Finding the right fit: A comparison of process assumptions underlying popular drift-diffusion models

Nathaniel J. S. Ashby[*1]

Marc Jekel[2]

Stephan Dickert[3,4]

Andreas Glöckner[2,5]

[1]Technion – Israel Institute of Technology, Israel

[2]University of Hagen, Germany

[3]WU Vienna University of Economics and Business, Vienna

[4]Linköping University, Sweden

[5]Max Planck Institute for Research on Collective Goods, Germany

[*]Correspondence concerning this article should be addressed to: Nathaniel J. S. Ashby (E-mail: nathaniel.js.ashby@gmail.com, Department of Industrial Engineering and Management, Technion – Israel Institute of Technology, Technion City, Haifa 32000, Israel

Abstract

Recent research makes increasing use of eye-tracking methodologies to generate and test process models. Overall, such research suggests that attention, generally indexed by fixations (gaze duration), plays a critical role in the construction of preference, although the methods employed to support this supposition differ substantially. In two studies we empirically test prototypical versions of prominent processing assumptions against one another and several base models. We find that general evidence accumulation processes provide a good fit to the data. An accumulation process that assumes leakage and temporal variability in evidence weighting (i.e. a primacy effect) fits the aggregate data, both in terms of choices and decision times, and does so across varying types of choices (e.g., charitable giving and hedonic consumption) and numbers of options well. However, when comparing models on the level of the individual, for a majority of participants simpler models capture choice data better. The theoretical and practical implications of these findings are discussed.

Keywords: eye-tracking, attention, drift-diffusion models, evidence accumulation, choice.

Finding the right fit: A comparison of process assumptions underlying popular drift-diffusion models

Classic economic theory (von Neumann & Morgenstern, 1944) posits that preferences are stable and consistent, however a growing body of work indicates that preferences are instead dynamic, possibly being constructed in the moment (Slovic, 1995). One class of process models known as drift-diffusion models (DDMs) have shown great promise in their ability to predict choices of consumer products (Krajbich, Armel, & Rangel, 2011; Krajbich & Rangel, 2010), risky prospects (Busemeyer & Townsend, 1993), as well as experts sport decisions (Raab & Johnson, 2007). The basic assumption underlying all DDMs is that preferences are constructed through a dynamic information acquisition process. Specifically, individuals' preferences are formed by sampling from available options until the evidence - generally defined as an options' value - supporting one option is strong enough to induce a judgment.

While all DDMs share this foundation, the models themselves vary substantially in their complexity and formulation, as do the data processing techniques (e.g., how fixations are identified) employed to construct and test them. This individualistic approach to process modeling presents a direct problem for the advancement of cognitive theory as the ability to implement differing models across studies, while directly testing the efficacy of their processing assumptions against one another, becomes an arduous task. In the current paper we extract the core processing assumptions of such models in order to compare them in terms of their ability to predict decisions and decision (reaction) times (Krajbich et al., 2011; Luce, 1986; Ratcliff & Smith, 2004).

The Contenders

Baseline Models

Before one can argue that attentional allocation (operationalized as fixation durations in the current studies) is crucial to predictions of choice it must be shown that directed attention is solely, or in combination with subjective ratings of value, a better predictor than simple utility maximization models. In order to test if this is the case we generate and compare three baseline models.

The first (Base$_v$), is a simple utility maximization model predicting that option *i* will be chosen if its value (V_i) is higher than all other options' values, where V_i is determined solely by an individuals' subjective valuation of that option (s_i):

For all options *i*: $V_i = s_i$ Equation 1

The second base model (Base$_a$), is a simple DDM that predicts that the option *i* which receives the most attention (i.e., is fixated on for the longest period of time) will be selected. Thus, the value (V_i) for an option is equal to the sum of the gaze durations (*e*) of all fixations (*f*) to this option. We define *e* to be measured in seconds here and in the following[1]:

For all options *i*: $V_i = \sum_{j=1}^{f} e_j$ Equation 2

The final baseline model we consider is a Standard Accumulator Model (SAM) which predicts that an options' subjective-value (s_i) is multiplied by the sum of its gaze durations (*e*) across all fixations (*f*) and the option *i* with the highest resulting value (*V*) is selected:

For all options *i*: $V_i = s_i \sum_{j=1}^{f} e_j$ Equation 3

Primacy-Recency Models

The first type of process model we explore is the primacy-recency variant (Glöckner et al., 2012; Raab & Johnson, 2007). Models of this type suggest that incoming evidence is weighted more (recency) or less (primacy) than previously accumulated evidence. The underlying theoretical basis for models of this type stems from memory research, which consistently reveals robust serial position effects in memory recall (for review see, Laming, 2010). For example, items encountered earlier and later in a list of to be remembered items are recalled more frequently than items presented in the middle of the list (Murdock, 1962).

[1]Different scalings had no effect.

Thus, for this DDM variant, greater weighting for already accumulated information reflects the fact that there is an increased likelihood for earlier encountered information to be remembered, while greater weighting for incoming information would reflect the fact that recently encountered information is more easily recalled.

For our purposes we generate our Primacy/Recency Model (PRM) for the value of an object as a weighted sum of the previous value and the current value using one free weighting parameter δ as follows:

If fixation to object i: $V_{i(j)} = \delta V_{i(j\text{-}1)} + s_i * e_j$ Equation 4a

For all unattended objects k≠i: $V_{k(j)} = V_{k(j\text{-}1)}$ Equation 4b

The value (V_i) for an option i starts with a value of zero prior to the first fixation ($V_{i(0)} = 0$). The V_i at each subsequent fixation j is calculated by the subjective-value (s_i) of the fixated option multiplied by the gaze-duration in seconds (e) at fixation j to option i, and the previous V_i multiplied by δ, which can take any value from 0 to 10 and indicates a primacy or recency effect. Specifically, if δ is found to be smaller than one, recency is indicated while if δ is larger than one primacy is indicated. For options that are not the focus of attention, V_i remains unchanged (Equation 4b). Importantly, the PRM equates to the SAM if δ is set equal to one and equates to the Base_a if all s_i had the same value.

Leaky Accumulator Models

Another popular DDM process is the leaky accumulator, which posits that accumulated evidence about a given option diminishes over time (Usher & McClelland, 2001, 2004). The leaky accumulator variants are also formulated on theories related to memory processes, in particular the activation and inhibition of neural networks tasked with processing and storing information (Rumelhart & McClelland, 1986). In the current context this implies that when information supportive of an option is being accumulated (is the focus of gaze), retained information about other options loses some activation in their respective neural networks, due to decreases in neuronal firing patterns and inhibitory signals being sent by neurons representing the option currently attended. We implement our Leaky Accumulator Model (LAM) to include only one free parameter ω which takes a value between zero and one as follows:

If fixation is to option i: $V_{i(j)} = V_{i(j-1)} + s_i * e_j$ Equation 5a

For all unattended options k≠i: $V_{k(j)} = \omega^{ej} * V_{k(j-1)}$ Equation 5b

Equation 5a posits that the value (V_i) of option *i* at fixation *j* is equal to the previous accumulated value of option *i* plus new information about option *i*, again defined as subjective-value (s_i) multiplied by gaze-duration (e_j), but only when option *i* is the current focus of attention. For all other options *k≠i* that are not the focus of attention, Equation 5b is used where ω decreases the value of option *k* as a function of how long it is not attended to in seconds. Thus, values of ω closer to zero indicate stronger information leaks whereas values of ω closer to one indicate slower leaks; if ω is set to one then the LAM is equivalent to the SAM as well as the PRM if δ was set to one.

