

A Computational Model of the Attention Process in Risky Choice

Joseph G. Johnson

Miami University and Hanse Wissenschaftskolleg,
Delmenhorst, Germany

Jerome R. Busemeyer

Indiana University

A key component of most decision making theories is the decision weight that reflects the importance of consequences when evaluating an action. In algebraic models such as prospect theory or configural weight theories, these are explicit value multipliers with functional forms fit to empirical data. In process models such as decision field theory, elimination-by-aspects, or decision-by-sampling, they represent the distribution of attention about which simplifying assumptions are often made. Still, little is known about the cognitive processes that produce these weights. The current work presents a computational model that derives decision weights from elementary attention processing mechanisms. The basic idea is that a decision weight corresponds to the proportion of times an outcome is predicted to occur when an action is mentally simulated. We demonstrate the model's success by (a) coherently explaining several robust phenomena with a single set of parameters, including some that utility theories cannot; (b) fitting common functional forms used in algebraic models such as prospect theory, allowing for cognitive processing reinterpretations of their parameters; and (c) quantitatively fitting the model to extensive data sets on choice and pricing.

Keywords: Allais paradox, branch independence, probability weighting function, decision making, stochastic dominance

One characterization of recent theoretical work in judgment and decision making is the shift from algebraic, outcome-based models of behavior to those that attempt to more directly specify the underlying cognitive and emotional processes that produce these behaviors (see [Johnson & Busemeyer, 2010](#), for a review). For

the most part, these models have focused on the deliberation and choice processes (e.g., [Brandstätter, Gigerenzer, & Hertwig, 2006](#); [Busemeyer & Townsend, 1993](#); [Diederich, 1997](#); [Stewart, Chater, & Brown, 2006](#)). In the current work, we extend this process-oriented focus to the notion of decision weighting, or the process of evaluating the relative impact of different choice attributes on the subsequent deliberation process.

The contribution of the current work can be considered as fulfilling a missing component of existing process theories of decision making. For example, [Busemeyer and Townsend \(1993\)](#) make simplifying assumptions about decision weights and how they influence attentional processes during deliberation. [Diederich \(1997\)](#) develops similar models that allow for more formal specification about such transitions in attention across attributes. However, these models do not detail the source of these decision weights or the exact patterns of attention shifting. Rather, they treat these constructs as parameters of the model from which the resulting evaluation and choice processes can be pre-

This article was published Online First February 11, 2016.

Joseph G. Johnson, Department of Psychology, Miami University and Hanse Wissenschaftskolleg, Delmenhorst, Germany; Jerome R. Busemeyer, Department of Psychological and Brain Sciences, Indiana University.

The authors are extremely grateful to Doug Wedell for first suggesting this application of computational models to decision weighting, and to Michael Birnbaum, Marius Usher, Elke Weber, and Nathaniel Wilcox for additional constructive comments. We thank George Wu and Richard Gonzalez for making their data available to us. The MATLAB codes used to generate the predictions and figures in this work are available from the corresponding author by request.

Correspondence concerning this article should be addressed to Joseph G. Johnson, Department of Psychology, Miami University, Oxford, OH 45056. E-mail: johnsojg@miamioh.edu

dicted. Here, we will present a computational model that specifies simple cognitive processes that use stimulus information and individual characteristics to determine these attention weights over time. Additionally, we will relate the patterns of attention produced by this model to classical notions of decision weight or probability weighting present in algebraic theories of choice (e.g., Birnbaum, 2008; Kahneman & Tversky, 1979; Tversky & Kahneman, 1992).

Representation of Risky Decision Making and Typical Approach

An introduction to the applications and the notation that we use is required before we present our model. For the past 60 years or so, decision theorists have examined questions about risky decision making using simple gambles.¹ For example, suppose you had a choice between the following two options:

- Option A: Win \$10 for sure.
- Option B: Win \$100 with 1/10 chance, or \$5 with 4/10 chances, or nothing with 5/10 chances.

More generally, consider a gamble with n outcomes that produces outcome x_1 with probability p_1 , outcome x_2 with probability p_2 , and so on. For brevity, we will denote this gamble as $G = (x_1, p_1; x_2, p_2; \dots; x_n, p_n)$, where $x_1 < x_2 < \dots < x_n$; so Option B above would be expressed as (\$0, 0.50; \$5, 0.40; \$100, 0.10). The question of central concern for the current work is how much weight should be given to each of the three payoffs for option B. One reasonable approach is to weight each payoff by its probability, and then sum the three weighted outcomes. In other words, the *probability* of an outcome determines the *weight* given to that outcome. This results in the expected value (EV) rule for decision making: $EV(B_0) = (.5) \cdot ($0) + (.40) \cdot ($5) + (.10) \cdot ($100) = 12 , which is better than the certain \$10 offered by A, and so this EV rule would lead one to choose the option B over A.

Decades of research have shown that the EV rule, although in some sense rational, does not accurately describe or predict the choices people actually make when presented with choices such as that between A and B. In response, theorists have focused on modifications to this basic framework; notably, to the subjective

evaluations of outcomes and the weights placed on these outcomes. That is, they have proposed transformations of the outcome and/or probability information, while retaining the assumption of a multiplicative combination of these transformed values, to better describe choices (e.g., Birnbaum & McIntosh, 1996; Edwards, 1962; Kahneman & Tversky, 1979; Tversky & Kahneman, 1992; Quiggin, 1992). Generally, this can be expressed by defining the *weighted expected utility* of a gamble G as follows:

$$\begin{aligned} WEU(G) = & w_G(p_1, x_1) \cdot u(x_1) \\ & + w_G(p_2, x_2) \cdot u(x_2) + \dots \\ & + w_G(p_n, x_n) \cdot u(x_n) \end{aligned} \tag{1}$$

Here $w_G(p_i, x_i)$ refers to the weight given to outcome x_i , which may depend on the probability p_i of that outcome and/or the value of that outcome as described below. The key difference among various approaches within this algebraic framework has been in their claims about forms of $w(\cdot)$ and $u(\cdot)$, often referred to as the probability weighting and value functions, respectively. (Hereafter we will suppress the subscript for the gamble G in the weighting function to reduce notation). It seems that there are functional forms that can explain most of the empirical results from this research program (see Birnbaum, 2008). Although we do not think algebraic maximization is an accurate depiction of human decision making, one way to connect the current work to this vast literature is to consider how our model of attentional processes in risky decision making could approximate the decision weights for use in Equation 1; we return to this later.

A Process-Oriented Approach to Risky Decision Making

Several contemporary theories of risky decision making have adopted a process-based explanation of risky decision making. That is, they

¹ Although we acknowledge arguments by Goldstein and Weber (1995), among others, about the dangers associated with this “gambling metaphor,” we nevertheless adopt it in the current work to align with the relevant theory and literature. A detailed assessment of the adequacy of the metaphor is far beyond the scope of the current work.

do not suppose that people perform probability and value transformations, and then calculate values according to Equation 1. For example, some theorists assume that people use simple rules (Brandstätter et al., 2006; Thorngate, 1980) when making risky choices. Here, we adopt an approach that attempts to directly model the information processing assumed to underlie risky choices.

In particular, we advocate the use of a sequential sampling framework to model and understand decision making. This framework has been very useful in explaining not only risky choice (Busemeyer & Townsend, 1993) but also other aspects of decision making as well (e.g., Berkowitsc, Scheibehenne, & Rieskamp, 2014; Diederich, 2003; Hotaling & Busemeyer, 2012; Johnson & Busemeyer, 2005; Rieskamp, 2008; Roe, Busemeyer, & Townsend, 2001; Usher & McClelland, 2004; Bhatia, 2013; Trueblood, Brown, & Heathcote, 2014). Although the details of the model specification depend on the application, this approach in general can be described as follows. First, a decision maker's attention shifts over time across various aspects of the decision options. Depending on the momentary focus, the attended information is used to perform an evaluation of the associated options. Over time, these momentary evaluations change as attention shifts across different attributes, some of which may favor an option and some of which may be disadvantageous. These momentary evaluations are accumulated over time to form an evolving preference state relating the preference for each possible choice option. A decision is made when sufficiently strong preference is accumulated to favor one option over the others.

This approach is not only successful at accounting for several robust decision phenomena (e.g., Berkowitsch et al., 2014; Busemeyer, Jessup, Johnson, & Townsend, 2006; Busemeyer & Townsend, 1993; Diederich, 2003; Diederich & Busemeyer, 1999; Jessup, Bishara, & Busemeyer, 2008; Johnson & Busemeyer, 2005), but in making novel predictions about response times and other processing variables about which utility-based theories are silent. However, as Busemeyer and Townsend (1993) have pointed out, these models can also mimic the predictions of utility-based models, and may have parameters for which there are direct analogues in the utility-based algebraic represen-

tations (see also Lee & Cummins, 2004). For example, with the appropriate assumptions, the average rate at which options proceed toward the decision threshold is directly related to the utility difference between options when each is calculated according to Equation 1. In this manner, the model can faithfully reproduce the choice predictions of utility-based models even though it never assumes there is any calculation similar to Equation 1, and no integration of probability weighting and outcome information. Rather, this result is obtained if the relative attention to outcomes over the course of the deliberation process described above is equal to the decision weights, $w(p, x)$, and the outcomes are evaluated according to $u(x)$.

In empirical applications, this assumption about relative attention has been sufficient to explain the observed results, even when an identity function has been assumed where $w(p, x) = p_i$ (e.g., Scheibehenne, Rieskamp, & González-Vallejo, 2009). Furthermore, for mathematical tractability, it is also usually assumed that attention shifts stochastically according to a zero-order process based on these probabilities (but see Diederich, 1997). Although these basic assumptions have not deterred the basic model from achieving considerable success, the time has come to more accurately develop the attentional model in the sequential sampling framework. We turn now to meeting this goal.

A Process Model of Attention in Risky Choice

For clarity, we will start by reiterating exactly what we are trying to accomplish in this section. Recall that the sequential sampling approach assumes that attention at each moment, t , is directed toward one of the features of each choice option, and at that moment the feature is evaluated in order to update the preference state. For gambles in a risky choice paradigm, we assume that the different features consist of the different gamble outcomes. The goal, then, is to determine for each moment t , which outcome is selected as the focus of attention for evaluative processing.

The basic idea we formalize in the current model is that the decision maker tries to anticipate or predict what outcomes will be generated by each gamble. These predictions are

based on mental simulations of the random device actually used to generate outcomes (akin to the “simulation heuristic” of Kahneman & Tversky, 1982). For each of these mental simulations, the predicted outcome is then the focus of attention for the corresponding gamble during moment t .² We formally model which outcome is predicted for each of these mental simulations.

Consider for example, a decision maker faced with the following choice pair:

A: (\$12, .10; \$90, .05; \$96, .85),

B: (\$12, .05; \$14, .05; \$96, .90).

In a typical experiment involving these independent gambles, participants might be informed that each gamble is represented by a separate urn, bingo cage, and so forth, and that the outcomes of a gamble are randomly sampled therefrom. Using this device, the decision maker makes separate and statistically independent predictions for each gamble. Initially, we focus on the problem of predicting a three outcome gamble such as A. If we could directly monitor the decision maker’s attention, then we imagine that the process of generating a prediction would operate in the following way.

The decision maker first considers or “looks” at one of the outcomes. If the outcomes are processed from lowest to highest, then perhaps \$12 is considered first. There is some small chance the person will even select this outcome as a prediction on this mental play and predict this outcome to occur. However the probability of the \$12 outcome is rather low, so in the next moment, the person is more likely to “look” at the next higher ranked outcome, which is the \$90 outcome. At this point, the person may predict the \$90 outcome (which is again unlikely because of its low probability), or move back to reconsider the \$12 outcome, or move up to consider the \$96 outcome. Suppose the latter occurs, then the \$96 outcome could be selected as a prediction (which is likely considering its high probability), or the person could move back down to consider the \$90 outcome, and so on. Ultimately, this process would lead the person to predict one of the three outcomes \$12, \$90, or \$96 and finish the prediction process for this time step in the decision model, t . Whichever outcome is predicted (say, \$96),

then the evaluation of this outcome (say, $u(\$96)$) enters the decision model as the momentary evaluation of this option. Then, the above prediction process begins anew for the next moment ($t + 1$).

Markov Model

Formally, we can model this attention process as a Markov chain model (Karlin & Taylor, 1975). Figure 1 and the derivation here will focus on the Markov model for three outcomes $0 \leq x_1 \leq x_2 \leq x_3$, without loss of generality. The model has three states, one for each possible outcome that can be sampled from the random device used to generate outcomes. The attention process starts by “looking” at one of these three states. Whereas in the example we assumed initial consideration of the lowest outcome, more generally we assume the probability of starting in state x_i is denoted $z_i \geq 0$ where $z_1 + z_2 + z_3 = 1$. These three probabilities are called the initial state probabilities, which are represented by the downward pointing arrows in Figure 1.³

Once the attention process is located at a particular state, x_i , one of three mutually exclusive and exhaustive events can happen: (E1) a prediction could be made for that state, that is, predict x_i and stop, which is a downward transition from a state in Figure 1; or (E2) no prediction occurs, but attention remains focused on the same state, that is, dwell on x_i for another time step, which is the inner return loop shown in Figure 1, or finally (E3) attention moves to another state. Note that for (E3), we make a random walk type of assumption that transitions

² We are assuming that a single prediction takes place during each moment t in the choice model. Obviously, generating each of these predictions is not instantaneous but also develops according to the dynamics detailed in this section. It may be useful to keep this distinction between two different time scales in mind; see also Appendix A.

