher in the RUM method than in the grid search method (d: 0.745 vs. 0.741; σ: 0.882 vs. 0.873; θ: 0.937 vs. 0.874; T_{er}: 0.550 vs. 0.538; each $p<10^{-15}$). Overall, estimating θ with the RUM method does not simply improve the estimation of θ, it also improves the precision of other estimated parameters and the overall goodness-of-fit.

4.2 Real data

To assess the performance of the RUM method on a real (i.e., non-simulated) dataset, we also fit the aDDM to the participant-level data from Krajbich et al. (2010). In this study, subjects ($N=39$) first rated how much they would like to eat 70 food items. Then, they made 100 incentivized choices between pairs of food items while their eye movements were tracked. We used subjects' self-reported valuations from the rating data (as X_i and X_j) while fitting the choice task data (comprising RTs, choices, and eye movements). We fit both with the traditional grid search and the RUM method (estimating θ with the model and then fitting the remaining parameters with a grid search). Here, the fit of the RUM and the grid search yielded similar values, with correlations of 0.74, 0.90, 0.57, 0.46 for d, σ, θ, T_{er}, respectively. It is also reassuring that the median/mean difference between parameter values is very close to zero in each case, indicating that the two methods agree in the aggregate as well (Fig. 3). A major advantage of the RUM method is that it estimates θ essentially instantaneously, allowing the overall fitting exercise to, again, finish in a fraction of the time of the grid search method.

The RUM method also provides standard errors on θ, which are key for inferential statistics (Fig. 4a). Using the RUM method, we find that 35 out of 39 subjects have $\theta<1$, with 26 out of 39 significant at the 0.05 level. Interestingly, 12 out of 39 subjects (6 significantly) have $\theta<0$, which falls outside of the theoretically predicted range of estimates. The implication of $\theta<0$ is that while attending to option i, the accumulator on option j decreases at a rate proportional to its underlying utility. A negative value of θ is typically not allowed by the aDDM, as it would imply that items that are more highly valued while being attended to are more negatively valued while not being attended to. One possibility is that the negative θ estimates are due to features of the data that are not captured by the aDDM. For instance, it is possible that unattended items leak evidence; leakage is a feature of some other sequential sampling models (e.g., leaky competing accumulator; Usher & McClelland 2001) but has yet to be incorporated in the aDDM framework. In the logit method, a negative θ arises because the logit places no boundaries on the estimate of θ. Such an estimate is precluded by the grid search approach, because it simply does not search in this domain of the parameter space (Fig. 4b).