Discounting Models

The next class of DDM we consider is one which specifies a process where all options have evidence about them constantly accumulated, but options that are currently the focus of attention have evidence accumulated at an increased rate (Krajbich et al., 2011; Krajbich & Rangel, 2010). The discounted accumulator models build on findings that attention appears to bias accumulation processes leading to preferences for options which receive more attention (Armel, Beaumel, Rangel, 2008; Shimojo et al., 2004), a process linked to modulated activation of the ventromedial prefrontal cortex (Hare, Camerer, & Rangel, 2009). Our Discounted Accumulator Model (DAM) captures this bias with one free parameter ϴ which can take any value between zero and one:

If fixation is to option i: $V_{i(j)} = V_{i(j-1)} + s_i * e_j$ Equation 6a

For all unattended option k≠i: $V_{k(j)} = V_{k(j-1)} + \Theta * s_k * e_j$ Equation 6b

The value (V_i) of option *i* in the DAM is identical to the LAM when that option is the focus of attention (Equation 6a). The DAM and LAM differ when an option is not attended to. Specifically, when an option is not being attended to, evidence (s_k*e_j) is still accumulated but at a discounted rate with the discount factor being equal to ϴ (Equation 6b). Therefore, a ϴ of zero indicates no accumulation of evidence for non-attended to options (a strong bias of attention), while a ϴ of one indicates a constant accumulation of evidence for all options irrespective of where attention is directed (no attentional bias). As with the previous models, if ϴ is set to zero the DAM is equivalent to the SAM, as well as the PRM and LAM if their parameters (ω and δ) were set equal to one.

Leaky Primacy-Recency Model

The last class of model we consider is a novel one which combines the processing assumptions underlying the PRM and LAM forming a hybrid Leaky Primacy-Recency Model (LPRM). The LPRM can be thought of as a generic neural network model. Like the leaky accumulator leakage in the LPRM represents the decrease in activation across a given neural network when an option is no longer being reinforced (attended to). The primacy-recency parameter can be thought of as reflecting greater (faster) activation for network nodes that are already active from previous fixations if primacy is found, while a recency effect might be interpreted similarly to the discounting models with there being a boost in information processing for currently fixated options. We define the LPRM with two free parameters:

If fixation is to option i: $V_{i(j)} = \delta V_{i(j-1)} + s_i * e_j$ Equation 7a

For all unattended options k≠i: $V_{k(j)} = \omega^{e_j} * V_{k(j-1)}$ Equation 7b

Each option *i* starts with a value of zero when no fixations have been made ($V_{i(0)} = 0$). The V_i at each subsequent fixation *j* is calculated by the subjective-value (s_i) of the fixated option multiplied by the gaze-duration (e_j; at fixation *j* to option *i*) and the previous V_i multiplied by δ (Equation 7a), with δ (bound between 0 and 10) indicating either a primacy (δ > 1) or recency (δ < 1) effect. When options do not receive attention, Equation 7b is used with ω (bound between zero and one) indicating that the value of option *k* decreases as a function of how long it is not attended to.

Levels of Comparison

We compare each of the models laid out in the previous section in three ways: First, each model can be thought of as being deterministic, predicting that the option with the highest value (*V* in Equations 1 – 7) will be selected, thus we compare the models in terms of the percentage of choices they correctly predict. Second, each model can be thought of as providing a probability that an option will be selected; the value (V_i) of a given option relative to all other available options (see Equation 8 below). Thus, we can compare how strong a model predicts a given option will be selected both when that prediction is correct and when it is incorrect. Finally, to investigate a process measure beyond preference we analyze the model's ability to predict observed decision times (reaction times). The analysis of decision time has long been proposed to provide additional insight into the cognitive processes involved in decisions (Cartwright & Festinger, 1943; Luce, 1986) and the ability for process models to predict decision times has been suggested as one way to further differentiate a given model's ability to reflect or capture underlying processes (Ratcliff & Smith, 2004).

By analyzing choices and response times, we aim to determine which of the basic mechanisms considered accounts best for behavioral data, which should provide important hints concerning the cognitive mechanisms involved. Still, it should be noted that the full-fledged models discussed in the literature contain further auxiliary parameters and assumptions which are not tested here. Hence, conclusions concerning the validity of specific models cannot be derived from our research.

Study 1

In Study 1 we investigated which model better predicted individuals' decisions to help others. Specifically, we obtained subjective valuations of need (s_i) for various children then measured fixations while participants choose which child to donate to, allowing us to directly compare the models introduced above in a novel domain (i.e., charitable giving, but see Study 2 for choices involving more common hedonic consumption decisions). We use need ratings as our s_i as perceived need has been shown to be linked to decision to donate (Loewenstein & Small, 2007). Awareness of need is the first step proposed to underlie charitable behavior (Bekkers & Wiepking, 2010), and together with empathic feelings and perceived impact of a donation one of the primary psychological drivers underlying donation decisions (Dickert, Sagara, & Slovic, 2011).

Methods

Participants and Design. Ninety-two participants (58% Female; $M_{\text{age}} = 21$) with normal to corrected-normal vision participated in Study 1. Two additional participants could not be calibrated to the eye-tracker and were not tested. Study 1 preceded two unrelated studies and the total time for all studies was one hour. Individuals received 12 Euro for their participation.

Materials and Procedure. The options in the study consisted of 54 images of children (~ 3° x 3° of visual angle) obtained from several charitable organizations. Participants were first presented with each of the images, one at a time in random order, and provided ratings of need for each image (child). Need ratings were indicated by increasing the diameter of a circle (0 to 300 pixels) with a larger circle indicating more need[2]; these ratings served as the s_i used in the models. Participants then engaged in an unrelated experiment lasting approximately 20 minutes before being calibrated to the eye tracker; eye movements were recorded using the Eyegaze binocular system (LC Technologies) with a remote binocular sampling rate of 120Hz and an accuracy of 0.45° of visual angle.

Participants were informed that they would be presented with varying numbers of children in need and that for each set of children they should select the one child they would

[2] s_i were scaled to be between 1 and 100 in Study 1 and Study 2 for convenience. Scaling had no impact on the results reported.

most like to donate to if given the chance (donations were hypothetical; see Study 2 for fully incentivized decisions). The number of options available within a choice was manipulated as a within-subjects factor (random order) and included set sizes of 2, 3, and 4 and the same child was never shown more than once. Each set of choices was unique and which children were shown in a given choice set was randomly determined for each participant. In total, participants made 18 selections (six trials in each set size) by using the mouse to move a yellow circle (displayed at the center of the screen at the start of each trial) so that it encompassed their selection and clicking the left mouse button (see *Figure 1*). After choosing a child, participants were asked to provide sympathy, need, and donation impact ratings for the child they selected using the same response method employed in the pre-test. These responses were included for exploratory reasons and we do not report analyses of these variables as we had no *a priori* hypothesis relating them to the processing assumptions tested here[3].

-Insert *Figure 1* Here-

Eye Tracking Analyses and Model Fitting Procedure. Regions of interest (ROIs) were defined as ~ 4° x 4° visual angle squares centered on each option. Trials in which no fixations to any options were detected assumed random choice (e.g., 50% likelihood for an option being picked from two options, 33% from three, and 25% from four). Consecutive fixations on the same ROI or blank space were collapsed into a single fixation.

Model parameters were fit on the level of participant using a box-constrained quasi-Newton method (Byrd, Lu, Nocedal, & Zhu, 1995) as implemented in R (2014) with starting values for free parameters determined from a loose grid-search within parameter boundaries.

[3]The pre-test and post-choice need ratings were strongly related (r = .72, p < .001), as were the ratings between pre-test need and post-choice sympathy (r = .46, p < .001) and impact ratings (r = .41, p < .001). These convergent correlations provide support for our use of need ratings as a subjective measure of the value (s_i) our participants saw in donating to a given child.