³ To minimize the number of free parameters, in the current applications the initial state probabilities are generated from a single free parameter denoted z , with $0 \leq z \leq 1$. This is achieved by defining the three initial state probabilities as follows: $z_1 = (1 - z)^2$, $z_2 = 2 \cdot z \cdot (1 - z)$, and $z_3 = z^2$. For example, if $z = 0$, then $z_1 = 1$, $z_2 = z_3 = 0$, and the process always starts with the lowest outcome; if $z = 1$, then $z_1 = z_2 = 0$ and $z_3 = 1$, and the process always starts with the highest outcome; and if $z = .5$, then $z_1 = .125$, $z_2 = .50$, and $z_3 = .125$, and the process is most likely to start with the middle outcome. A wide variety of initial distributions can then be generated using a single initial state parameter, z .

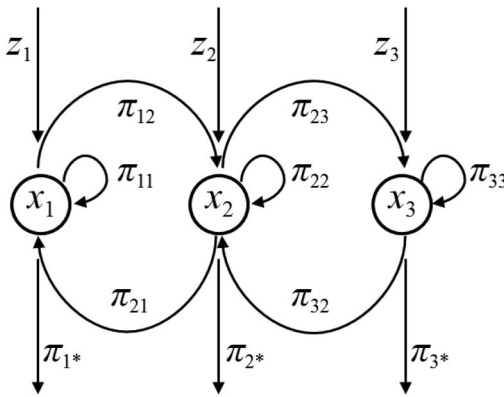


Figure 1. Schematic of the process weighting model. Circular nodes represent outcomes of a single gamble, and arrows represent possible probabilistic transitions in attention during a single “mental play.” The probability z_i of first considering each outcome x_i is input to the model, along with the objective (stated) probabilities $\pi_i^* = p_i = \text{Pr}[x_i]$. The parameter β of the model determines π_{ii} for each outcome, with the remaining probability partitioned among transitions to adjacent outcomes, $\pi_{i(i+1)}$ and/or $\pi_{i(i-1)}$.

may occur only to adjacent states, either one step to the right from state x_1 , one step to the left from x_3 (generally, x_n), or one step in either direction from x_2 (generally, any x_i , for $i \neq 1, n$). One can of course make other transition assumptions, but this random walk process provides a mechanism for drifting up and down the utility scale. This mechanism is similar to other perturbation models, such as those producing systematic errors in order and position information in short term memory (Estes, 1972).

To fully specify the model, we must determine the probability of each event. First, the probability of event E1 is denoted π_i^* in Figure 1, which we will assume to be equal to each outcome’s probability, $\pi_i^* = p_i$. In other words, given that an outcome is being considered, the probability of it being predicted is simply equal to its objective probability (similar, to the probability matching law for prediction; Estes, 1959). For Gamble A, this means $\pi_1^* = .10$, $\pi_2^* = .05$, and $\pi_3^* = .85$. Next, it follows that considering an outcome will *not* produce a prediction with probability $(1 - \pi_i^*)$, in which case we assume there is a constant “dwell rate” $0 \leq \beta \leq 1$ to describe event E2. This is the probability that this person’s attention remains focused on outcome x_i for another time step; denoted π_{ii} in

Figure 1, and equal to $\pi_{ii} = \beta \cdot (1 - \pi_i^*)$. Finally, the probability for event E3 depends on whether attention is focused on one of the end states or an intermediate state. If attention is focused on the lowest payoff x_1 , then the probability of shifting focus up to x_2 must equal $\pi_{12} = (1 - \pi_1^*) \cdot (1 - \beta)$; similarly, if attention is focused on the highest payoff x_3 , then the probability of shifting focus down to x_2 must equal $\pi_{32} = (1 - \pi_3^*) \cdot (1 - \beta)$. These are both required because the three events are mutually exclusive and exhaustive, so the transition probabilities out of a state must sum to one. If attention is focused on an intermediate state, x_2 in Figure 1, then the probability of transitioning to the adjacent state equals $\pi_{21} = \pi_{23} = (1 - \pi_2^*) \cdot (1 - \beta)/2$. That is, we assume that attention is equally likely to step up or down the outcome rank. Generally, $\pi_{i,i-1} = \pi_{i,i+1} = (1 - \pi_i^*) \cdot (1 - \beta)/2$, for all $i \neq 1, n$. In sum, all of the state transitions are computed from the objective outcome probabilities (p_1, p_2, \dots, p_n) and a single parameter, the dwell rate β .

Deriving Model Predictions

It is straightforward to directly compute the final probabilities of predicting each outcome (i.e., the attention weights) from the Markov model using simple matrix operations. That is, we can assume that the process described above serves as an ongoing attentional input to the decision model, but simplify the resulting predictions by deriving the expected attention weights at each moment t in that model. We continue to derive these formulas using the three-outcome case for illustrative purposes.

First, we form an initial state probability row vector as $\mathbf{Z}' = [z_1 \ z_2 \ z_3]$ containing the probability of first considering each outcome. Second, we form a transient matrix \mathbf{Q} and an absorbing matrix \mathbf{R} from the outcome probabilities (p_1, p_2, p_3) and the dwell rate β :

$$\mathbf{Q} = \begin{bmatrix} \beta(1 - p_1) & (1 - \beta)(1 - p_1) & 0 \\ (1 - \beta)(1 - p_2)/2 & \beta(1 - p_2) & (1 - \beta)(1 - p_2)/2 \\ 0 & (1 - \beta)(1 - p_3) & \beta(1 - p_3) \end{bmatrix},$$

$$\mathbf{R} = \begin{bmatrix} p_1 & 0 & 0 \\ 0 & p_2 & 0 \\ 0 & 0 & p_3 \end{bmatrix}.$$

In this case, the attention weights can then be easily computed from the Markov model by the simple matrix equation (see [Karlin & Taylor, 1975](#)):

$$\mathbf{W} = \mathbf{Z}' \cdot (\mathbf{I} - \mathbf{Q})^{-1} \cdot \mathbf{R} \tag{2}$$

The matrix \mathbf{I} is a 3×3 identity matrix (ones on the diagonals, zeroes on the off-diagonals), and $(\mathbf{I} - \mathbf{Q})^{-1}$ is the inverse of the matrix $(\mathbf{I} - \mathbf{Q})$. The probabilities of predicting each of the three outcomes (i.e., the attention weights) are contained in the row vector $\mathbf{W} = [w_1 \ w_2 \ w_3]$. For example, if we apply this model to gamble A with $z_1 = 1$ and $\beta = .70$, we obtain the following:

$$\begin{aligned} & [w(\$12) \ w(\$90) \ w(\$96)] \\ &= [1 \ 0 \ 0] \begin{bmatrix} 1 - (.7)(.9) & -(.)3(.9) & 0 \\ -(.)3(.95)/2 & 1 - (.7)(.95) & -(.)3(.95)/2 \\ 0 & -(.)3(.15) & 1 - (.7)(.15) \end{bmatrix}^{-1} \\ & \times \begin{bmatrix} .10 & 0 & 0 \\ 0 & .05 & 0 \\ 0 & 0 & .85 \end{bmatrix} \\ &= [.40 \ .16 \ .44]. \end{aligned}$$

It is interesting to note that if we use the typical parameters from [Birnbau’s \(2008\)](#) TAX model ($\delta = 1$ and $\pi(p) = p^{.7}$), then the TAX model produces similar decision weights for gamble A: $w(\$12) = .37$, $w(\$90) = .26$, $w(\$96) = .37$.

[Equation 2](#) easily can be extended for use with any arbitrary number of outcomes, n (see [Appendix B](#) for details). For the simple but important case of binary outcomes, $n = 2$, then for the larger of the two outcomes [Equation 2](#) reduces to (see [Appendix B](#)):

$$w_2 = \frac{(1 - \beta)z_1p_2^2 + (1 - p_2\beta)z_2p_2}{(1 - \beta)p_1 + (1 - \beta)p_2^2 + \beta p_1p_2} \tag{3}$$

Note that attention weights sum to one, so the weight for the other (lower) outcome is $w_1 = (1 - w_2)$. Also, the initial probabilities must sum to one, so we set $z_1 = (1 - z)$ and $z_2 = z$. Therefore, the model has only two free parameters (β and z). Even when there are more than three outcomes, we can compute all of the initial state probabilities in \mathbf{Z} from a single free parameter, such as by assuming that this initial probability has a binomial distribution (see

[Footnote 3](#) and [Appendix B](#) for details). In sum, the general model for n outcomes still has only *two* free parameters, the initial state parameter z and the dwell rate parameter β .

Psychological Interpretation of Parameters

We stress that never do we assume that [Equations 2](#) and [3](#) are ever computed by the decision maker. Instead, these equations are the mathematically expected consequences of the attention process summarized in [Figure 1](#) at any moment t in the decision process. There are only two free parameters in the attention process model: z determines the initial distribution \mathbf{Z} , and β determines the tendency to dwell on a particular outcome or state. Individual differences in the initial state parameter, z , and/or dwelling probability β can produce various mappings of probabilities to decision weights. This is important because individuals actually display a very wide range of decision weight patterns (see [Luce, 2000, Ch. 3](#)).

Conceptually, one can think of the attention process as undergoing a “search,” where the initial probabilities z_i affect where the search starts, the dwelling probability β represents how resistant the search is to move along the outcome scale, and p_i represents the tendency to stop the search with a prediction. Thus, correspondence between an outcome’s objective probability and its decision weight depends on the likelihood that the process starts out considering the outcome. Simply put, the process is more likely to terminate close to where it started than far from where it started, all else being equal. Note that this general principle is in line with research on “coherent arbitrariness” that shows the influence of (even irrelevant) anchors on decision processes (e.g., [Ariely, Loewenstein, & Prelec, 2003](#)). Furthermore, this is magnified if the process is likely to dwell (resistant to adjustment). This property of the model follows from the random walk arrangement of transition probabilities.

This feature of the model is illustrated in [Figure 2](#), which plots the decision weight given to the higher outcome in a binary gamble as a function of the two free parameters. In [Figure 2a](#), the process is assumed to always start at the lower outcome, $z_1 = 1$; in this case, the model produces consistent underweighting of the higher outcome, where the marginal effect (cur-

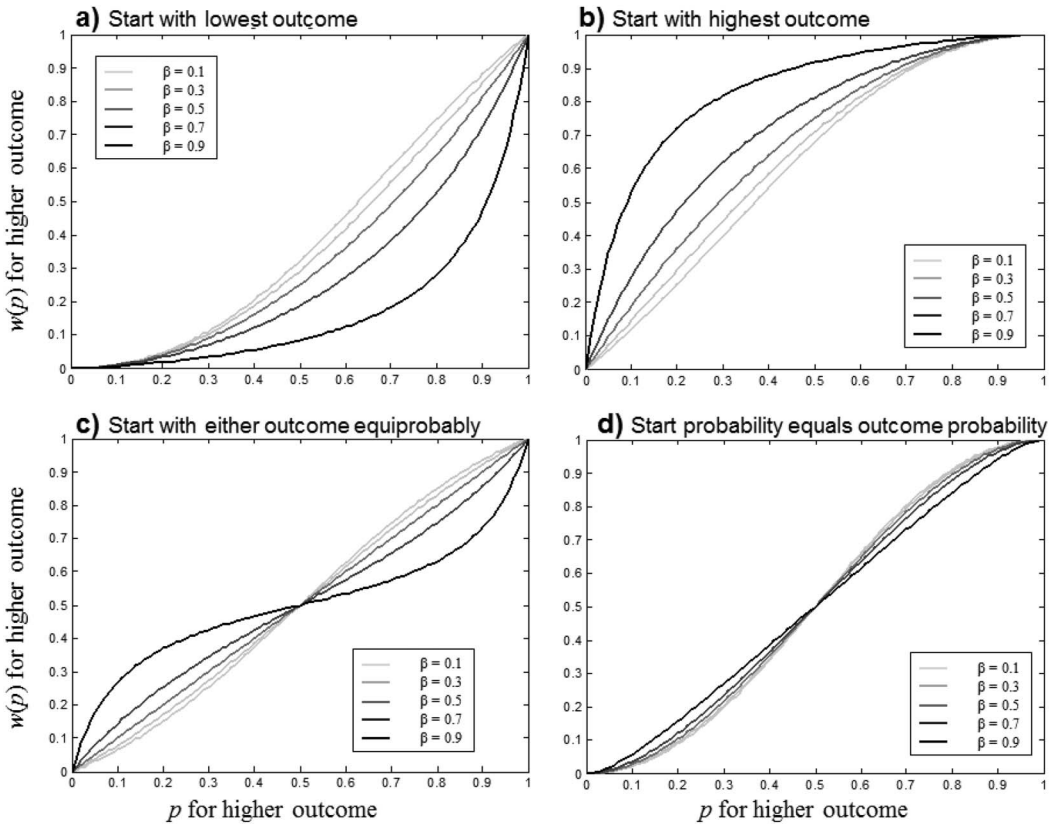


Figure 2. Representative plots of the process weighting model. Plots show value of the weight given to the larger outcome, $w(p_n)$ against p_n for $n = 2$, where different lines represent different values of β , and different panels reflect different assumptions about the initial state of the model. (a) Deterministic starting with the lower outcome, $z_1 = 1, z_2 = 0$. (b) Deterministic starting with the highest outcome, $z_2 = 1, z_1 = 0$. (c) Uniform starting distribution, $z_i = 1/n = [1/2]$, for $i = 1, 2$; and (d) Starting probabilities equal to objective probabilities, $z_i = p_i = \text{Pr}[x_i]$, for $i = 1, 2$.