If a researcher so desired, a positivity constraint on θ can simply be added to the maximum likelihood procedure used to estimate the logit model. This can be accomplished by putting positivity constraints on the two parameters (γ and d, where $\theta = \frac{\gamma}{d}$) estimated via the logit maximum likelihood procedure. The subjects with nonnegative θ in the unconstrained regression displayed very little difference between the constrained and unconstrained regressions (mean deviance $=0.018$, SE$=0.0075$; Pearson's correlation$=0.99$, $p<0.0001$). On the other hand, the subjects with negative θ in the unconstrained regression were best fit with $\theta\approx0$ in the constrained regression. Overall, constraining the regression does not change our conclusions about the improved estimation of θ with the RUM.
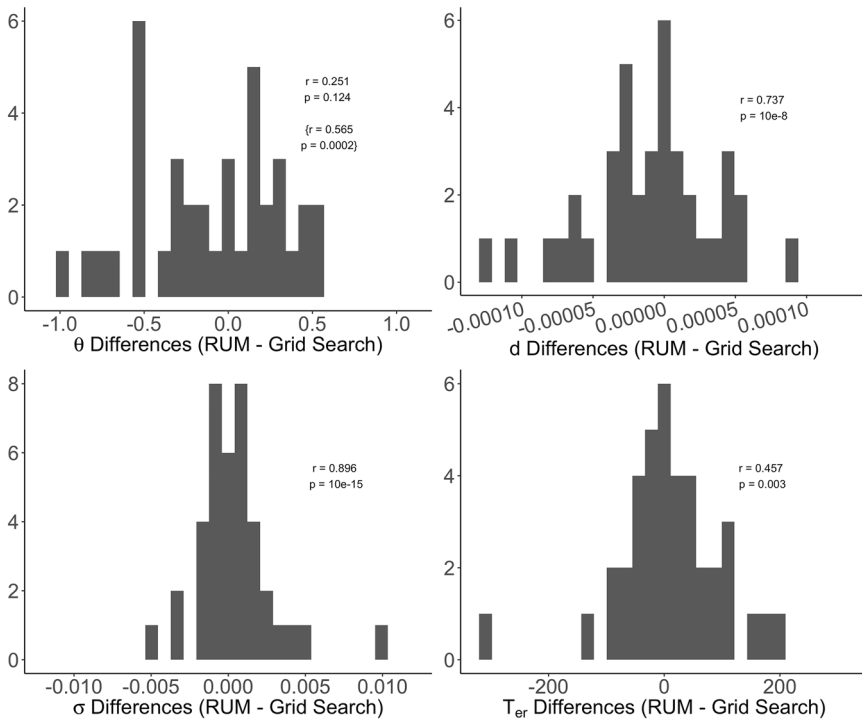
Fig. 3 Histograms of the difference between RUM and grid search estimated parameters, using real data. Inset numbers refer to Pearson correlation coefficients and p values. Note: for θ, one observation (-20.8) was removed from the graph as an outlier. Coefficients in brackets refer to the correlation without the outlier
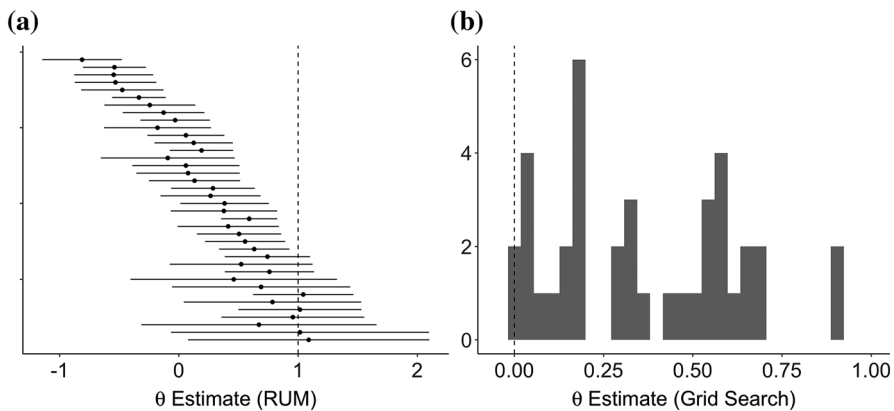

Fig. 4 **a** Inference test of RUM θ estimates. Again, the outlier (-20.7) was removed from the graph for clarity. **b** Distribution of θ from the grid search method

5 Conclusion

In this paper, we have demonstrated the benefits of applying the RUM to estimate the effects of attention on choice. In doing so, we have established that the role of attention in choice can be accurately and quickly estimated. We find improvement in parameter recovery using a mere fraction of the computational time of existing methods. Moreover, the RUM rests on a well-known, firmly structured analysis (i.e., a logit regression), while existing methods are plagued by their flexibility and lack of standardization. The binning process, use of eye-tracking data, number of simulations, and range of parameter combinations are all left to the discretion of the researcher. Finally, parameters estimated via existing methods are not amenable to conducting inference tests. These are all problems the RUM method solves.

There are, of course, limitations to this approach. The effectiveness of the DDM in modeling choice behavior is rooted in its capability to make use of RT, in addition to choices. However, with the RUM approach, RTs are ignored in the estimation of the attentional parameter θ. This could mean that valuable information is lost, but it could also mean that the RUM method is more robust to assumptions about other components of the model. With eye trackers now commercially available for ~ \$100, and phone manufacturers and retailers deploying mobile eye-tracking tools, the use of eye-movement data in research will likely soon become commonplace. It is therefore important to continue to develop better methods for analyzing that data.