Formally, for each model we used the Luce choice rule (Luce, 1959) to estimate the evidence speaking in favor of a selected option (Equation 8):

Luce choice rule: $$P(V_{i(f)}) = \frac{V_{i(f)}}{\sum_{l=1}^{L} V_{l(f)}}$$ Equation 8

Where the model's prediction for the probability (*P*) of option *i* being picked is its value $V_{i(f)}$ at choice, that is after the final fixation *f*, divided by the sum of values ($V_{l(f)}$) of all options L. In the maximum likelihood estimation the set of parameters is determined that maximizes (the sum of the log-transformed) likelihoods of all observed choices (i.e., $\sum\ln(P(V_{i(f)}))$). Parameters were searched on the level of participant (per person) simultaneously taking into account choices between two, three, and four options[4].

Results

Correct Prediction Comparison. We first compared each fitted model's ability to make correct predictions in a deterministic implementation. We find that all models perform well above the overall chance level of 36% (i.e. [1/2 + 1/3 + 1/4]/3; see *Figure 2,* upper row). However, models containing an influence of attention in addition to value made more correct predictions than the $Base_v$ model (all $t(91)$s > 4.83, all $ps < .001$)[5] and $Base_a$ model (all $t(91)$s > 3.60, all $ps < .001$). Comparing the LPRM (the model making the most correct predictions) to the LAM (the model making the second most correct predictions) we find the LPRM makes significantly more correct predictions on average ($t(91) = 3.42$, $p < .001$). Examining the best fitting model on the level of the individual we find a similar pattern of results with the LPRM making the most correct predictions for 20% of participant followed by PRM (9%), the LAM (8%), $Base_a$ (8%), $Base_v$ (7%), and DAM (1%): Note these totals are rounded and thus sum to greater than 100%. For nearly half of the participants (49%) no clear classification was

[4]Conducting separate fittings for each choice set size returned highly similar results.
[5]All of the following *t*-tests are one-sided paired *t*-tests.

possible since they could be best described by at least two models (LPRM and LAM: 10%, LPRM and PRM: 9%, PRM, SAM, and DAM: 4%, LPRM, PRM, LAM DAM, and SAM: 6%, all else < 4%).

-Insert *Figure 2* Here-

Controlling for the number of options (2, 3, or 4) presented to participants, the odds for a correct prediction of the LPRM increases by a factor of 1.79 in comparison to $Base_v$ (*Table 1*) and the Odds for a correct prediction when using LPRM versus the LAM increased by a factor of 1.16 (i.e. 1.79/1.54). Thus, in terms of our first measure of comparison we find the LPRM makes more correct predictions on the aggregate level and performs as well as PRM on the individual level, though its advantage over other models is meager.

-Insert *Table 1* Here-

Likelihood Comparison. To compare the fit of each model to the data in terms of log-likelihoods (the strength of prediction) while taking into account the number of free parameters, we computed the Akaike Information Criterion (Akaike, 1974) for model k according to:

$$AIC_k = 2*param_k - 2*\sum(\ln(P(V_{i(f)})))$$

The factor *param* equals the number of parameters of the model k (i.e., either 0, 1 or 2) when comparing the fit of models within participants and the number of parameters times the number of participants when comparing the overall fit. We then compared AICs using delta AIC: $\Delta_k = AIC_k - AIC_{min}$. A model k having values less than 2 in comparison to the best fitting model with the lowest AIC has substantial support, values between 4 and 7 indicate considerable support, while values greater than 10 suggest essentially no support for model k (Burnham & Anderson, 2002, p. 446). It is worth noting that the SAM is nested in the LAM and the PRM which are both nested in the LPRM. Thus in case a less complex model fits better, AIC differences can never be higher than the punishment term for the number of free parameters (i.e., 2 between SAM and LAM/PRM, 4 between SAM and LPRM, and 2 between LAM/PRM and LPRM).

The bottom rows of *Figure 2* display the delta LLs and AICs for each model. Examining overall AICs we find the LPRM outperforms each of the other models ($\Delta_{AICs} >$ 36), while when considering the best fitting model on the level of the individual the LPRM only fits 16% of participants best; the SAM fits the majority best (30%), followed by the LAM (21%), $Base_a$ (13%), PRM (12%), $Base_v$ (8%), and lastly the DAM (0%). Computing Δ_{AIC} for those individuals fit best by the LPRM to their next best fitting model we find very weak support for the LPRMs superiority (M = 1.92, Median = 1.06); for those fit best by other models the LPRM did on average about as good as they did (M = 5.17, Median = 2.36). Therefore, in terms of AIC, we find that the LPRM fits the data on the aggregate level best, while when categorizing on the level of the individual less complex sub-models of the LPRM (i.e., LAM, PRM, SAM) can describe 63% of participants better and models assuming no accumulation process can describe the remaining 21% of participants best.

-Insert *Figure 2* Here-

Decision Time Comparison. To examine the extent to which reaction times reflected s_i in general we first examined whether decision times were faster when the selected option had a higher s_i relative to other options as one would expect that when the selected option greatly dominated other options it would lead to faster decisions. Correlating skew corrected decision times[6] on $Base_v$ model's predicted choice probabilities (P_{Base_v}) per individual we find there to be a negative relationship indicating that the more one option dominated another in terms of *si* (i.e., at the start of the decision) the faster a decision was made (see the leftmost black dot in *Figure 3*).

[6]As would be expected decision times showed significant positive skew (Mardia Skew = 5.34, $\chi(1)^2$ = 1448.04, p < .001). We corrected for this by employing the following log correction equation which reduced skew to ~ .001: ln(Decision Time + .0872227).

To see how well each of the accumulator models could predict reaction times we first constructed absolute choice probabilities which corrected for the starting probability an option would be picked (P_{abs} = abs|P - P_{Base_v}|). We use P_{abs} rather than P because as indicated above s_i influences decision times, thus by removing P_{Base_v} we control for this effect allowing the P_{abs} of each model to reflect changes in the value of the selected option over the course of the decision rather than initial differences in s_i[7]. We expect that if a model is capturing the accumulation process a larger P_{abs} should predict longer decision times as it reflects greater changes in the value of the option relative to its s_i. The black dots in *Figure 3* display the median of individual correlations between the P_{abs} of each model and skew corrected decision times. For the SAM and DAM these relationships are negative and weak. While a small positive relationship was found between the PRMs P_{abs} and decision times, both the LAM and LPRM showed larger positive relationships between their P_{abs} and decision times. Thus, in terms of the models tested, only the PRM, LAM and LPRM show relationships in line with the predictions of the general evidence accumulation framework. For an overall evaluation of fit for these three models, we also regressed P_{abs} as predictor (with a random intercept for participants) on normalized decision times. The regression model for LPRM shows the best fit in terms of log-Likelihood (LPRM: -1526.69, LAM: -1533, PRM: -1539).

-Insert *Figure 3* Here-

Estimated Parameters. *Figure 4* displays the parameter estimates for each model. For the PRM we find that the majority of participants (77%) showed a primacy effect (*M* = 2.47; Median = 1.62; $CI_{95\%}$ [1.97, 2.95]). In addition, 39% of participants showed very strong primacy with estimated δ being greater than two. These primacy effects could be the result of

[7]Strict interpretation of the drift rate as an options subjective value would predict negative relationships between P and decision time for all models. From such a perspective, models with the strongest negative relation would provide the best fit to the response time data (i.e., the SAM in the current analysis). The last author of this manuscript supports such an interpretation.

initial fixations driving choice - choosing the first option seen. This is unlikely to be the case however, as initial fixations to an option were no more predictive than chance when two (*M* = .50; $CI_{95\%}$ [.46, .55]), three (*M* = .37; $CI_{95\%}$ [.33, .41]), or four (*M* = .22; $CI_{95\%}$ [.19, .26]) options were available.