vature) increases with increased dwelling (β). In contrast, if the process always begins with consideration of the higher outcome ($z_2 = 1$, Figure 2b), then there is global overweighting of this outcome, with a similar dependence on β for the marginal effect. Assuming a uniform distribution across initial states ($z_1 = z_2 = .5$, Figure 2c), the model's qualitative behavior depends predictably on the dwelling parameter. Specifically, for $\beta > 0.5$, the model exhibits overweighting of small to moderate probabilities, and underweighting of moderate to large probabilities; for $\beta < 0.5$, this trend is reversed; and linear weighting is produced when $\beta = 0.5$. Finally, in Figure 2d, the initial probabilities are set equal to the objective outcome probabilities

($z_i = p_i$), and the model exhibits mild underweighting of small to moderate probabilities (because the process is less likely to start with small probabilities when $z_i = p_i$) and overweighting of moderate to large probabilities, with minimal dependence on β (Figure 2d).⁴

Empirically, the existing evidence is mixed regarding which shape of probability weighting is most plausible; this may also depend on the

⁴ It is interesting to note a strong, testable prediction of the model in these last two cases: that an inflection point always occurs at an objective probability $p = .5$ for two-outcome gambles. However, this does not necessarily require the "fixed point" of the model to equal 0.5 (all $p < .5$ are overweighted and all $p > .5$ are underweighted).

specific task and stimuli. Early work (Preston & Baratta, 1948) and recent studies (Gonzalez & Wu, 1999) suggest the inverse S shape, whereas other evidence suggests either a strictly concave (Edwards, 1955; Tversky, 1967) or strictly convex (Kahneman & Tversky, 1979) form. Stott (2006) conducted a comprehensive examination of various choice, value, and decision weighting functions including a review of previous literature as well as a simultaneous comparison of 256 functional combinations; his results seem to generally support the inverse S shape. In a similar approach, Booij, van Praag, and van de Kuilen (2010) also review the published literature and also collect data from nearly 2,000 people from the general public supporting the inverse S shape.

Connections to Algebraic Utility Models

To be clear, our motivation was to provide a dynamic processing mechanism for the shifting attention assumed in sequential sampling models of risky choice such as the decision field theory of Busemeyer and Townsend (DFT; Busemeyer & Townsend, 1993). On the one hand, this mechanism represents a “missing” component used to determine the attention weights for deriving predictions in DFT or other process models such as decision by sampling theory (Stewart et al., 2006). On the other hand, this mechanism could be used to determine the decision weights that are inserted into weighted utility rules of risky decisions, thereby providing a cognitive processing account of the decision weighting functions typically used in those theories. Therefore, we can easily conceive of the attention process as a general explanation for decision weights, which could be used in various types of decision rules or choice mechanisms. For example, the weights produced by the model could easily be substituted for each $w(p, x)$ in Equation 1.

In this context, it may be interesting to note that this attention process model can in fact reproduce the trends associated with several popular functions relating objective probability p to probability weight $w(p, x)$; see Figure 3 for characteristic examples. Most prominent forms of probability weighting functions can be characterized by two key psychological properties, here called *discriminability* and *attractiveness*, following Gonzalez and Wu (1999; Tversky &

Kahneman, 1992; Tversky & Wakker, 1995).⁵ Discriminability refers to the marginal effects of probability, or sensitivity to increases or decreases in probability. This property suggests greater slope or curvature in weighting functions. Attractiveness refers to global tendencies of the weight applied to probabilities, such that greater attractiveness corresponds to assignment of higher weights. This property suggests greater intercepts or elevation of weighting functions.

Generally, these two psychological properties can be modeled by two parameters. For example, some empirical applications have used the following two-parameter function (e.g., Lattimore et al., 1992; Tversky & Fox, 1995):

$$w(p, x) = \frac{\beta p^\gamma}{\beta p^\gamma + (1 - p)^\gamma} \quad (4)$$

Another two-parameter form has been offered by Prelec (1998; Eq. 3.2), which has also been shown to work well in empirical applications (e.g., Sneddon & Luce, 2001):

$$w(p, x) = e^{-\beta(\log(p))^\gamma} \quad (5)$$

Finally, it is possible to use only one parameter in defining a probability weighting function that still exhibits the shape of Equations 4 and 5. In fact, a special form of Equation 5, with $\beta = 1.0$, seemed to be the most successful function based on the work of Stott (2006). Another one-parameter form was proposed in the original cumulative prospect theory paper (Tversky & Kahneman, 1992) and has also been used quite extensively (e.g., Camerer & Ho, 1994; Wu & Gonzalez, 1996):

$$w(p, x) = \frac{p^\gamma}{(p^\gamma - (1 - p)^\gamma)^{\frac{1}{\gamma}}} \quad (6)$$

Of course, the process model can reproduce decision weights calculated from Equations 4 through 6 in a straightforward manner. If we assume $\beta = 1$, and then set z_i equal to $w(p_i, x_i)$

⁵ Some researchers, following Quiggin (1992), characterize probability weighting in terms of optimism (overweighting) and pessimism (underweighting).

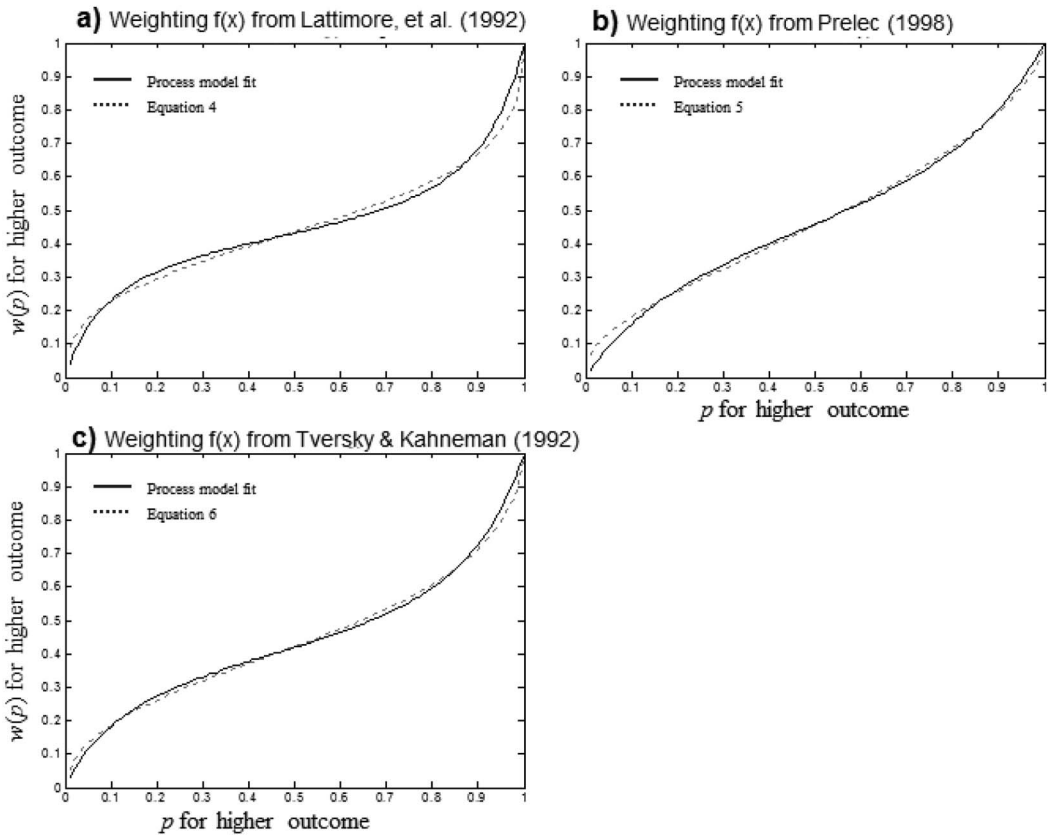


Figure 3. Fit of process model to popular probability weighting functions. Solid lines show process model fits to weighting functions (dashed lines) given in (a) Equation 4, (b) Equation 5, and (c) Equation 6. See text and Appendix C for details and model parameters.

from any probability weighting function, the resulting decision weights will correspond exactly to the given function. Our goal here, however, is to go beyond these trivial mappings and instead to see what psychological interpretations of our parameters correspond to resultant weights from other functions. To do so, we determined the value of z_1 and β (where $z_2 = 1 - z_1$) that minimized the squared residuals between the weight w_2 to the higher outcome in binary gambles predicted by our process model and the comparison model (see Appendix C for details regarding the model fitting procedures). Figure 3 shows the fit of the process model to the weights generated from Equation 4 using the parameters that were reported to best fit the median data by Gonzalez and Wu (1999), a special form of Equation 5 with $\beta = 1$, and

Equation 6 using the original parameter estimate from Tversky and Kahneman (1992).

There is not a great deal of variance across the process model parameters that produce the three plots shown in Figure 3—not surprising when one considers the similarity of the algebraic forms. The probability of attending to the lower outcome first, z_1 , ranged from 0.40 to 0.45; the value of β had a slightly larger range, from 0.79 to 0.91. In other words, the process model could very closely reproduce each and every of the algebraic functions by assuming a slight bias of initial attention to the higher outcome, and a moderately high probability of dwelling on a given outcome when it is considered.

In summary, we have shown that our process model can reproduce the weights achieved

through popular algebraic weighting functions. On one hand, this suggests it still makes perfect sense to simply use algebraic “resultant weights” for modeling simplicity in research applications. That is, although we describe a process, if it mimics other algebraic accounts and supports their assumptions about resultant decision weights, then in practice it is still justifiable to use the algebraically derived weights. On the other hand, however, perhaps our approach also accounts for data and/or phenomena that the algebraic equations cannot (and/or more parsimoniously accounts for similar data). To evaluate this possibility, we next consider some formal empirical applications of the model.

Applications of the Attention Weight Process Model to Empirical Findings

The evolution of algebraic utility-based decision weighting (or probability weighting) functions has been driven largely by their ability to account for a number of puzzling phenomena in people’s decision making that questioned the simple EV rule as an adequate descriptive model. As a first empirical assessment of our process model, it should be able to account for these robust phenomena. Although we obviously cannot apply the model in a single report to all possible decision phenomena, we choose here to focus on some of the most well-studied phenomena from the past 60 years: the common consequence and common ratio effects, and violations of branch independence and stochastic dominance.

Common Consequence and Common Ratio Effects

Consider a choice between the following gambles (M stands for million):

$$A_1: (\$1M, 1.0)$$

$$B_1: (\$0M, .01; \$1M, .89; \$5M, .10)$$

A person may prefer option A_1 because he or she is sure to get a million dollars, and option B_1 has some risk of producing nothing. This result alone is not peculiar; it simply describes risk aversion. However, the independence and cancellation axioms of many utility theories require that the choice between A_1 and B_1 above

is consistent with the choice between A_1^* and B_1^* below (see [Busemeyer & Johnson, 2008](#), for an overview):

$$A_1^*: (\$0M, .89; \$1M, .11) = (\$0M, .89; \$1M, .01; \$1M, .10)$$

$$B_1^*: (\$0M, .90; \$5M, .10) = (\$0M, .89; \$0M, .01; \$5M, .10)$$

Note that the gambles on the left hand sides of the equal signs for A_1^* and B_1^* are logically equivalent (according to the rules of probability) to the gambles on the right hand side. Allow us to rewrite gamble A_1 to make comparisons clear:

$$A_1: (\$1M, 1.0) = (\$1M, .01; \$1M, .89; \$1M, .10)$$

Viewing the right hand sides, we see that all we did across the pairs of gambles $\{A_1, B_1\}$ and $\{A_1^*, B_1^*\}$ is change the \$1M outcome with probability .89 (common between A_1 and B_1) to a \$0M outcome with probability .89 (common between A_1^* and B_1^*). According to the EU rule, these common outcomes with a common probability just cancel out in the comparison, and so the independence and cancellation axioms require that (a) if a person chooses A_1 over B_1 then that person must also choose A_1^* over B_1^* ; (b) alternatively, if a person chooses B_1 over A_1 then that person must also choose B_1^* over A_1^* .

[Allais \(1953\)](#) actually invented the gambles shown above, and when he presented those on the left hand sides, he generally found that people chose A_1 over B_1 but they switched and chose B_1^* over A_1^* , thus violating the independence axiom. This initial research was replicated and extended later by [Kahneman and Tversky \(1979\)](#), who called this finding the *common consequence effect*.