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*. https://doi.org/10.1038/s41562-019-0537-2

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

Armel, K. C., Beaumel, A., & Rangel, A. (2008). Biasing simple choices by manipulating relative visual attention. *Judgment and Decision Making, 3*(5), 396–403.

Ashby, N. J. S., Jekel, M., Dickert, S., & Glöckner, A. (2016). Finding the right fit: A comparison of process assumptions underlying popular drift-diffusion models. *Journal of Experimental Psychology: Learning, Memory, and Cognition*. https://doi.org/10.1037/xlm0000279.

Bagdziunaite, D., Nassri, K., Clement, J., & Ramsøy, T. Z. (2014). An added value of neuroscientific tools to understand consumers' in-store behaviour. In *EMAC 2014*.

Bergstra, J., & Bengio, Y. (2012). Random search for hyper-parameter optimization. *The Journal of Machine Learning Research, 13*(1), 281–305.

Cavanagh, J. F., Wiecki, T. V., Kochar, A., & Frank, M. J. (2014). Eye tracking and pupillometry are indicators of dissociable latent decision processes. *Journal of Experimental Psychology: General, 143*(4), 1476–1488. https://doi.org/10.1037/a0035813.

Chen, W. J., & Krajbich, I. (2017). Computational modeling of epiphany learning. *Proceedings of the National Academy of Sciences, 114*(18), 4637–4642. https://doi.org/10.1073/pnas.1618161114.

Chiong, K., Shum, M., Webb, R., & Chen, R. (2018). *Split-second decision-making in the field: Response times in mobile advertising*. SSRN Working Paper.

Devetag, G., Di Guida, S., & Polonio, L. (2016). An eye-tracking study of feature-based choice in one-shot games. *Experimental Economics, 19*(1), 177–201.

Echenique, F., & Saito, K. (2017). Response time and utility. *Journal of Economic Behavior & Organization, 139,* 49–59. https://doi.org/10.1016/j.jebo.2017.04.008.

Fehr, E., & Rangel, A. (2011). Neuroeconomic foundations of economic choice—Recent advances. *The Journal of Economic Perspectives, 25*(4), 3–30.

Fisher, G. (2017). An attentional drift diffusion model over binary-attribute choice. *Cognition, 168,* 34–45. https://doi.org/10.1016/j.cognition.2017.06.007.

Fudenberg, D., Strack, P., & Strzalecki, T. (2018). Speed, accuracy, and the optimal timing of choices. *American Economic Review, 108*(12), 3651–3684. https://doi.org/10.1257/aer.20150742.

Gabaix, X. (2017). *Behavioral Inattention (No. w24096)*. Cambridge, MA: National Bureau of Economic Research. https://doi.org/10.3386/w24096.

Gossner, C., Steiner, J., & Stewart, C. (2018). Attention Please!*, 34. Working paper, University of Toronto Department of Economics

Harwood, T., & Jones, M. (2014). Mobile eye-tracking in retail research. In M. Horsley, M. Eliot, B. A. Knight, & R. Reilly (Eds.), *Current trends in eye tracking research* (pp. 183–199). Cham: Springer. https://doi.org/10.1007/978-3-319-02868-2_14.

Knoepfle, D. T., Tao-yi Wang, J., & Camerer, C. F. (2009). Studying learning in games using eye-tracking. *Journal of the European Economic Association, 7*(2–3), 388–398. https://doi.org/10.1162/JEEA.2009.7.2-3.388.

Konovalov, A., & Krajbich, I. (2016). Gaze data reveal distinct choice processes underlying model-based and model-free reinforcement learning. *Nature Communications, 7,* 12438. https://doi.org/10.1038/ncomms12438.