The DAM's parameter ϴ was found to be at the lower end of the parameter space (*M* = .06; $CI_{95\%}$ [.02, .10]) providing insight into why the DAM fit poorly compared to other models, namely since there was limited evidence of discounting with a parameter close to zero. The LAMs ω parameter was less than one for 85% of participants (equal to 1 for 15% of participants) indicating a leak of information (about 30% per second) over time (*M* = .71; Median = .76; $CI_{95\%}$ [.66, .76]). The LPRM model combining a primacy/recency (δ: *M* = 3.37; Median = 2.14; $CI_{95\%}$ [2.72, 4.01]) and leakage (ω: *M* = .66; Median = .69; $CI_{95\%}$ [.61, .72]) parameter revealed similar parameter estimates as the PRM and LAM: 79% of participants showed a recency effect and 86% a leak of information over time. The correlation between the δ and ω was negative and strong which might suggest some dependency between them, suggesting caution in interpreting their estimated parameters ($r(90) = -.47$, $p < .001$).

-Insert *Figure 4* Here-

Discussion

In Study 1 we tested what improvements in fit could be obtained over a simple value maximization model by employing models integrating attention and various processing assumptions. In line with the general contentions of the DDM framework we find strong support for both attention and subjective value being important predictors of choice, with the SAM outperforming the other base models. In terms of the various processing assumptions tested we do not find very strong support for any single model: Each model captured some share of participants' data better than the others. Nevertheless, the AIC and percentage of correct predictions indicate that the LPRM provides the best overall fit, while both the LAM and LPRM were able to correctly predict decision times. Thus, while the results are mixed, there is some evidence that leaky processes, perhaps combined with primacy-recency processes, provide better fits to the data and increase predictive accuracy.

Study 2

While Study 1 provides support for the general role of attention in the preference construction processes, and weaker support for a leaky-primacy process providing the best fit (account) of this process, one might question the robustness of these findings and the ability of each model to cross-predict fully incentivized choices between different types of decisions (choices) and increased numbers of options. Therefore, in Study 2 we test the generalizability of the models by: (1) Expanding the number of options to choose from, (2) including different types of choices, and (3) making choices consequential (providing incentivization). In addition, drawing on the rich literature of out-of-sample predictions (Erev et al., 2010*;* Busemeyer & Wang, 2000) we also examine whether fixed parameter versions of each model can capture these new decisions. The fixed parameters were the median parameter values from Study 1: PRM (δ = 1.62), LAM (ω = .76), DAM (Θ = 0)[8], and LPRM (δ = 2.14; ω = .69).

Methods

Thirty-two new participants (59% female; M_{age} = 21) with normal or corrected to normal vision participated in Study 2 which lasted approximately 45 minutes. Participants first provided ratings of need for 70 children and appetitiveness ratings for 70 candy bars using a horizontal slider (500 pixels wide). As in Study 1, these ratings served as the subjective values (s_i) used in each model. Following a short filler task participants made choices between sets of either two, four, or eight candy bars (consumption choices) or children (donation choices as in Study 1): Five trials of each choice set size and type, amounting to 30 choices in total were made and trails were presented in random order for

[8]Note that DAM with Θ = 0 is identical to SAM; thus both models fit the data equally well when parameters are fixed.

each participant. Participants were informed that one of the children they chose would be selected at random and have 2.00 Euro donated to them on their behalf and one of the candy bars they chose would be randomly selected and given to them at the end of the experiment. Choices were made by moving a mouse wheel up/down - highlighting the options one by one with a white outline - and clicking the left mouse button once the option they wished to select was highlighted: Post-choice questions were not included. All other aspects were identical to Study 1.

Results

We fit each model using the same procedures described in Study 1. In addition, we assessed the fits of each model using the estimated median values of the parameter values from Study 1.

Correct Prediction Comparison. As in Study 1, we find that the LPRM makes the most correct predictions for both consumption and donation decisions (*Figure 2,* second and third column upper row). We find that all models perform well above overall chance level 29% (i.e., [1/2 + 1/4 + 1/8]/3) (see *Figure 2,* upper row). Comparing the LPRM to the next best fitting model, the LAM, the LPRM makes significantly more correct predictions for donation, $t(31) = 1.95$, $p < .05$, and consumption decisions, $t(31) = 2.79$, $p < .005$. In terms of each model's ability to predict on the level of the individual we find the LPRM makes the most correct predictions for the largest share of participants for donation decisions (25%) followed by LAM (9%), PRM (3%), and Base_v (3%). The majority of participants donation decisions (59%) could be best described by at least two models (LPRM and LAM: 13%, LPRM and Base_a: 6%, LPRM, LAM, and PRM: 6%, all else < 4%). For consumption choices, the LPRM also predicts most participants' choices best (19%) followed by Base_a (6%), and PRM (3%). However, the majority of participants consumption choices (72%) could be best described by at least two models (LPRM and LAM: 16%, LPRM and PRM: 9%, LPRM, LAM, and PRM: 9%, SAM, PRM, and DAM: 6%, all else < 4%). We find that the LPRM provides the most correct predictions (83%) when using the median values of the parameters estimated in Study 1 for consumption choices and performs similarly to the best model LAM (76% for LPRM versus 77% for LAM) for donation choices. The percentage of correctly predicted choices is almost identical between models with fitted versus fixed parameters providing some evidence that the current results are not simply the result of over-fitting. Thus, in line with the results of Study 1, we find the LPRM makes more correct predictions on the aggregate, using both estimated and out-of-sample parameter values, though on the level of the individual its support is relatively small over simpler models.

Likelihood Comparison. The bottom row, second and third column, of *Figure 2* displays the Delta AICs (i.e. the lowest AIC subtracted from values) for each model separately for consumption and donation decisions. As in Study 1 the LPRM outperforms each of the other models for consumption ($\Delta_{AICs} > 2.80$) and donation decisions ($\Delta_{AICs} > 25.41$). While the LAM fits the most participants best for consumption decisions (44%), the LPRM fit the most participants best for donation decisions (44%). Nevertheless, the improvements in fit for those fit better by the LAM in consumption decisions ($M = 1.15$; Median = 1.19) and those fit better by the LPRM in donation decisions ($M = 1.94$; Median = 1.02) over the next best fitting model was barely worth mention. For participants fit best by other models little decrease in fit was found comparing the best fitting model to the LAM in consumption ($M = 2.30$; Median = 1.73) and LPRM in donation ($M = 2.15$; Median = 1.70) decisions.

Examining the fits of each model using the median values of parameters estimated in Study 1 we also find that the LPRM provides the best fit for both consumption and donation decisions in terms of log-Likelihoods (shaded bars, *Figure 2*, second row). Thus, the LPRM provides the best overall fit in Study 2 whether parameters were fitted individually to choices or used median values of parameters from Study 1. Importantly, this again suggests that the good fit of the LPRM does not result from over-fitting.

Decision Time Comparison. As in Study 1 we examined whether decision times were faster when the selected option had a higher s_i by correlating skew-corrected decision times[9] per person on P_{Base_v} (black triangles and diamonds in *Figure 3*). As in Study 1, a negative relationship emerged such that decisions were made faster when the selected option had higher s_i.

For the SAM and DAM, the median of the individual correlations was only significantly different from 0 for consumption decisions. The PRM, LAM, and LPRM all showed positive relationships between their P_{abs} and decision times. For an overall evaluation of fit we also regressed P_{abs} as predictor (with a random intercept by participant) on normalized decision times as in Study 1 for donation and consumption choices separately. Similar to Study 1, the LPRM had the highest log-Likelihood value in comparison to all other evidence accumulation models (all differences above 14 for consumption choices and above 4 for donation choices). Importantly, these results replicate when using the median values of parameters estimated in Study 1 (*Figure 3*, grey symbols).