Another example of a violation of the independence axiom is the *common ratio effect* ([Allais, 1953; Kahneman & Tversky, 1979](#)). First consider a choice between the following:

$$A_2: (\$0M, 0; \$3M, 1.0)$$

$$B_2: (\$0M, .20; \$5M, .80).$$

Most people choose A_2 over B_2 , which again is fine with utility theory. But according to the independence axiom, this implies that A_2^* should be chosen over B_2^* below:

$$A_2^*: (\$0M, .80; \$3M, .20) = (A_2, .20; 0, .80)$$

$$B_2^*: (\$0M, .84; \$5M, .16) = (B_2, .20; 0, .80).$$

Note that A_2^* is just a .20 chance to play A_2 otherwise nothing; and B_2^* is also a .20 chance to play B_2 otherwise nothing. Again, the common consequence of 0 with probability .80 cancels out in the comparison of A_2^* with B_2^* , and the common probability .20 divides out in this comparison. So the choice between A_2 and B_2 must be consistent with the choice between A_2^* and B_2^* . In contrast, Allais (1953) and Kahneman and Tversky (1979) found that most people prefer A_2 over B_2 and yet they also choose B_2^* over A_2^* , violating the independence axiom (again when presented in the format on the left-hand side of the equal sign).

To examine very clearly the ability of the attention process model to explain these qualitative patterns, we use a very stringent protocol to generate the predictions. First, we use a single set of parameters to derive predictions for all of the gambles, rather than allowing free parameters for each gamble or application. Specifically, we assume that the process always begins with the lowest outcome ($z_1 = 1$, or $z = 0$ per Footnote 3), and we use a moderate level for the dwelling probability ($\beta = 0.70$). Second, rather than relying on the DFT choice model which we theoretically endorse, we want to show that the attention model alone is accountable for predicting the results. Therefore, we based the choices on a simple weighted utility formula from Equation 1, where $w(x_i)$ is now computed from Equation 2.⁶ Third, we put the full explanatory burden on the attention weighting process model by using a strictly linear utility assumption, $u(x) = x$. By using this simple protocol, we show that our proposed weighting process—rather than parameter flexibility, the DFT choice mechanism, or utility valuation—can account for the results.

Some justification for starting with the lowest outcome is needed. This tends to increase the decision weights for lower rank outcomes above their objective probabilities (see Figure 2). First of all, similar results are obtained if z is low but not exactly zero, and setting $z = 0$ provides the simplest example. Second, we could introduce a risk aversion utility parameter (e.g., $u(x)^\alpha$, $\alpha < 1$), and then we would not need the initial state to start at such a low value. However, we want to avoid adding extra utility parameters, and so z needs to be low to produce some risk aversion. Third, this is consistent with the hypothesis of other popular models, such as

Table 1
Process Model Predictions for Pairwise Choices Given in Text

Gamble	WEU	w_1	w_2	w_3	w_4
Common consequence					
A₁	1.00	.00	1.00		
B ₁	.99	.03	.96	.00	
A ₁ [*]	.01	.99	.01		
B₁[*]	.04	.99	.01		
Common ratio					
A₂	3.00	.00	1.00		
B ₂	2.64	.47	.53		
A ₂ [*]	.10	.97	.03		
B₂[*]	.11	.98	.02		
Branch independence					
A₃	15.61	.68	.27	.05	
B ₃	14.94	.68	.27	.05	
A ₃ [*]	33.61	.68	.27	.05	
B₃[*]	39.41	.68	.27	.05	
Stochastic dominance					
A₄	61.77	.40	.16	.44	
B ₄	59.52	.24	.20	.56	
A ₄ [*]	49.19	.27	.28	.12	.33
B₄[*]	50.47	.27	.28	.12	.33

Note. Expected utility theories require choices of A and A*, or of B and B*, for all choice pairs {A, B} and {A*, B*} within each of the first three choice sets; rank-dependent utility theories require choice of A and A*, or of B and B*, in the last set, “Stochastic dominance.” In all sets, empirical evidence suggests majority choices for A and B* (in bold), which are predicted when substituting the process model weights in Equation 1, with $u(x) = x$, to produce $WEU(A) > WEU(B)$ and $WEU(B^*) > WEU(A^*)$ within each set. The values of w_i show the predicted weight to outcome i of each gamble, when outcomes are rank-ordered. See text for properties of choice options in each choice set. Weights may not sum to one in all rows because of rounding.

the TAX model (Birnbbaum, 2008) assumption that the lowest outcome takes attention away from higher outcomes, or that decision makers start by comparing the lowest payoffs in the priority heuristic (Brandstätter, Gigerenzer, & Hertwig, 2006).

As can be seen in Table 1, the process model correctly predicts each of the pairwise choices that give rise to the common consequence and common ratio effects. That is, the inconsistent choice pattern (A, B*) is predicted for each pair

⁶ Note that the predictions of DFT in this case would produce the same ordinal relations, although they would provide additionally indications of preference strength (choice probabilities).

of choices. Table 1 also gives the predicted weights in each situation.

Violations of Branch Independence and Stochastic Dominance

Birnbaum and colleagues (Birnbaum & McIntosh, 1996; see also Birnbaum & Chavez, 1997; Birnbaum & Veira, 1998) have shown how violations of *branch independence* refute perhaps the most popular algebraic model, prospect theory (Kahneman & Tversky, 1979). When presented with the choice $\{A_3, B_3\}$:

A_3 : (\$10, 1/3; \$24, 1/3; \$45, 1/3)

B_3 : (\$10, 1/3; \$12, 1/3; \$96, 1/3),

most people prefer A_3 over B_3 , presumably because A_3 gives only one chance for producing a ‘very poor’ outcome but B_3 provides two chances. However, consider the following:

A_3^* : (\$24, 1/3; \$45, 1/3; \$100, 1/3)

B_3^* : (\$12, 1/3; \$96, 1/3; \$100, 1/3).

Now, most people reversed and chose B_3^* over A_3^* because now B_3^* gives two chances of producing a ‘very good’ outcome and A_3^* gives only one chance. This is a violation of branch independence. Note that A_3 and B_3 share a common outcome with a common probability (1/3 chance of \$10), and this common term should cancel out for prospect theory. Note also that A_3^* and B_3^* simply replace the common outcome of \$10 with a common outcome of \$100 dollars, and again this common outcome would cancel out. After cancelling out the common terms, there is no difference between these pair of choices. Thus prospect theory (as well as many others) must satisfy branch independence (select A_3 and A_3^* , or select B_3 and B_3^*), which fails to explain these findings.

Branch independence violations are problematic for all theories in which the weight given to an outcome only depends on the subjective probability assigned to that outcome. Contrary to this assumption, the violations of branch independence indicate that the weight given to a 1/3 probability associated with the smallest outcome differs from the weight given to the same 1/3 probability when attached to the largest outcome. Instead, the weight given to an outcome seems to depend, not only its probability, but also on its *rank* order. Both rank-dependent utility models (RDU;

e.g., Quiggin, 1992; Tversky & Kahneman, 1992) and configural weight models (CWU; e.g., Birnbaum, 2008) can explain these effects.

However, it turns out that RDU theories are also deficient for explaining human choice behavior (Birnbaum, 2008). In particular, the RDU rule cannot explain violations of *stochastic dominance* (Birnbaum, 2005; Birnbaum & Navarrete, 1998; Birnbaum, Patton, & Lott, 1999; Loomes, Starmer, & Sugden, 1992). Consider the pair of gambles $\{A_4, B_4\}$:

A_4 : (\$12, .10; \$90, .05; \$96, .85)

B_4 : (\$12, .05; \$14, .05; \$96, .90).

People choose A_4 over B_4 presumably because A_4 has two ways to produce a ‘very good’ outcome but B_4 has only one way (as if the outcomes are equally likely even though they are not). However, B_4 stochastically dominates A_4 , which can be seen by considering the next pair of gambles:

A_4^* : (\$12, .05; \$12, .05; \$90, .05; \$96, .85),

B_4^* : (\$12, .05; \$14, .05; \$96, .05; \$96, .85).

Now people almost always choose B_4^* over A_4^* . This is because it is clear that B_4^* always produces at least as good or better result than A_4^* with the same probability, and so B_4^* stochastically dominates A_4^* . However, A_4^* is constructed from A_4 simply by splitting the \$12 with probability .10 into two parts: each part is \$12 with probability .05. Similarly, B_4^* is constructed from B_4 simply by splitting the \$96 with probability .95 into two parts: \$96 with probability .05 and \$96 with probability .90. Thus B_4 also stochastically dominates A_4 . People tend to violate stochastic dominance with the pair $\{A_4, B_4\}$ but satisfy it with the pair $\{A_4^*, B_4^*\}$ (Birnbaum, 2005; Birnbaum & Navarrete, 1998; Birnbaum, Patton, & Lott, 1999). Birnbaum (2008) endorses the TAX model, which succeeds in explaining the appropriate violations of stochastic dominance. Can our attention process model, as well?

To apply the current model, we used the exact same protocol and parameters described in the preceding application. The only changes were the gamble stimuli on which the process was applied. Table 1 again shows the ability of the model to accurately explain violations of branch independence and stochastic dominance.

Importantly, our model makes the correct prediction when the stochastic dominance is transparent, as in $\{A_4^*, B_4^*\}$.

Figure 4 shows how sensitive the model predictions for all four effects in Table 1 are to the exact parameter values used here. Specifically, the figure shows the parameter value combinations that produce each of the four effects considered in this section, given the protocol used here. It is interesting to note that the process model parameter values which predict these four effects to co-occur suggest a strictly convex weighting function (Figure 2a). It is an open question whether this type of weighting function is empirically supported among participants that simultaneously show all four effects.

Additional Predictions

Now we examine some new effects that manipulate attention without changing the outcomes or their probabilities. Note that none of the traditional decision models mentioned above, including Birnbaum's (2008) TAX model, have posited any formal mechanisms for explaining these manipulations.

Advertisement or "attention grabbing" effects. The attention process model makes predictions concerning the effects of advertisements on decision weighting, which grab attention but provide no logical information about outcome probabilities. For example, Weber and Kirsner (1997) found that increasing the perceptual salience (e.g., font size) of gamble out-

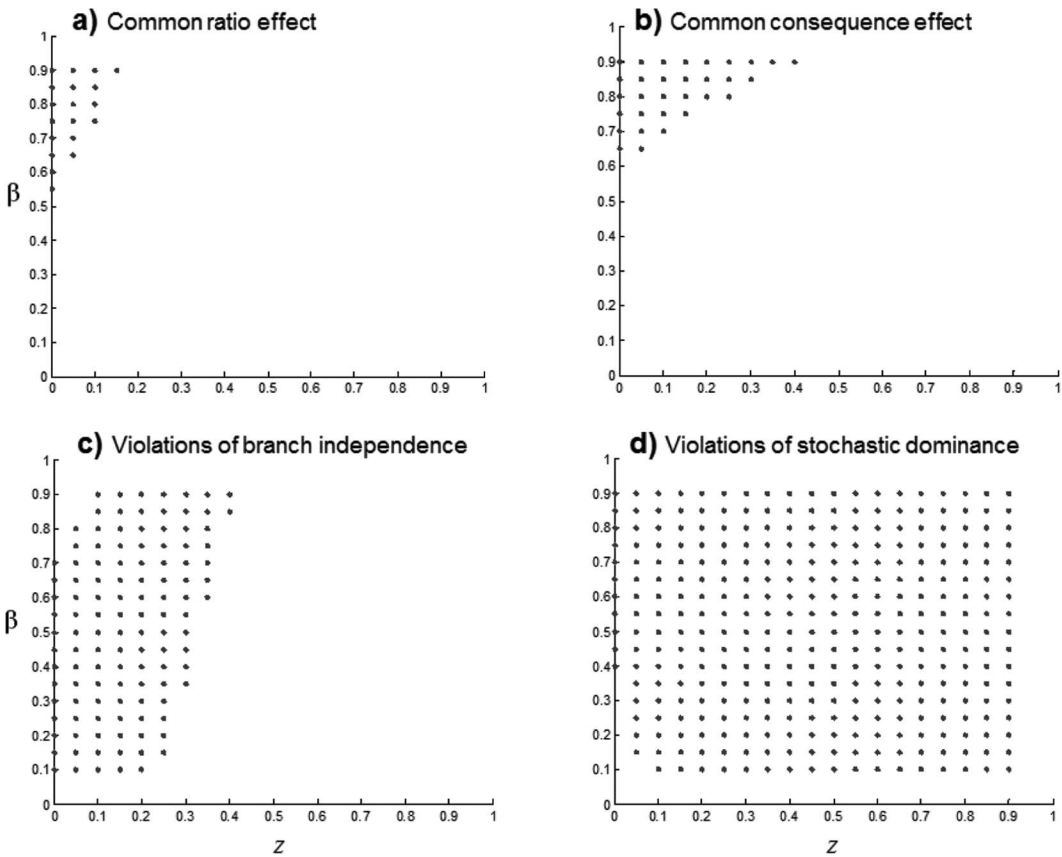


Figure 4. Parameter space map producing effects for corresponding stimuli in Table 1. Points indicate the (z, β) parameter combinations, to a precision of 0.05, that produce (a) the common ratio effect, (b) the common consequence effect, (c) violations of branch independence, and (d) violations of stochastic dominance.

comes affected choices, and could be modeled by increased weight given to those outcomes. This agrees with our conceptualization of decision weight as a product of shifting attention, and could be accounted for by assuming an increase in initial attention to salient outcomes. That is, increased salience of a specific outcome would suggest an increased probability of initial attention to the outcome, which would in turn produce increased weight to the outcome (see also Shah & Oppenheimer, 2007; and Armel, Beaumel, & Rangel, 2008, for related results that can be similarly explained). To support this explanation, work using process-tracing paradigms such as eye-tracking indeed show increased initial attention to salient features (see Orquin & Mueller Loose, 2013, for a review). Future work could potentially incorporate process-tracing in tasks such as that of Weber and Kirsner (1997) to directly test the specific effect regarding probability weighting. In fact, evidence indicates that there is a strong link between frequency and duration of information views (i.e., attention) and the decision weight given to the information (Wedell & Senter, 1997).