Krajbich, I., Armel, C., & Rangel, A. (2010). Visual fixations and the computation and comparison of value in simple choice. *Nature Neuroscience, 13*(10), 1292–1298. https://doi.org/10.1038/nn.2635.

Krajbich, I., Lu, D., Camerer, C., & Rangel, A. (2012). The attentional drift-diffusion model extends to simple purchasing decisions. *Frontiers in Psychology*, *3*. Retrieved from http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3374478/.

Krajbich, I., Oud, B., & Fehr, E. (2014). Benefits of neuroeconomic modeling: New policy interventions and predictors of preference. *The American Economic Review, 104*(5), 501–506.

Krajbich, I., & Rangel, A. (2011). Multialternative drift-diffusion model predicts the relationship between visual fixations and choice in value-based decisions. *Proceedings of the National Academy of Sciences, 108*(33), 13852–13857. https://doi.org/10.1073/pnas.1101328108.

Lahey, J. N., & Oxley, D. (2016). The power of eye tracking in economics experiments. *American Economic Review, 106*(5), 309–313. https://doi.org/10.1257/aer.p20161009.

Pärnamets, P., Johansson, P., Hall, L., Balkenius, C., Spivey, M. J., & Richardson, D. C. (2015). Biasing moral decisions by exploiting the dynamics of eye gaze. *Proceedings of the National Academy of Sciences, 112*(13), 4170–4175.

Polonio, L., Di Guida, S., & Coricelli, G. (2015). Strategic sophistication and attention in games: An eye-tracking study. *Games and Economic Behavior, 94,* 80–96. https://doi.org/10.1016/j.geb.2015.09.003.

Reutskaja, E., Nagel, R., Camerer, C. F., & Rangel, A. (2011). Search dynamics in consumer choice under time pressure: An eye-tracking study. *American Economic Review, 101*(2), 900–926. https://doi.org/10.1257/aer.101.2.900.

Shi, S. W., Wedel, M., & Pieters, F. G. M. (2013). Information acquisition during online decision making: A model-based exploration using eye-tracking data. *Management Science, 59*(5), 1009–1026.

Sims, C. A. (2003). Implications of rational inattention. *Journal of Monetary Economics, 50*(3), 665–690. https://doi.org/10.1016/S0304-3932(03)00029-1.

Smith, S. M., & Krajbich, I. (2018). Attention and choice across domains. *Journal of Experimental Psychology: General*. https://doi.org/10.1037/xge0000482.

Smith, S. M., & Krajbich, I. (2019). Gaze amplifies value in decision making. *Psychological Science, 30*(1), 116–128. https://doi.org/10.1177/0956797618810521.

Stewart, N., Hermens, F., & Matthews, W. J. (2015). Eye movements in risky choice. *Journal of Behavioral Decision Making*. Retrieved from http://onlinelibrary.wiley.com/doi/10.1002/bdm.1854/full.

Towal, R. B., Mormann, M., & Koch, C. (2013). Simultaneous modeling of visual saliency and value computation improves predictions of economic choice. *Proceedings of the National Academy of Sciences, 110*(40), E3858–E3867.

Vaidya, A. R., & Fellows, L. K. (2015). Testing necessary regional frontal contributions to value assessment and fixation-based updating. *Nature Communications, 6,* 10120. https://doi.org/10.1038/ncomms10120.

Wang, J. T., Spezio, M., & Camerer, C. F. (2010). Pinocchio's pupil: Using eye tracking and pupil dilation to understand truth telling and deception in sender-receiver games. *American Economic Review, 100*(3), 984–1007. https://doi.org/10.1257/aer.100.3.984.

Webb, R. (2019). The (neural) dynamics of stochastic choice. *Management Science, 65*(1), 230–255. https://doi.org/10.1287/mnsc.2017.2931.

Woodford, M. (2014). Stochastic choice: An optimizing neuroeconomic model. *The American Economic Review, 104*(5), 495–500.