Estimated Parameters. The parameter estimates for both consumption and donation decisions are shown in *Figure 5*. For the PRM (δ) we find that the majority of participants show a primacy effect in both consumption (88%; M = 4.44; Median = 3.06; $CI_{95\%}$ [3.28, 5.60]) and donation decisions (91%; M = 3.51; Median = 2.17; $CI_{95\%}$ [2.43, 4.58]), though the relationship between parameter values across choice types was non-significant, $r(30) = -.09$, $p = .63$. As in Study 1, many participants (75% in consumption and 50% in donation decisions) showed very strong primacy with δ being estimated as > 2. Also as in Study 1, initial fixations were no more predictive than chance for consumption (49%, 24%, and 10%) and donation decisions (44%, 18%, and 12%) in each set size suggesting they were not driving the strong primacy estimates. The estimates of the DAMs Θ were found to be at the lower end of the

[9]Decision times showed significant skew (Mardia Skew = 56.78, $\chi(1)^2 = 9141.85$, $p < .001$) which we corrected for as in Study 1: ln(Decision Time - 1.018482).

parameter space for consumption (0 for all participants) and donation decisions (*M* = .003; Median = 0; $CI_{95\%}$ [-.001, .007]) as in Study 1. The LAMs ω parameter was less than one for the vast majority of participants for consumption (94%; *M* = .43; Median = .44; $CI_{95\%}$ [.33, .54]) and donation decisions (100% *M* = .41; Median = .39; $CI_{95\%}$ [.30, .51]) indicating a leak of information over time as in Study 1; the relationship in parameters across domains was non-significant, $r(30) = .19$, $p = .30$. For the LPRMs δ all participants showed a primacy effect for both consumption (100%; *M* = 5.59; Median = 4.60; $CI_{95\%}$ [4.44, 6.73]) and most for donation decisions (91%; *M* = 5.56; Median = 5.05; $CI_{95\%}$ [4.23, 6.87]) though the relationship across domains was non-significant, $r(30) = .18$, $p = .34$. For the ω parameter the majority of participants in both consumption (94%; *M* = .46; Median = .49; $CI_{95\%}$ [.35, .56]) and donation (100%; *M* = .45; Median = .46; $CI_{95\%}$ [.35, .55]) decisions had parameter values below one indicating an information leak over time; the relationship across domains was non-significant, $r(30) = .17$, $p = .36$. The correlations between δ and ω were negative and significant for consumption decisions ($r(30) = -.65$, $p < .001$), but non-significant for donation decisions ($r(30) = -.28$, $p = .12$. Thus, in terms of parameter values we find estimated parameters are similar across domains, fall in line with the average estimates of Study 1, but are not highly stable within an individual across domains for any model.

-Insert *Figure 5* Here-

Discussion

In Study 2 we replicate the finding that models which include an impact of attentional allocation provide a better fit than those relying on subjective values alone, for consumption and donation decisions, providing support for the proposition that attentional allocation plays a role in the preference construction process. In terms of our process comparisons for fitted models we find that overall the LAM and LPRM provide the best fits to the data, for both consumption and donation decisions, though their improvement in fit was not very large on the level of the participant. For the fixed parameter values estimated from Study 1 we find the

LPRM provides the best fit, even outperforming the PRM and DAM using parameters estimated from the data, though again the increase in fit was rather small. Furthermore, while we find average parameter values to be similar for both kinds of decisions in each model the within-participant stability was not strong across choice types. In terms of the ability to predict decisions times we again find that the LAM and LPRM provide the most predictive power. Thus, while a leaky accumulation process provides the best fit overall only minimal support is provided as to its superiority over simple accumulation processes.

General Discussion

In two studies we find evidence for the role of directed attention in the preference construction process, with models incorporating attentional allocation providing large increases in predictive accuracy for choices and a superior fit to decision times. These findings bolster assertions that DDMs can increase predictive accuracy for choices over classic value maximization models, and that evidence accumulation processes can at least in part be captured (reflected) by directed visual attention (i.e. fixation durations; Just & Carpenter, 1984).

In each study we find that a process which assumes a leak of accumulated evidence for options that are not currently the focus of attention, and an increased weight for earlier accumulated evidence, fits both pro-social and hedonic consumption decisions and decision times better on average, even when using parameters estimated from a different data set (i.e., out of sample prediction; Erev et al., 2010) and incentivizing decisions. Nevertheless, in terms of our results speaking directly to the superiority of leaky *plus* primacy processes (i.e., the LPRM model) it must be noted that the improvement in fit they provided was not very large and its greater fit on the aggregate was primarily driven by the result that the LPRM fit between 16% (Study 1) to 44% (Study 2 donation) participants better than any other model. Furthermore, the majority of participants in both studies was described best by a less complex evidence accumulation model that either did not include the primacy/recency process (LAM), the leaky process (PRM), or both (SAM). As such we do not wish to suggest that the LPRM model is “better” than all other models, or that any of the models we examined here are superior for that matter. Instead, each model appears to capture some participants’ behavior better than others and less complex accumulation models (i.e., LAM, PRM, SAM) can generally describe a majority of participants best, indicating that there is some variation in the types of cognitive processes underlying individuals’ decisions. Thus, the current results suggest that future explorations comparing DDMs, or any other computational models for that matter, are best examined on the level of the individual rather than on the aggregate.

In addition, it is important to note that we reduced previously developed models to their central processing assumptions. Thus, these models were not tested in their more complex implementations. This was done as we were most interested in providing a clear comparison of their fundamental processing mechanisms rather than specific model implementations (e.g., inclusion of temperature or noise parameters, or decision thresholds). However, it is possible that with more complex implementations other processes might perform as well, or even outperform, the processes underlying the LAM and LPRM. Therefore, future comparisons of these processes, which further refine the models housing them, are necessary before strong conclusions are drawn about the processes that underlie these types of decisions. It is also important to examine how well these models and their underlying processes perform when options are defined by multiple attributes (see for review, Gwo-Hshiung, 2010). Such investigations would allow for greater insight into the abilities of each model to capture the information accumulation and storage of both positive and negative attribute information in addition to capturing the impact of attributes with differing degrees of importance.

The studies presented here provide a test of the process assumptions underlying popular accumulation models (Krajbich, et al., 2011; Raab & Johnson, 2007; Usher & McClelland, 2001) and find that on average a process in which accumulated information diminishes over time, and weights earlier accumulated information heavier provides a fairly good account for choices and decision times. This finding is relevant not only for those interested in process models, but more generally to anyone interested in the accumulation, storage, and use of information during choice. For instance, the leaky process combined with greater weighting for earlier accumulated evidence is roughly in line with the findings of decreasing excitation and inhibitory/excitatory activations in neural networks (McClelland et al., 2010; Thagard, 1989; Glöckner & Betsch, 2008). Thus, the current findings not only provide support for the general DDM framework and the ability of directed attention to capture the preference construction process, they motivate continued comparisons of such processes, particularly on the level of the individual, and the neurological aspects that underlie them.

References

Akaike, H. (1974). A new look at the statistical model identification. *Automatic Control, IEEE Transactions on*, *19*(6), 716-723.

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

Bates, D., Maechler, M., Bolker B., & Walker, S. (2014*). lme4: Linear mixed-effects models using Eigen and S4.* R package version 1.1-7.

Bekkers, R., & Wiepking, P. (2011). Testing mechanisms for philanthropic behaviour. *International Journal of Nonprofit and Voluntary Sector Marketing*, *16*, 291-297.

Burnham, K. P., & Anderson, D. R. (2002). *Model selection and multimodel inference: a practical information-theoretic approach*. Springer.

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

Busemeyer, J. R., & Wang, Y. M. (2000). Model comparisons and model selections based on generalization criterion methodology. *Journal of Mathematical Psychology*, *44*(1), 171-189.

Byrd, R. H., Lu, P., Nocedal, J. and Zhu, C. (1995) A limited memory algorithm for bound constrained optimization. SIAM J. Scientific Computing, 16, 1190–1208.Dickert, S., Sagara, N., & Slovic, P. (2011). Affective motivations to help others: A two-stage model of donation decisions. *Journal of Behavioral Decision Making, 24*, 361-376.