Emotional or “affect-rich” outcomes. The process model can also provide psychologically plausible interpretations for phenomena that would seem to contradict algebraic accounts. For example, Rottenstreich and Hsee (2001) discovered that when choices involve emotionally laden outcomes rather than risky monetary gambles the probability weighting function seems to be flatter across the middle range of probabilities, or more step-like. At best, the algebraic approach would be silent on this issue. More likely, the standard algebraic account would specify decreased discriminability or sensitivity for emotional outcomes, because this is the only way to produce the qualitative result (Gonzalez & Wu, 1999; Tversky & Kahneman, 1992; Tversky & Wakker, 1995). However, this seems counterintuitive because people should be more sensitive to emotional outcomes than to those that are not affect-rich. By focusing instead on the attention process, our approach offers an alternative explanation that is more congruent. By simply assuming that one dwells more on emotional (compared to nonemotional) outcomes—by increasing the dwell parameter β —the model predicts exactly the decreased slope for intermediate probabili-

ties found by Rottenstreich and Hsee (2001), as can be seen by comparing across lines within a given panel in Figure 2. More recently, Pachur, Hertwig, and Wolkewitz (2014) found some additional, albeit mixed, evidence for differences in shape for affect-rich situations. Furthermore, they used a paradigm that could potentially determine whether our process-based explanation is valid. Although they did not report detailed statistics for the process-tracing data, including the critical data on dwell times that would support our possible explanation, this remains an exciting avenue for future tests of the model.

Effects of motivational focus. The previous example showed how changes in the dwell probability, independent of any changes in (i.e., given a constant) initial probability distribution, could produce reported effects for emotional outcomes. Alternatively, changes in only the initial probability distribution (i.e., with a constant dwell probability) can also be used to explain other results. For example, Kluger, Stephan, Ganzach, and Hershkovitz (2004) elicited participants’ decision weights for many probability values from various nongamble stimulus domains, such as the perceived probability of a change in social order or victimization in a terrorist attack, given the objective probability of such events. Kluger et al. (2004) found that domains they posited to reflect a “promotion” motivational focus showed more elevated weighting functions, relative to neutral domains and those which engage a “prevention” motivational focus (such as the social order and victimization examples). This increase in elevation can be produced in the process model by assuming an increase in the initial attention to larger outcomes, which agrees conceptually with the notion of a promotion focus (Higgins, 1998). In a similar manner, the process model can use differential initial attention to larger versus smaller outcomes to explain differences between “security-minded,” “potential-minded,” and “cautiously-hopeful” individuals described by Lopes (1995).

Quantitative Comparisons

The previous sections focused on qualitative predictions of the attention process model. In this section we summarize two quantitative tests of the model. First, we compare the predictions

about aggregate choice proportions generated by the attention process model with predictions generated by perhaps the most popular algebraic model, cumulative prospect theory (CPT; [Tversky & Kahneman, 1992](#)). Second, we compare the model predictions to the empirically determined probability weighting functions derived from 10 individual participants (see [Appendix C](#) for details of both procedures). In this manner, we are able to examine both the aggregate- and individual-level performance of the model.

For the first quantitative test, we used data from a well-known study concerning decision weights conducted by [Wu and Gonzalez \(1996\)](#). Their 105 participants each made 40 pairwise choices in an 8(Probability) \times 5(Expected Value) factorial design, manipulating the probability of winning an outcome common across the two gambles, as well as the expected win from each gamble (held relatively equal across each pair). The data thus consisted of choice probabilities (across participants) for each of 40 binary choice conditions between three-outcome gambles. This study was designed to carefully investigate the curvature of the decision weighting function within a CPT framework using the weighting function in [Equation 6](#); see [Appendix C](#) for details.

A generalization criterion ([Busemeyer & Wang, 2000](#)) was used to compare the CPT as applied in the original source with our new process model. This procedure involves fitting model parameters to data from a subset of calibration conditions, and then using these same parameters to make a priori predictions for data from a subset of generalization test conditions. Using this method, the predictions for the generalization test phase do not involve fitting any free parameters, but rather models are tested on their ability to generalize to new conditions.

We first fit both the attention process model and the CPT model presented in [Wu and Gonzalez \(1996\)](#) to the four lowest common outcome probabilities (20 conditions), and then we used the resulting parameter estimates to predict choices among the remaining 20 conditions with relatively high common outcome probabilities. In this generalization test, the attention process model (SSE = .117, $R^2 = 0.58$) outperformed the CPT model (SSE = .135, $R^2 = .51$).⁷ Next, we reversed the procedure: we fit the models to choices involving the higher set of probability conditions, and then predicted choices for the lower set of

conditions. Once again, the attention process model (SSE = .068, $R^2 = 0.56$) outperformed the CPT model (SSE = .128, $R^2 = 0.16$).⁸ Again, it is important to recognize that these are a priori quantitative predictions for the relative frequencies obtained from 20 *new* conditions using *no* free parameters. These results provide empirical evidence that the attention process model not only explains the qualitative findings in the previous section better than the RDU model, but also makes better quantitative predictions than a standard RDU model, cumulative prospect theory.

In applying the process model of weighting to a second study, the data reported in [Gonzalez and Wu \(1999\)](#), there are two key differences from the previous procedure. First, for this application the experimental design provides enough data to perform model fits for each individual, rather than fitting summary measures (mean estimated weights across individuals). This is indeed more in line with the level of analysis of our process model, which specifies attentional processes that occur within the individual.

Second, [Gonzalez and Wu \(1999\)](#) used a pricing (certainty equivalent) task as opposed to choices. We have developed a computational model for pricing data ([Johnson & Busemeyer, 2005](#)), and the pricing model used by [Gonzalez and Wu \(1999\)](#) is a special case of our model (produced by allowing one of our parameters to approach zero in a limit). Our computational model is essential if one wishes to account for preference reversals between choices and prices (see [Johnson & Busemeyer, 2005](#)). However, choices were not included in this second study, and so preference reversals are not the focus of the current application. Therefore, we will

⁷ We examined the use of a maximum likelihood procedure as well, which produced very similar results. We report SSE in following [Wu and Gonzalez \(1996\)](#) and for ease of interpretability.

⁸ The performance of the models in fitting the calibration data suggests the same pattern of results. For the set of lower probability conditions, the process model obtained $R^2 = 0.56$ and the RDU model obtained $R^2 = 0.24$; for the higher probability conditions, the process model obtained $R^2 = 0.58$ and the RDU model obtained $R^2 = 0.57$. The lower probability conditions seem harder to fit, in general. Also, note that the performance of the process model changed little from the calibration set to the generalization set, whereas the RDU model performance decreased substantially in the second application, a possible sign of overfitting by the latter but not the former ([Roberts & Pashler, 2000](#)).

use the limiting case of our model which reduces to the same pricing model used by Gonzalez and Wu (1999) to make the predictions more comparable.⁹

Gonzalez and Wu (1999) elicited certainty equivalents from 10 participants for each of 165 two-outcome gambles (15 outcome pairs crossed with 11 levels of probability). To fit the data, we followed the same general procedure of Gonzalez and Wu (1999). Once again, we fit a total of three parameters—one utility parameter and two decision weight parameters—to each individual participant as well as the median data. Results for all participants are reported in Table 2. The median data were best fit with $z = 0.44$ (which in this binary case corresponds to z_1 , the probability of first considering the lower outcome) and $\beta = 0.93$, with an inflection point near 0.40. For the median data, the best-fitting process model produced $R^2 = 0.94$, with a mean R^2 of 0.90 across all 10 participants. Note that the model was able to account for qualitatively different weighting trends by the individual participants, such as “typical” inverse-S weighting (Participant 3), nearly linear weighting (Participant 6), and nearly step weighting (Participant 8).

Discussion

During the past 60 years, decision scientists working within an algebraic utility framework have assumed a weighted combination of proba-

bility and value information drive human risky decision making. The subjective evaluation of value has been well-captured by a utility function that includes psychological notions such as diminishing marginal sensitivity and reference dependence. The evaluation of probability information that weights this utility has also been the subject of much research. However, in this vein, the ideas started out simple but have become increasingly complex and—we would argue—further removed from the underlying cognitive psychology. Successfully describing the catalog of empirical results from the past 60 years has evolved from assuming that the weight simply equals the outcome probability, to the assumption that weight represents a difference between two nonlinearly transformed rank-dependent decumulative probabilities (e.g., Tversky & Kahneman, 1992; Birnbaum, 2008). In short, the theory of decision weights has become very complicated, which makes one wonder: where do these complicated formulas come from?

We have tried to answer this question using a cognitive modeling approach. The basic idea is that a decision weight corresponds to the proportion of times (or the probability) that an outcome is predicted to occur when a gamble is played. An attention process model, illustrated in Figure 1, is then used to describe the process used to generate these predictions. The resulting process model produces decision weights that reproduce properties postulated by decision theorists.

The attention process model also has several distinct advantages. It does a better job of explaining the empirical qualitative findings than most of the previous decision weight theories (specifically, theories that must satisfy stochastic dominance), and it also makes more accurate quantitative predictions than the weights commonly used in cumulative prospect theory. It can produce weights that approximate those obtained from Birnbaum’s TAX model, which is one of the most successful weighted expected utility models to date. More importantly, it makes new predictions for manipulations of attention (holding outcomes and probabilities

Table 2
Parameter Estimates and Fit Statistics for Individual Participants and Median Data From Gonzalez and Wu (1999)

Participant	z	β	α	SSE	R^2
1	.30	.97	.68	98891	.88
2	.52	.87	.35	129560	.91
3	.43	.92	1.00	116550	.94
4	.17	1.00	.59	24388	.93
5	.84	.98	.14	238060	.76
6	.63	.63	.67	15817	.99
7	.30	1.00	.57	59718	.86
8	.31	.99	.41	88122	.84
9	.36	.61	.58	48739	.97
10	.41	.88	.51	150450	.88
Median	.44	.93	.48	58309	.94

Note. SSE = sum of squared residuals between reported data and model predictions.

⁹ Using the unrestricted pricing model of Johnson and Busemeyer (2005), with any additional parameters set to values reported therein, the fits increased marginally. For the median data: SSE = 22,652; $R^2 = 0.98$. The resulting attention process model parameters were $z = 0.11$; $\beta = 0.38$; and $\alpha = .15$.

fixed) that are supported by experiments yet cannot be explained by utility theories. For example, the attention process model explains how decision weights may change depending on “advertising” or “attention grabbing” manipulations, even when the outcome probabilities remain the same (e.g., Weber & Kirsner, 1997). Finally, the basic idea that a decision weight reflects the percentage of times one considers an outcome is supported by process tracing research that finds “looking time” to be highly correlated with estimated decisions weights (Wedell & Senter, 1997).

Alternative Versions of the Process Model

Beyond changing the free parameters (z , β) of the model, it is possible to conceive of other procedural changes as well. For example, perhaps transitions can occur not only to neighboring outcomes, but to any other outcomes in a gamble. That is, at any given moment, perhaps one may consider any outcome rather than just the outcomes adjacent (in rank order) to the outcome considered in the preceding moment. This would require reapportioning transition probabilities among viable states for event E3, such as follows:

$$\pi_{ij} = \frac{(1 - \pi_{i*})(1 - \beta)}{n - 1}, \text{ for all } j \neq i$$

However, this process of jumping to any outcome fails to reproduce the findings reviewed above. The random walk (transitions only to adjacent states) used in Figure 1 produces a slow drift across the outcome scale, which is needed to reproduce the findings summarized here. One can imagine other possible modifications as well, such as alternative formulations for the initial probability distribution. In our exploration of model variants, we considered uniform probabilities to begin with any outcome, as well as an initial probability distribution set equal to the objective outcome probabilities (see Figure 2). Considering these alternative procedural formulations and/or parameter settings, the proposed attention process is not so much a specific model as it is a family of related models, each making different assumptions about the exact details of the process.

Even though only the model as shown in Figure 1 has been found to be completely successful in accounting for all of the important

findings covered here, it is not necessarily the case that a single model version need account for all findings simultaneously. The applications we have used here were collected from decades of research, and we are not aware of a single study that simultaneously produces all of these effects. Thus, just as the parameter values used between our qualitative and quantitative sections differed, the exact parameter values and model formulation that might best be suited to any specific empirically obtained result remains an open question. For example, Fiedler and Glöckner (2012) report increased attention to gamble outcomes as their value and probability increase. These effects could be captured in our model by a biased start toward higher outcomes, and the relation between objective probability and exit probability inherent in the model, respectively. Thus, although we are confident and supportive of the modeling framework proposed herein, we do not make any strong claims about the generality of the specific parameters used in our “proof of concept” here to produce the behavioral trends.