Cartwright, D., & Festinger, L. (1943). A quantitative theory of decision. *Psychological Review, 50*(6), 595-621.

Dickert, S., Sagara, N., & Slovic, P. (2011). Affective motivations to help others: A two-stage model of donation decisions. *Journal of Behavioral Decision Making*, *24*(4), 361-376.

Erev, I., Ert, E., Roth, A. E., Haruvy, E., Herzog, S. M., Hau, R., ... & Lebiere, C. (2010). A choice prediction competition: Choices from experience and from description. *Journal of Behavioral Decision Making, 23*(1), 15-47.

Glöckner, A., & Betsch, T. (2008). Modeling option and strategy choices with connectionist networks: Towards an integrative model of automatic and deliberate decision making. *Judgment and Decision Making, 3*(3), 215–228.

Glöckner, A., Heinen, T., Johnson, J. G., & Raab, M. (2012). Network approaches for expert decisions in sports. *Human Movement Science, 31*, 318-333.

Gwo-Hshiung, T. (2010). Multiple attribute decision making: methods and applications. *Multiple Attribute Decision Making: Methods and Applications*.

Hare, T. A., Camerer, C. F., & Rangel, A. (2009). Self-control in decision-making involves modulation of the vmPFC valuation system. *Science*,*324*(5927), 646-648.

Just, M. A., & Carpenter, P. A. (1984). Using eye fixations to study reading comprehension. *New methods in reading comprehension research*, 151-182.

Krajbich, I., Armel, C., & Rangel, A. (2011). Visual fixations and the computation and comparison of value in simple choice. *Nature Neuroscience*, 13, 1292-1298.

Krajbich, I., & Rangel, A. (2010). A multi-alternative drift-diffusion model predicts the relationship between visual fixations and choice in value-based decisions. *PNAS,* 108, 13852-13857.

Laming, D. (2010). Serial position curves in free recall. *Psychological Review, 117*(1), 93.

Loewenstein, G., & Small, D. A. (2007). The Scarecrow and the Tin Man: The vicissitudes of human sympathy and caring. *Review of General Psychology*, *11*, 112-126.

Luce, R. D. (1959). On the possible psychophysical laws. *Psychological review, 66*(2), 81.

Luce, R. D. (1986). *Response Times: Their Role in Inferring Elementary Mental Organization3* (No. 8). Oxford University Press.

McClelland, J. L., Botvinick, M. M., Noelle, D. C., Plaut, D. C., Rogers, T. T., Seidenberg, M. S., & Smith, L. B. (2010). Letting structure emerge: connectionist and dynamical systems approaches to cognition. *14,* 348-356.

Murdock Jr, B. B. (1962). The serial position effect of free recall. *Journal of Experimental Psychology*, *64*(5), 482.

R Core Team (2014). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna: Austria.

Raab, M., & Johnson, J. G. (2007). Expertise-based differences in search and option-generation strategies. *Journal of Experimental Psychology-Applied, 13,* 158-170.

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

Rumelhart, D. E., & McClelland, J. L. (1986). The PDP Research Group: Parallel distributed processing: Explorations in the microstructure of cognition. *Foundations, 1*.

Shimojo, S., Simion, C., Shimojo, E., & Scheier, C. (2003). Gaze bias both reflects and influences preference. *Nature neuroscience*, *6*(12), 1317-1322.

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

Thagard, P. (1989). Explanatory coherence. *Behavioral and Brain Sciences, 12*(3), 435-502.

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

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

von Neumann, J., & Morgenstern, O. (1944). *Theory of Games and Economic Behavior, 1st edition*. Princeton Univ. Press, 2nd edition, 1947.

Table 1. Logistic regression with random intercept per participant (Bates et al. 2014) with correct prediction (1 = yes, 0 = no) regressed on models and number of options presented.

	b	*se*(*b*)	*z*	Factor Odds (i.e., e^b)
Intercept	0.62	0.08	7.95***	1.86
Model (control = Base$_v$)				
Base$_a$	0.03	0.07	0.48	1.03
SAM	0.28	0.07	3.75***	1.32
PRM	0.39	0.08	5.19***	1.48
LAM	0.43	0.08	5.72***	1.54
DAM	0.29	0.07	3.90***	1.34
LPRM	0.58	0.07	7.58***	1.79
Set size (control = Two options)				
Three options	-0.21	0.050	-4.26***	0.81
Four options	-0.28	0.049	-5.72***	0.76

Note. Regression weights *b*s are log(Odds) for a correct prediction. Base$_V$ and two options are set as controls for dummy coding. *** $p < .001$.

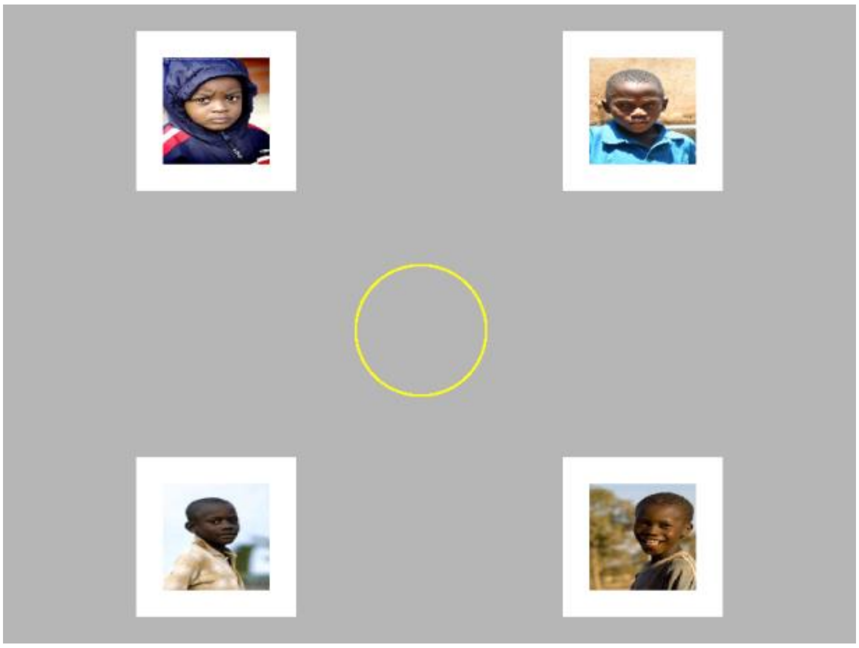

Figure 1. Screenshot of a selection trial with four options in Study 1; white outlines indicate ROIs which were not visible to participants; photographs were presented in color.

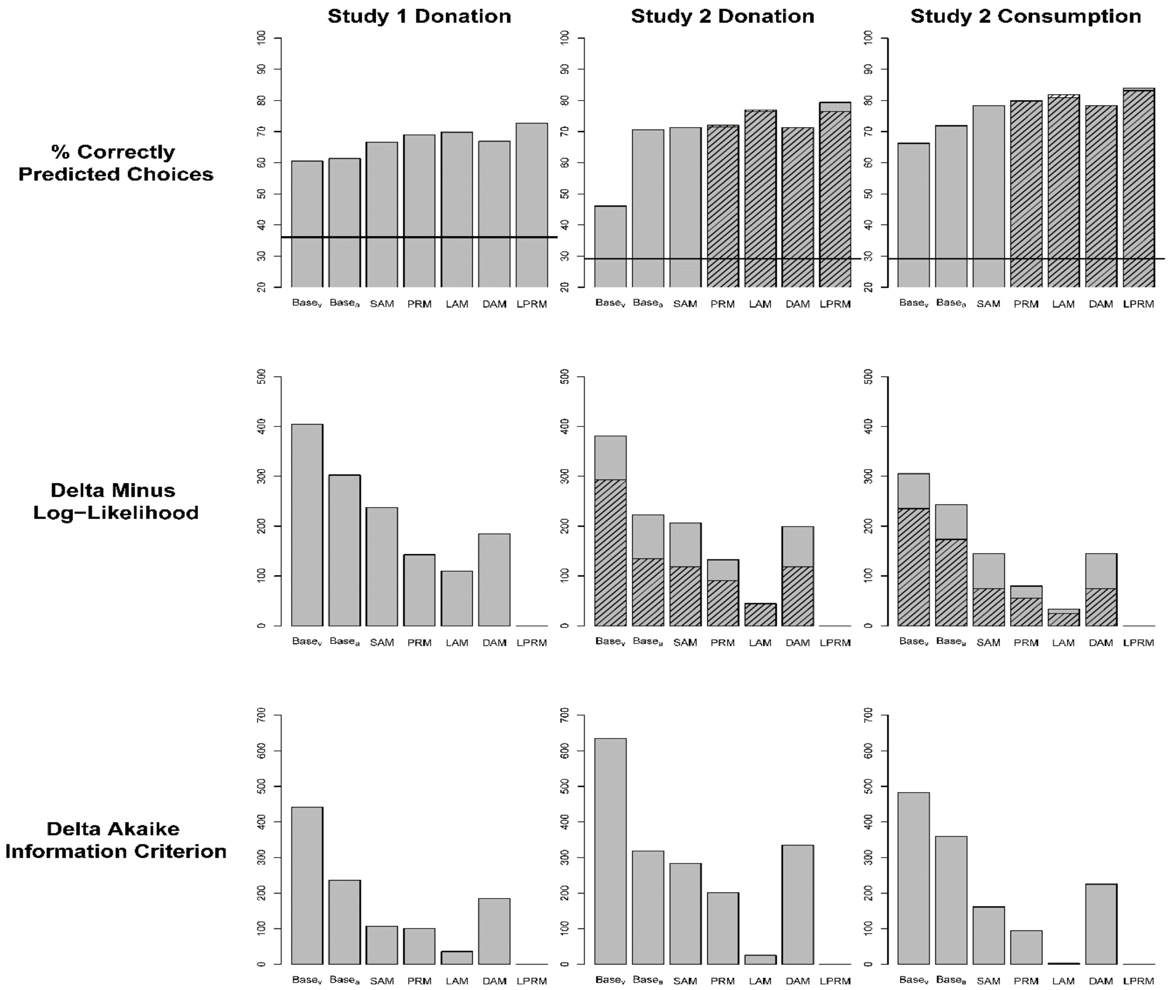

Figure 2. Bar charts displaying Percentage of Correctly Predicted Choices (first row; black horizontal lines indicating overall chance-performance across choice set sizes), Delta Minus Log-Likelihoods (second row), and Delta Akaike Information Criterion (third row) for Base_v, Base_a, the Standard Accumulator Model (SAM), Primacy Recency Model (PRM), Leaky Accumulator Model (LAM), Discounted Accumulator Model (DAM), and the Leaky Primacy-Recency Model (LPRM), for fits to decisions in Study 1 (left column), Study 2 Consumption (middle column), and Study 2 Donation (right column). Grey bars indicate evaluations of models with fitted parameters and overlapping shaded bars indicate evaluations of models when free parameters in Study 2 are fixed using best-fitting median parameters from Study 1. Note that Delta Minus Log-Likelihoods for models with no parameters also change when parameters are fixed since values result from subtraction with LPRM.

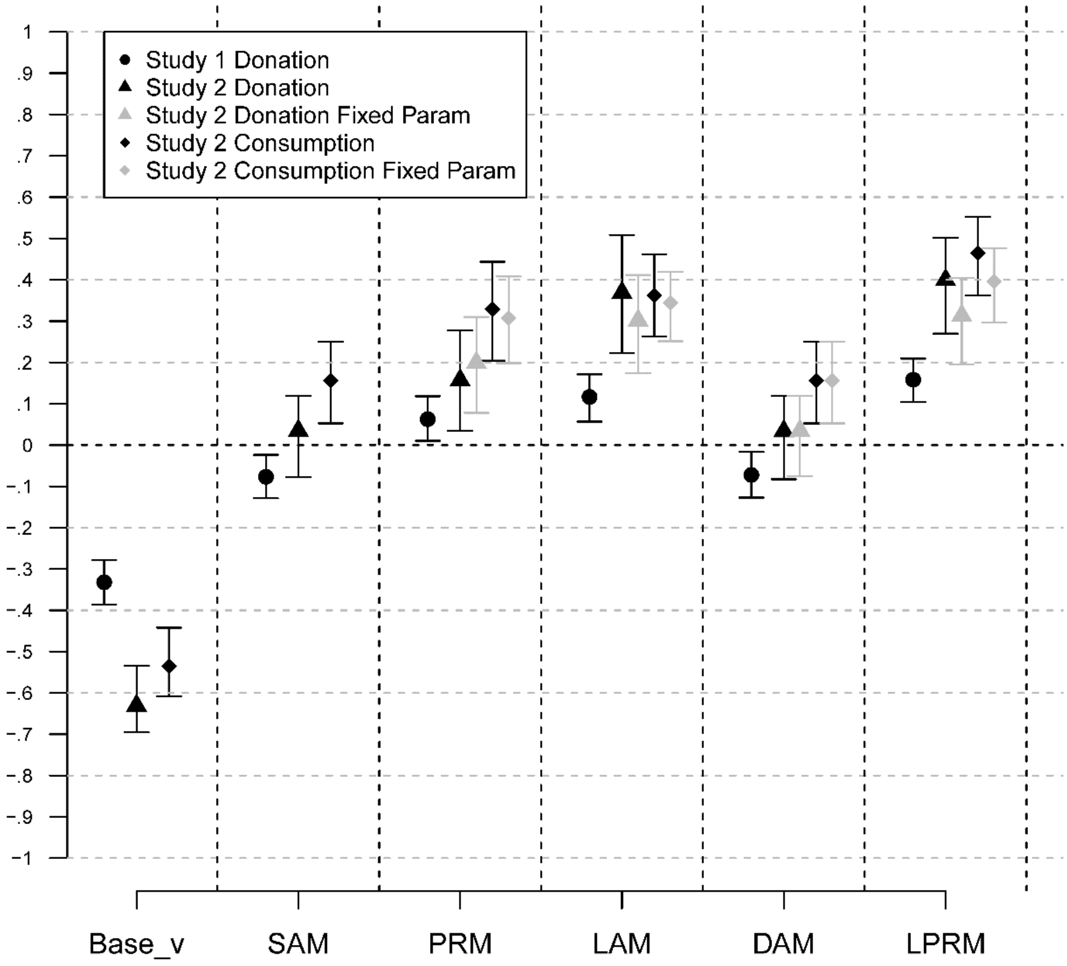

Figure 3. Estimator of the median and a nonparametric 95%-confidence intervals of individual correlations between normalized decision times and predicted decision times for models in both studies with fitted (black symbols) and fixed (grey symbols) parameters estimated from Study 1.