Predictions for Independent Versus Dependent Gambles

The attention process model can also be formulated to operate on gambles when the outcomes are dependent, as opposed to the assumed independence used in the reported applications (and almost exclusively in the existing literature). For example, assume that there are 99 colored marbles in each of two urns representing the outcomes of the gambles below:

F: (\$10, 1/3; \$24, 1/3; \$45, 1/3)

G: (\$10, 1/3; \$12, 1/3; \$96, 1/3),

If a ball is drawn from each urn, for each gamble, any outcome from F can occur with any outcome of G, so that if both gambles were realized, \$24 could occur with F and \$96 could occur with G, for example. In this case, it is assumed that an *independent* prediction is made for each gamble as introduced in our model originally.

If the outcomes of each gamble are determined jointly, such as by a single draw of a colored marble from a single urn, then they will be *dependent*. In this case, perhaps there are equal numbers of red, blue, and green marbles in the urn that jointly determine the lowest, middle, and highest outcomes, respectively. Then, if a red marble is

drawn, it produces \$10 for F and \$10 for G, and a blue marble produces \$24 for F and \$12 for G, and a green marble produces \$45 for F and \$96 for G. In particular, compared with the case of independent draws, a \$24 outcome from F could not occur with \$96 from G, and instead \$24 from F could only occur with \$12 from G.

If some common event, such as drawing a green marble, produces outcomes for both gambles, then the attention model would mentally simulate these events, rather than predicting independent outcomes for each gamble. In other words, attention would shift among the different colored marbles, producing a prediction for a particular color (such as green), which would then determine the prediction simultaneously for both gambles. It is worth noting that, except for regret-based theories (Loomes et al., 1992), none of the algebraic utility-based decision weight models introduced in the current work make any distinction between independent and dependent gambles (Andraszewicz, Rieskamp, & Scheibehenne, 2015).

To illustrate the model in this situation, consider the simplest case where a single fair coin flip will determine the payoffs for two different binary outcome gambles. In this case, the outcomes depend on a common event, that is, the coin flip. The notation $(\$x, H; \$y, T)$ represents a payoff of $\$x$ should the coin land on heads, and $\$y$ should tails occur, where $\Pr[\text{Head}] = \Pr[\text{Tail}] = 0.5$. Consider a choice between the pair:

F: $(\$100, H; -\$100, T)$,

G: $(\$99, H; -\$101, T)$.

Obviously, F stochastically dominates G, and people almost always choose F (see Diederich & Busemeyer, 1999, for related work with different stimuli). Next consider a choice between the pair:

F: $(\$100, H; -\$100, T)$;

G*: $(-\$101, H; \$99, T)$.

Note that in terms of probabilities and outcomes, G* is identical to G and so F still stochastically dominates G*. However, people are almost indifferent between these two options, and the probability of choosing F drops down close to .50 in some cases (for the more complicated stimuli used by Diederich & Busemeyer, 1999). This juxtaposition of equally likely events with payoffs has a major effect on choices. Again, none of the traditional decision theories reviewed earlier pre-

dict this juxtaposition effect, because in these theories the utility of a gamble is based solely on the outcomes and probabilities of the gamble, and so the choice between F and G is identical to the choice between F and G*.

According to the attention process model, attention is now focused on the possible outcomes of the coin flip. On the one hand, when faced with a choice between F and G, the only two possible predicted outcomes are \$100 for F and \$99 for G (if attention is focused on the event heads); or $-\$100$ for F and $-\$101$ for G (if attention is focused on the event tails). Gamble F is clearly better in both circumstances, there is no conflict at all, and so F is always chosen. The decision maker modeled by our process model would never predict a pair of outcomes such as \$100 for F and $-\$101$ for G, because this combination of outcomes is impossible for this pair of actions. On the other hand, when faced with a choice between F and G*, the only two possible predicted outcomes are \$100 for F and $-\$101$ for G* (if attention is focused on the event heads); or $-\$99$ for F and \$100 for G* (if attention is focused on the event tails). The choice is now very unclear because what one should do depends strongly on whether a head or tail occurs. This introduces a great deal of conflict, and makes preferences more uncertain and choices more random. In short, if the gamble outcomes are yoked to specific events (i.e., heads or tails) then attention does not shift independently among outcomes, but instead these predictions are coordinated by the events. Indeed, assuming $\beta = 0.70$ and linear utility, $u(x) = x$, as in the other qualitative applications, equal probability of attending first to either heads or tails ($z = 0.5$), and using decision field theory to predict choices (in order to produce probabilistic choice; see Appendix A), the process model predicts $\Pr[F | \{F, G\}] > 0.99$, and $\Pr[F | \{F, G^*\}] = 0.51$, in line with the empirical result.

To summarize, a reasonable extension of the attention process model explains how decision weights may change depending on the nature of the random device used to generate gamble outcomes, even when the outcome probabilities remain the same—that is, depending on whether outcomes are generated by statistically independent or dependent devices. Previous decision theories assume that the weights depend only on the outcome probabilities (and possibly value ranks) for each gamble, and they are insensitive to the statistical dependence of the outcomes

between gambles. The effect of statistical dependence was empirically demonstrated in the study by Diederich and Busemeyer (1999) as well as a more recent study by Andraszewicz, Rieskamp, and Scheibehenne (2015).

Other Extensions and Future Directions

The attention process model also makes predictions regarding the time it takes to generate each sample, and thus the overall decision time. Specifically, factors that increase the dwell time parameter, β , should increase the time required to generate a prediction, and thus increase the total decision time (see Appendix A for a mathematical analysis). Recall that we explain the effect of emotional outcomes by assuming that the dwell rate parameter is larger for these types of outcomes. If this assumption is correct, then we are forced to predict that decision times should be longer for these emotional outcomes. Because empirical work for decades has been driven by utility theories that are silent on predictions such as information search and response time, there is not yet an abundance of data to critically examine these aspects of computational models of decision making.

The general phenomenon of probability weighting is likely to extend beyond situations such as the admittedly artificial gambling tasks presented in the current work. For example, Ungemach, Chater, and Stewart (2009) study decision weighting in situations where outcome probabilities of gambles are learned from experience rather than directly stated. Glaser, Trommershäuser, Mamassian, and Maloney (2012) extend this type of situation to a visual task (in addition to traditional gambles) and find similar trends of overweighting small probabilities and underweighting large ones via Equation 6. Kusev, van Schaik, Ayton, Dent, and Chater (2009) applied Equation 4 across five experiments comparing traditional gambles with more “real world” decisions such as buying insurance against lost luggage and found differential probability weighting across different problem contexts, even with identical probabilities and values. Future applications of the process model to more ecologically valid situations such as these would be especially beneficial in understanding the mechanisms involved. For example, assuming a greater dwell probability in the current model in the context of insurance decisions could apparently explain the effects re-

ported by Kusev et al. (2009). Finally, it would be worthwhile to compare the current model to other recent accounts of decision weighting, both in terms of their explanatory power as well as in their proposed perceptual and cognitive mechanisms. Bordalo, Gennaioli, and Shleifer (2012) propose a theory of probability weighting that depends on the context-dependent saliency of outcomes, whereby outcomes that are more attractive in contrast to other outcomes garner more weight. Fennell and Baddeley (2012) propose that Bayesian updating can produce decision weighting effects by assuming either knowledgeable (beta) or uninformative priors which are updated with the probability information stated in a decision task. It would be interesting not only to explore the theoretical and conceptual similarities and distinctions between the current model and these alternative perspectives, but also to develop experimental tests that could discriminate among them.

Just as empirical violations of theoretical properties have led to the current incarnations of utility theory (such as TAX, arguably the most successful form), results that contradict these theories should pave the way for embracing new, alternative models. Some researchers prefer to continue adjusting the utility framework as necessary, leading to increasingly opaque and theoretically complex forms. We adopt a different approach through computational modeling of the cognitive processes underlying decision behavior. The number and scope of applications of computational models to decision making are growing at a rapid pace (see Busemeyer & Johnson, 2004; Johnson & Busemeyer, 2010). These models account for results that have challenged traditional theories while adding novel predictions, parsimony, and psychological plausibility.

References

- Allais, M. (1953). Le comportement de l'homme rationnel devant le risque: Critique des postulats et axiomes de l'école américaine [Rational man's behavior in the presence of risk: Critique of the postulates and axioms of the American school]. *Econometrica*, 21, 503–546. <http://dx.doi.org/10.2307/1907921>
- Andraszewicz, S., Rieskamp, J., & Scheibehenne, B. (2015). How outcome dependencies affect decisions under risk. *Decision*, 2, 127–144. <http://dx.doi.org/10.1037/dec0000028>

- Ariely, D., Loewenstein, G., & Prelec, D. (2003). Coherent arbitrariness: Stable demand curves without stable preferences. *The Quarterly Journal of Economics*, *118*, 73–106.
- Armel, K. C., Beaumel, A., & Rangel, A. (2008). Biasing simple choices by manipulating relative visual attention. *Judgment and Decision Making*, *3*, 396–403.
- Berkowitsch, N. A., Scheibehenne, B., & Rieskamp, J. (2014). Rigorously testing multialternative decision field theory against random utility models. *Journal of Experimental Psychology: General*, *143*, 1331–1348. <http://dx.doi.org/10.1037/a0035159>
- Bhatia, S. (2013). Associations and the accumulation of preference. *Psychological Review*, *120*, 522–543. <http://dx.doi.org/10.1037/a0032457>
- Birnbaum, M. H. (2005). A comparison of five models that predict violations of first-order stochastic dominance in risky decision making. *Journal of Risk and Uncertainty*, *31*, 263–287. <http://dx.doi.org/10.1007/s11166-005-5103-9>
- Birnbaum, M. H. (2008). New paradoxes of risky decision making. *Psychological Review*, *115*, 463–501. <http://dx.doi.org/10.1037/0033-295X.115.2.463>
- Birnbaum, M. H., & Chavez, A. (1997). Tests of theories of decision making: Violations of branch independence and distribution independence. *Organizational Behavior and Human Decision Processes*, *71*, 161–194. <http://dx.doi.org/10.1006/obhd.1997.2721>
- Birnbaum, M. H., & McIntosh, W. R. (1996). Violations of branch independence in choices between gambles. *Organizational Behavior and Human Decision Processes*, *67*, 91–110. <http://dx.doi.org/10.1006/obhd.1996.0067>
- Birnbaum, M. H., & Navarrete, J. B. (1998). Testing descriptive utility theories: Violations of stochastic dominance and cumulative independence. *Journal of Risk and Uncertainty*, *17*, 49–79. <http://dx.doi.org/10.1023/A:1007739200913>
- Birnbaum, M. H., Patton, J. N., & Lott, M. K. (1999). Evidence against rank-dependent utility theories: Tests of cumulative independence, interval independence, stochastic dominance, and transitivity. *Organizational Behavior and Human Decision Processes*, *77*, 44–83. <http://dx.doi.org/10.1006/obhd.1998.2816>
- Birnbaum, M. H., & Veira, R. (1998). Configural weighting in judgments of two- and four-outcome gambles. *Journal of Experimental Psychology: Human Perception and Performance*, *24*, 216–226. <http://dx.doi.org/10.1037/0096-1523.24.1.216>
- Booij, A. S., Van Praag, B. M., & Van De Kuilen, G. (2010). A parametric analysis of prospect theory's functionals for the general population. *Theory and Decision*, *68*, 115–148. <http://dx.doi.org/10.1007/s11238-009-9144-4>
- Bordalo, P., Gennaioli, N., & Shleifer, A. (2012). Salience theory of choice under risk. *The Quarterly Journal of Economics*, *127*, 1243–1285. <http://dx.doi.org/10.1093/qje/qjs018>
- Brandstätter, E., Gigerenzer, G., & Hertwig, R. (2006). The priority heuristic: Making choices without trade-offs. *Psychological Review*, *113*, 409–432. <http://dx.doi.org/10.1037/0033-295X.113.2.409>
- Busemeyer, J. R., Jessup, R. K., Johnson, J. G., & Townsend, J. T. (2006). Building bridges between neural models and complex decision making behaviour. *Neural Networks*, *19*, 1047–1058. <http://dx.doi.org/10.1016/j.neunet.2006.05.043>
- Busemeyer, J. R., & Johnson, J. G. (2004). Computational models of decision making. In D. Koehler & N. Harvey (Eds.), *Blackwell handbook of judgment and decision making* (pp. 133–154). Oxford, UK: Blackwell Publishing Co. <http://dx.doi.org/10.1002/9780470752937.ch7>
- Busemeyer, J. R., & Johnson, J. G. (2008). Micro-process models of decision making. *Cambridge handbook of computational psychology* (pp. 302–321). Cambridge, UK: Cambridge University Press.
- Busemeyer, J. R., & Townsend, J. T. (1992). Fundamental derivations for decision field theory. *Mathematical Social Sciences*, *23*, 255–282. [http://dx.doi.org/10.1016/0165-4896\(92\)90043-5](http://dx.doi.org/10.1016/0165-4896(92)90043-5)
- 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. <http://dx.doi.org/10.1037/0033-295X.100.3.432>
- Busemeyer, J. R., & Wang, Y. M. (2000). Model comparisons and model selections based on generalization criterion methodology. *Journal of Mathematical Psychology*, *44*, 171–189. <http://dx.doi.org/10.1006/jmps.1999.1282>
- Camerer, C. F., & Ho, T. H. (1994). Violation of the betweenness axiom and nonlinearity in probability. *Journal of Risk and Uncertainty*, *8*, 167–196. <http://dx.doi.org/10.1007/BF01065371>
- Diederich, A. (1997). Dynamic stochastic models for decision making under time constraints. *Journal of Mathematical Psychology*, *41*, 260–274. <http://dx.doi.org/10.1006/jmps.1997.1167>
- Diederich, A. (2003). MDFT account of decision making under time pressure. *Psychonomic Bulletin & Review*, *10*, 157–166. <http://dx.doi.org/10.3758/BF03196480>
- Diederich, A., & Busemeyer, J. R. (1999). Conflict and the stochastic dominance principle of decision making. *Psychological Science*, *10*, 353–359. <http://dx.doi.org/10.1111/1467-9280.00167>
- Diederich, A., & Busemeyer, J. R. (2003). Simple matrix methods for analyzing diffusion models of choice probability, choice response time, and simple response time. *Journal of Mathematical Psy-*

- chology, 47, 304–322. [http://dx.doi.org/10.1016/S0022-2496\(03\)00003-8](http://dx.doi.org/10.1016/S0022-2496(03)00003-8)
- Dror, I. E., Basola, B., & Busemeyer, J. R. (1999). Decision making under time pressure: An independent test of sequential sampling models. *Memory & Cognition*, 27, 713–725. <http://dx.doi.org/10.3758/BF03211564>
- Edwards, W. (1955). The prediction of decisions among bets. *Journal of Experimental Psychology*, 50, 201–214. <http://dx.doi.org/10.1037/h0041692>
- Edwards, W. (1962). Subjective probabilities inferred from decisions. *Psychological Review*, 69, 109–135. <http://dx.doi.org/10.1037/h0038674>
- Estes, W. K. (1959). Component and pattern models with Markovian interpretations. In R. R. Bush and W. K. Estes, *Studies in mathematical learning theory* (pp. 9–53). Stanford, CA: Stanford University Press.
- Estes, W. K. (1972). An associative basis for coding and organization in memory. In A. W. Melton & E. Martin (Eds.) *Coding processes in human memory* (pp. 161–190). Washington, DC: Winston.
- Fennell, J., & Baddeley, R. (2012). Uncertainty plus prior equals rational bias: An intuitive Bayesian probability weighting function. *Psychological Review*, 119, 878–887. <http://dx.doi.org/10.1037/a0029346>
- Fiedler, S., & Glöckner, A. (2012). The dynamics of decision making in risky choice: An eye-tracking analysis. *Frontiers in Psychology*, 3, 335. <http://dx.doi.org/10.3389/fpsyg.2012.00335>
- Glaser, C., Trommershäuser, J., Mamassian, P., & Maloney, L. T. (2012). Comparison of the distortion of probability information in decision under risk and an equivalent visual task. *Psychological Science*, 23, 419–426. <http://dx.doi.org/10.1177/0956797611429798>
- Goldstein, W. M., & Weber, E. U. (1995). Content and discontent: Indications and implications of domain specificity in preferential decision making. *Psychology of Learning and Motivation*, 32, 83–136. [http://dx.doi.org/10.1016/S0079-7421\(08\)60308-4](http://dx.doi.org/10.1016/S0079-7421(08)60308-4)
- Gonzalez, R., & Wu, G. (1999). On the shape of the probability weighting function. *Cognitive Psychology*, 38, 129–166. <http://dx.doi.org/10.1006/cogp.1998.0710>
- Higgins, E. T. (1998). Promotion and prevention: Regulatory focus as a motivational principle. In M. P. Zanna (Ed.), *Advances in experimental social psychology* (Vol. 30, pp. 1–46). London, UK: Elsevier Academic Press.
- Hotaling, J. M., & Busemeyer, J. R. (2012). DFT-D: A cognitive-dynamical model of dynamic decision making. *Synthese*, 189, 67–80. <http://dx.doi.org/10.1007/s11229-012-0157-0>
- Jessup, R. K., Bishara, A. J., & Busemeyer, J. R. (2008). Feedback produces divergence from prospect theory in descriptive choice. *Psychological Science*, 19, 1015–1022. <http://dx.doi.org/10.1111/j.1467-9280.2008.02193.x>
- Johnson, J. G., & Busemeyer, J. R. (2005). A dynamic, stochastic, computational model of preference reversal phenomena. *Psychological Review*, 112, 841–861. <http://dx.doi.org/10.1037/0033-295X.112.4.841>
- Johnson, J. G., & Busemeyer, J. R. (2010). Decision making under risk and uncertainty. *Wiley Interdisciplinary Reviews: Cognitive Science*, 1, 736–749. <http://dx.doi.org/10.1002/wcs.76>
- Kahneman, D., & Tversky, A. (1979). Prospect theory: An analysis of decision under risk. *Econometrica*, 47, 263–291. <http://dx.doi.org/10.2307/1914185>
- Kahneman, D., & Tversky, A. (1982). The simulation heuristic. In D. Kahneman, P. Slovic, & A. Tversky (Eds.), *Judgment under uncertainty: Heuristics and biases* (pp. 201–208). New York, NY: Cambridge University Press. <http://dx.doi.org/10.1017/CBO9780511809477.015>
- Karlin, S., & Taylor, H. M. (1975). *A first course in stochastic processes* (Vol. I). New York, NY: Academic Press.
- Kluger, A. N., Stephan, E., Ganzach, Y., & HersHKovitz, M. (2004). The effect of regulatory focus on the shape of probability-weighting function: Evidence from a cross-modality matching method. *Organizational Behavior and Human Decision Processes*, 95, 20–39. <http://dx.doi.org/10.1016/j.obhdp.2004.05.003>
- Kusev, P., van Schaik, P., Ayton, P., Dent, J., & Chater, N. (2009). Exaggerated risk: Prospect theory and probability weighting in risky choice. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 35, 1487–1505. <http://dx.doi.org/10.1037/a0017039>
- Lattimore, P. K., Baker, J. R., & Witte, A. D. (1992). The influence of probability on risky choice: A parametric examination. *Journal of Economic Behavior & Organization*, 17, 377–400. [http://dx.doi.org/10.1016/S0167-2681\(95\)90015-2](http://dx.doi.org/10.1016/S0167-2681(95)90015-2)
- Lee, M. D., & Cummins, T. D. (2004). Evidence accumulation in decision making: Unifying the “take the best” and the “rational” models. *Psychonomic Bulletin & Review*, 11, 343–352. <http://dx.doi.org/10.3758/BF03196581>
- Loomes, G., Starmer, C., & Sugden, R. (1992). Are preferences monotonic? Testing some predictions of regret theory. *Econometrica*, 59, 17–33.
- Lopes, L. L. (1995). Algebra and process in the modeling of risky choice. In J. R. Busemeyer, R. Hastie, & D. Medin (Eds.), *Decision making from a cognitive perspective* (pp. 177–220). New York, NY: Academic Press. [http://dx.doi.org/10.1016/S0079-7421\(08\)60310-2](http://dx.doi.org/10.1016/S0079-7421(08)60310-2)
- Luce, R. D. (2000). *Utility of gains and losses*. Mahwah, NJ: Erlbaum.

- Orquin, J. L., & Mueller Loose, S. (2013). Attention and choice: A review on eye movements in decision making. *Acta Psychologica*, *144*, 190–206. <http://dx.doi.org/10.1016/j.actpsy.2013.06.003>
- Pachur, T., Hertwig, R., & Wolkewitz, R. (2014). The affect gap in risky choice: Affect-rich outcomes attenuate attention to probability information. *Decision*, *1*, 64–78. <http://dx.doi.org/10.1037/dec0000006>
- Prelec, D. (1998). The probability weighting function. *Econometrica*, *66*, 497–527. <http://dx.doi.org/10.2307/2998573>
- Preston, M. G., & Baratta, P. (1948). An experimental study of the auction-value of an uncertain outcome. *The American Journal of Psychology*, *61*, 183–193. <http://dx.doi.org/10.2307/1416964>
- Quiggin, J. (1992). *Generalized expected utility theory: The rank dependent model*. New York, NY: Springer.
- Rieskamp, J. (2008). The probabilistic nature of preferential choice. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *34*, 1446–1465. <http://dx.doi.org/10.1037/a0013646>
- Roberts, S., & Pashler, H. (2000). How persuasive is a good fit? A comment on theory testing. *Psychological Review*, *107*, 358–367. <http://dx.doi.org/10.1037/0033-295X.107.2.358>
- Roe, R. M., Busemeyer, J. R., & Townsend, J. T. (2001). Multialternative decision field theory: A dynamic connectionist model of decision making. *Psychological Review*, *108*, 370–392. <http://dx.doi.org/10.1037/0033-295X.108.2.370>
- Rottenstreich, Y., & Hsee, C. K. (2001). Money, kisses, and electric shocks: On the affective psychology of risk. *Psychological Science*, *12*, 185–190. <http://dx.doi.org/10.1111/1467-9280.00334>
- Scheibehenne, B., Rieskamp, J., & González-Vallejo, C. (2009). Cognitive models of choice: Comparing decision field theory to the proportional difference model. *Cognitive Science*, *33*, 911–939. <http://dx.doi.org/10.1111/j.1551-6709.2009.01034.x>
- Shah, A. K., & Oppenheimer, D. M. (2007). Easy does it: The role of fluency in cue weighting. *Judgment and Decision Making*, *2*, 371–379.
- Sneddon, R., & Luce, R. D. (2001). Empirical comparisons of bilinear and nonbilinear utility theories. *Organizational Behavior and Human Decision Processes*, *84*, 71–94. <http://dx.doi.org/10.1006/obhd.2000.2919>
- Stewart, N., Chater, N., & Brown, G. D. (2006). Decision by sampling. *Cognitive Psychology*, *53*, 1–26. <http://dx.doi.org/10.1016/j.cogpsych.2005.10.003>
- Stott, H. P. (2006). Cumulative prospect theory's functional menagerie. *Journal of Risk and Uncertainty*, *32*, 101–130. <http://dx.doi.org/10.1007/s11166-006-8289-6>
- Thorngate, W. (1980). Efficient decision heuristics. *Behavioral Science*, *25*, 219–225. <http://dx.doi.org/10.1002/bs.3830250306>
- Townsend, J. T., & Wenger, M. J. (2004). The serial-parallel dilemma: A case study in a linkage of theory and method. *Psychonomic Bulletin & Review*, *11*, 391–418. <http://dx.doi.org/10.3758/BF03196588>
- Trueblood, J. S., Brown, S. D., & Heathcote, A. (2014). The multiattribute linear ballistic accumulator model of context effects in multialternative choice. *Psychological Review*, *121*, 179–205. <http://dx.doi.org/10.1037/a0036137>
- Tversky, A. (1967). Utility theory and additivity analysis of risky choices. *Journal of Experimental Psychology*, *75*, 27–36. <http://dx.doi.org/10.1037/h0024915>
- Tversky, A., & Fox, C. R. (1995). Weighing risk and uncertainty. *Psychological Review*, *102*, 269–283. <http://dx.doi.org/10.1037/0033-295X.102.2.269>
- Tversky, A., & Kahneman, D. (1992). Advances in prospect theory: Cumulative representations of uncertainty. *Journal of Risk and Uncertainty*, *5*, 297–323. <http://dx.doi.org/10.1007/BF00122574>
- Tversky, A., & Wakker, P. (1995). Risk attitudes and decision weights. *Econometrica*, *63*, 1255–1280. <http://dx.doi.org/10.2307/2171769>
- Ungemach, C., Chater, N., & Stewart, N. (2009). Are probabilities overweighted or underweighted when rare outcomes are experienced (rarely)? *Psychological Science*, *20*, 473–479. <http://dx.doi.org/10.1111/j.1467-9280.2009.02319.x>
- Usher, M., & McClelland, J. L. (2004). Loss aversion and inhibition in dynamical models of multialternative choice. *Psychological Review*, *111*, 757–769. <http://dx.doi.org/10.1037/0033-295X.111.3.757>
- Weber, E. U., & Kirsner, B. (1997). Reasons for rank-dependent utility evaluation. *Journal of Risk and Uncertainty*, *14*, 41–61. <http://dx.doi.org/10.1023/A:1007769703493>
- Wedell, D. H., & Senter, S. M. (1997). Looking and weighting in judgment and choice. *Organizational Behavior and Human Decision Processes*, *70*, 41–64. <http://dx.doi.org/10.1006/obhd.1997.2692>
- Wu, G., & Gonzalez, R. (1996). Curvature of the probability weighting function. *Management Science*, *42*, 1676–1690. <http://dx.doi.org/10.1287/mnsc.42.12.1676>

(Appendices follow)

Appendix A

Integrating the Attention Process Model into Decision Field Theory

Consider a choice between two gambles, labeled *A* and *B*. The preferences for *A* and *B* at time *t* are denoted $P_A(t)$ and $P_B(t)$, respectively. In a simple version of DFT, the preference for an option evolves according to the following random walk equation: $P_i(t + 1) = P_i(t) + V_i(t + 1)$, where the $V_i(t)$ is called the valence of option *i* at time *t*. The valence for option *i* is defined as $V_i(t) = [U_i(t) - U_j(t)]$, where $U_i(t)$ and $U_j(t)$ represent the affective evaluations of the outcomes that are anticipated for options *i* and *j* at time step *t*.