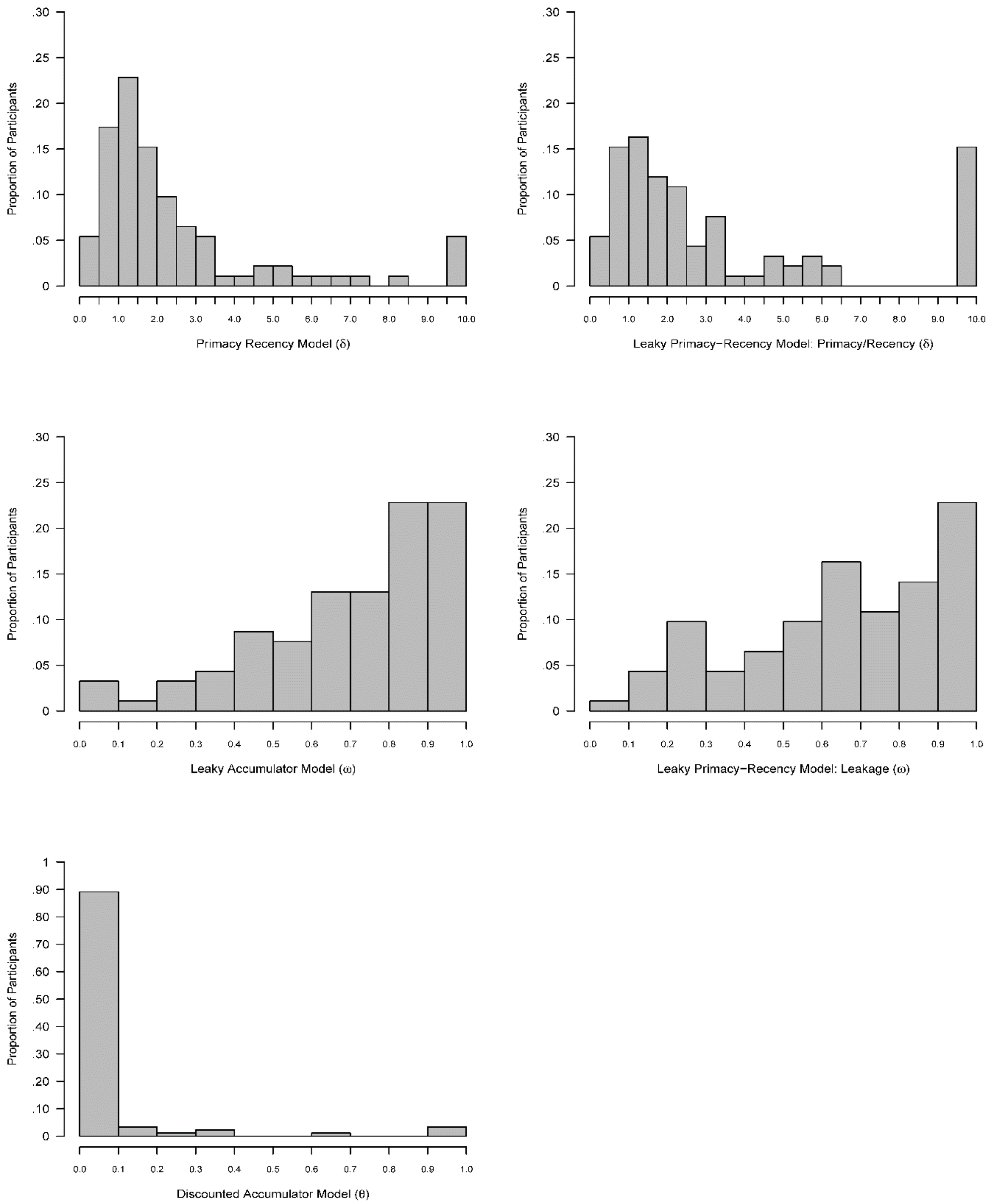

Figure 4. Histograms of best fitting parameters for each model in Study 1. Proportion (y-axis) of participants fit best by a given parameter value (x-axis) for the Primacy Recency Model (PRM), the Leaky Accumulator Model (LAM), the Discounted Accumulator Model (DAM), and the primacy recency parameter (LPRM - Primacy/Recency) and leakage parameter (LPRM – Leak) for the Leaky Primacy Recency Model.

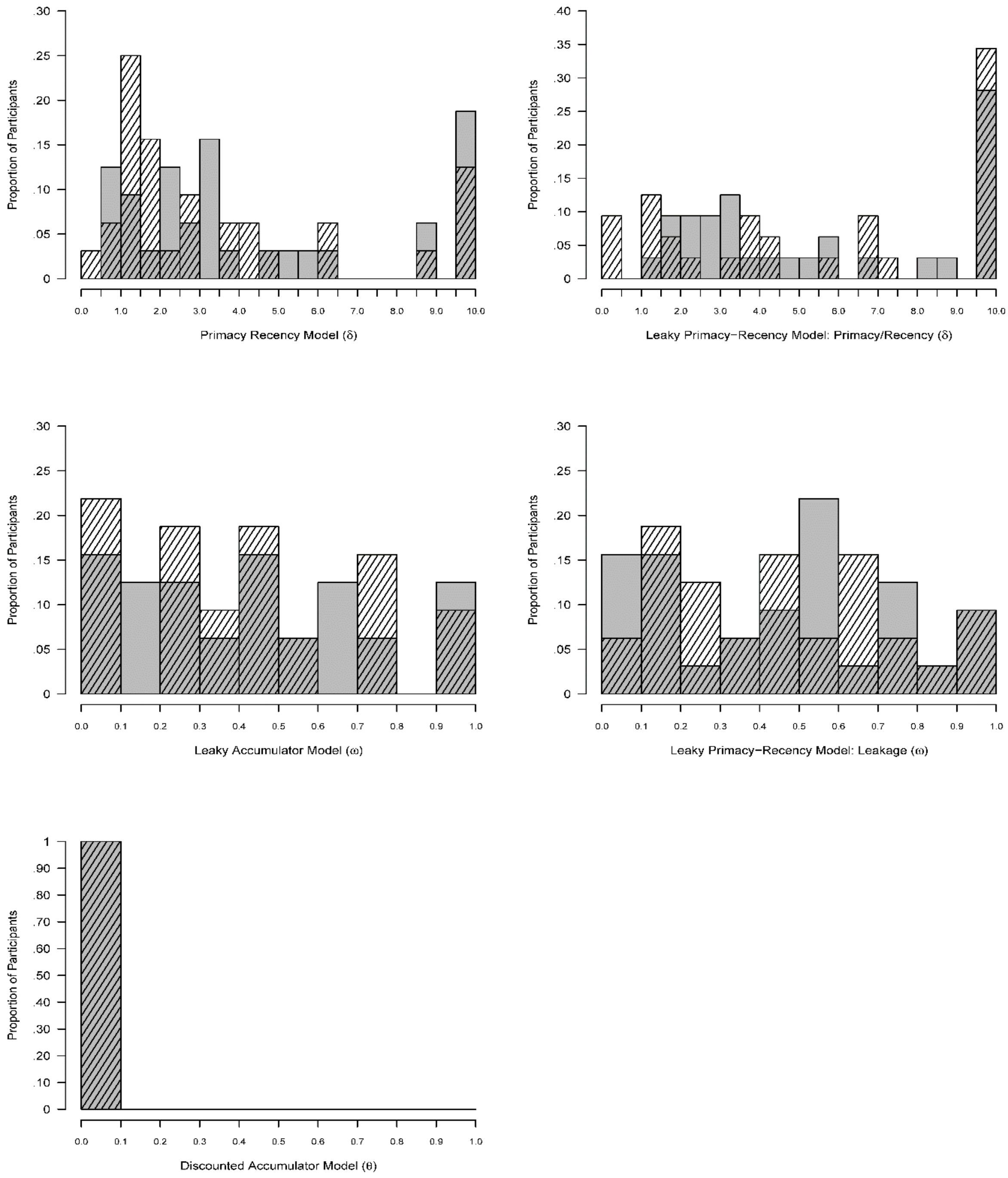

Figure 5. Histograms of best fitting parameters for each model in Study 2 over donation (shaded bars) and consumption (grey bars) decisions. Proportion (y-axis) of participants fit best by a given parameter value (x-axis) for the Primacy Recency Model (PRM), the Leaky Accumulator Model (LAM), the Discounted Accumulator Model (DAM), and the primacy recency parameter (LPRM - Primacy/Recency) and leakage parameter (LPRM – Leak) for the Leaky Primacy Recency Model.