For example, consider the choice between the following pair given in the Introduction:

A: (\$12, .10; \$90, .05; \$96, .85),

B: (\$12, .05; \$14, .05; \$96, .90).

At each time step, a valence is generated by considering a possible outcome from A, and also considering a possible outcome from B, and then comparing the difference in affective values between these two anticipated outcomes. What is the probability of anticipating a particular outcome from each gamble to determine the valence at each time step?

Here is where the attention process model makes a contribution. We now use Equation 1 to determine the probability of considering an outcome from a gamble at each time step. For example, considering gamble A, and using the parameters $z = 0$ and $\beta = .70$, then according to Equation 2 or Equation 3, the probabilities of considering the outcomes \$10, \$90, \$96 equal (.40, .16, .44) at each time step.

Choice Probability Predictions From DFT

To compute the choice probability predictions, we need to derive the mean and variance of the valence (see Busemeyer & Townsend, 1992, for derivations). For a binary choice between two statistically independent gambles A and B, these are given by:

$$E[V(t)] = \mu = E[U_A(t)] - E[U_B(t)], \quad (A1)$$

$$\text{Var}[V(t)] = \text{Var}[U_A(t)] + \text{Var}[U_B(t)] \text{ and}$$

$$\sigma = \sqrt{\text{Var}[V(t)]} \quad (A2)$$

The individual means and variances are determined as follows, for $j = \{A, B\}$ and outcomes $k = \{1, 2, \dots, n\}$:

$$E[U_j(t)] = \sum_k w(x_{kj}) \cdot u(x_{kj}), \quad (A3)$$

$$\text{Var}[U_j(t)] = \sum_k w(x_{kj}) \cdot u(x_{kj})^2 - E[U_j(t)]^2 \quad (A4)$$

Here we have inserted the decision weights, $w(x_{kj})$, given by Equation 2 to determine the probabilities of considering each outcome from each gamble. These parameters are used together to compute a discriminability index, $d = \mu/\sigma$, much like the d' index used in signal detection theory. With this formulation, if we specify a threshold level θ indicating the value of accumulated preference necessary to make a choice, and assume that $P(0) = 0$ so the choice process starts unbiased, then the probability of selecting gamble A equals (see, e.g., Busemeyer & Townsend, 1993):

$$\text{Pr}[A | \{A, B\}] = \frac{1}{1 + \exp(-2 \cdot d \cdot \theta)}$$

Dependent Gambles

In the Discussion, we alluded to the possibility of using our process model for situations where gamble outcomes are dependent. The model operates in the same manner, but now the attention is focused jointly on one outcome from each gamble (i.e., outcome pairs). In this case, the variance of the valence needs to take

(Appendices continue)

into account the relationship between the gambles. Specifically, for the case of two paired outcomes as in the coin flip example, we use (where H denotes the first outcome, Heads; T denotes the second outcome, Tails):

$$\text{Var}[V(t)] = w(H)[x_{1H} - x_{2H}]^2 + w(T)[x_{1T} - x_{2T}]^2 - \mu^2 \quad (\text{A5})$$

The weights $w(H)$ and $w(T)$ can easily be found using Equation 3 where the states Heads and Tails are arbitrarily but consistently defined. The rest of the derivation remains consistent with the case for independent gambles, with the computation of $d = \mu/\sigma$ after defining μ via Equations A1 and A3, and the probabilistic choice rule $\text{Pr}[A | \{A, B\}]$.

Decision Time

One advantage of developing a process model for decision weights is that this process also describes the time it takes to make a decision. Of course, the original version of DFT is a process model for choice, and in previous work we have tested choice response time predictions derived from this choice process (Busemeyer & Townsend, 1993; Dror, Basola, & Busemeyer, 1999; Diederich, 2003). However, the addition of the attention process now requires us to consider two distinct conceptual time scales on which our hypothesized choice process operates. Each macro time step of the DFT choice process is now comprised of several micro time steps required to generate predicted outcomes of the gambles. Several moments pass in the attention model, corresponding to transitions among outcomes, before a prediction is made, and this prediction then produces a single step in the DFT choice process. The whole process of generating predictions and updating preferences is repeated until the threshold is reached, and the final time to make a decision is determined by the time this entire process takes.

Consequently, the new addition of the attention process to the previous DFT choice process has new implications for predictions regarding choice time. In previous applications to choice response time, we assumed that each time step of the preference accumulation process was constant. Now each time step is stochastically determined by the attention process. Thus the mean choice time is determined by the mean number of time steps required to reach threshold from the choice model, multiplied by the mean time to generate each step from the attention model. Of course all of these derivations depend on the assumption of serial processing of predictions from the attention process and updates of the preference accumulation process. Methodologies for identifying mental architecture (e.g., see Townsend & Wenger, 2004, for a review) could be used to test these serial processing assumptions. In previous applications, we derived the mean time for the DFT model to reach the choice threshold (Busemeyer & Townsend, 1992; Diederich & Busemeyer, 2003). Below we derive the mean time to generate a prediction from the attention model.

Prediction time for any single option t_i^* is a random variable, and the mean time required to produce each prediction can be found by (Diederich & Busemeyer, 2003; Busemeyer & Townsend, 1992):

$$E[t_i^* | \mathbf{W}] = \mathbf{Z}'(\mathbf{I} - \mathbf{Q})^{-2} \mathbf{R}h ./ \mathbf{W} \quad (\text{A6})$$

Each element in the n -dimensional vector \mathbf{t}_i^* thus gives the conditional mean time required to predict the corresponding outcome for the gamble i . The operator $./$ denotes element division of matrices, and the constant h is a scalar time unit used to translate the number of attention transitions into real time. Without h , Equation A5 produces the mean number of transitions undertaken by the attention model before predicting the corresponding outcome.

(Appendices continue)

Appendix B

Relevant Derivations and Solutions

Suppose a gamble has n outcomes ordered ($x_n > x_{n-1} > \dots > x_{i+1} > x_i > x_1$); then the attention weighting process can be, at any point in time, in one of n states, where the states are ordered by outcome values. The initial probability distribution across these n states is represented by the $1 \times n$ row vector Z , containing the probability z_i that outcome i is the first one considered. In the current applications, the distribution over initial states, z_i contained in Z is defined using a binomial distribution that depends on the parameter z :

$$z_i = \frac{(N-1)!}{(i-1)!(N-i-2)!} \cdot z^{i-1}(1-z)^{N-i-2}$$

Given that one is “looking” at a state, we assume that the exit or prediction probability for each state is equal to the objective probability of obtaining the associated outcome, $\pi_{i^*} = p_i$.

The joint probability of not predicting state i but continuing to dwell on this state is then equal to $\pi_{ii} = (1 - \pi_{i^*}) \cdot \beta$, for all outcomes i . This corresponds to the probability of staying on the same node in Figure 1 for consecutive moments. If an exit has not occurred, and the person does not dwell there, then a transition to another state must occur with some total probability equal to $(1 - \pi_{i^*}) \cdot (1 - \beta)$. For now, we adopt the simple random walk assumption that transitions occur only to neighboring states (adjacent rank-ordered outcomes); then, for n outcomes, we obtain:

$$\begin{aligned} \pi_{12} &= (1 - \pi_{1^*})(1 - \beta) \\ \pi_{n(n-1)} &= (1 - \pi_{n^*})(1 - \beta) \\ \pi_{i(i-1)} = \pi_{i(i+1)} &= \frac{(1 - \pi_{i^*})(1 - \beta)}{2}, \text{ for } i \neq 1, n. \end{aligned} \tag{B1}$$

Derivation of Process Weighting Model Solution for Binary Case (Equation 3)

For the binary case, $n = 2$, then the transition matrix Q in Equation 1 becomes as follows:

$$\begin{aligned} Q_2 &= \begin{bmatrix} (1-p_1)\beta & (1-p_1)(1-\beta) \\ (1-p_2)(1-\beta) & (1-p_2)\beta \end{bmatrix} \\ &= \begin{bmatrix} p_2\beta & p_2(1-\beta) \\ p_1(1-\beta) & p_1\beta \end{bmatrix}. \end{aligned}$$

The latter matrix takes advantage of the relation $p_1 + p_2 = 1$. This transition matrix is used to create the second term in Equation 2 by subtraction from a 2×2 identity matrix and taking the inverse:

$$(I - Q_2)^{-1} = \begin{bmatrix} \frac{1 + (p_2 - 1)\beta}{p_1(1 - \beta) + p_2(1 - \beta) + p_1p_2(2\beta - 1)} & \frac{(p_1 - 1)(\beta - 1)}{p_1(1 - \beta) + p_2(1 - \beta) + p_1p_2(2\beta - 1)} \\ \frac{(p_2 - 1)(\beta - 1)}{p_1(1 - \beta) + p_2(1 - \beta) + p_1p_2(2\beta - 1)} & \frac{1 + (p_1 - 1)\beta}{p_1(1 - \beta) + p_2(1 - \beta) + p_1p_2(2\beta - 1)} \end{bmatrix}$$

We define the initial state vector generally, $Z_2 = [z_1 \ z_2]'$, and assume that the response or absorbing probabilities in R are equal to the objective probabilities (see text):

$$R_2 = \begin{bmatrix} p_1 & 0 \\ 0 & p_2 \end{bmatrix}$$

Then w_2 can be found as the second element of the matrix product in Equation 2:

$$w_2 = \frac{(1 - \beta)z_1p_2^2 + (1 - p_2\beta)z_2p_2}{(1 - \beta)p_1 + (1 - \beta)p_2^2 + \beta p_1p_2}$$

(Appendices continue)

Appendix C

Fitting the Process Model

Procedure for Fitting Process Weighting Model to Algebraic Functions

We used the domain of $p_2 = [0, 0.01, 0.02, \dots, 0.98, 0.99, 1.00]$ to generate predictions $w(p_2)$ for the appropriate algebraic function, and then minimized the sum of squared residuals (SSE) between our model prediction of $w(p_2)$ and these values. Specifically, we determined the value of z_1 (where $z_2 = 1 - z_1$) and β in Equation 3 that minimized SSE. The parameters used to generate the algebraic functions are as follows: in Equation 4, $\delta = 0.77$, $\gamma = 0.44$, from Gonzalez and Wu (1999); in Equation 5, $\beta = 1.0$, $\gamma = 0.65$, from Prelec (1998); and in Equation 6, $\gamma = 0.61$, from Tversky and Kahneman (1992).

Application to Data From Wu and Gonzalez (1996)

The data reported by Wu and Gonzalez (1996) consist of choice probabilities (pooled across 105 participants) for each of 40 binary choices. They used a choice procedure among five different base “risky” and “safe” gamble pairs. For each pair, a “ladder” with eight “rungs” was created by incrementally increasing the probability of obtaining a common outcome shared by the two gambles. Their hypothesized concavity/convexity conditions suggest a pattern of choices that shows increasing preference (as the common outcome probability increases) for the risky option up to an inflection point, followed by increasing preference for the safer option (see Wu & Gonzalez, 1996, for theoretical and procedural details).

We derived choice predictions for DFT using the equations in Appendix A and compared them with the empirical data. Following Wu

and Gonzalez, we used a power utility function to represent an outcome evaluation. This function allows us to model decreasing marginal sensitivity and (as a result) incorporate risk aversion. Specifically, we used Equation A3 with $u(x_i) = x_i^\alpha$, where α was a free parameter fit to the data. We derived each w_i using Equation 2. The other parameter of the DFT choice mechanism, the decision threshold bound, was preset and held fixed at a value (three) obtained from previous applications of that model. This produced a total of three free parameters, α , β , and z , that were found so as to minimize the SSE between the observed and predicted choice probabilities.

We derived choice predictions for CPT following the procedure of Wu and Gonzalez (1996; see especially p. 1685). Specifically, for each gamble we calculated the rank-dependent utility by summing across two nonzero outcomes:

$$\begin{aligned} \text{RDU}(A) = & w(p_2)u(x_2) \\ & + [w(p_1 + p_2) - w(p_2)]u(x_1). \end{aligned} \quad (\text{C1})$$

where $w(p_i)$ is defined using Equation 6 and $u(x_i) = x_i^\alpha$. We then used a logistic function to determine choice probabilities:

$$\text{Pr}[A | \{A, B\}] = 1 / (1 + e^{\text{RDU}(B) - \text{RDU}(A)}). \quad (\text{C2})$$

This produced a total of two free parameters, α , and γ , that were found so as to minimize the SSE between the observed and predicted choice probabilities.

(Appendices continue)

First we fit one set of parameters to all five choice ladders simultaneously. The best fitting parameters for the attention process were $\beta < .01$ and $z = .02$ (with $\alpha = .19$). The fit value of z resulted in initial state probabilities of .9604, .0392, and .0004 for the lowest, moderate, and highest outcomes of each ternary gamble, respectively. For the combined fit, the model achieved $R^2 = 0.57$ (SSE = .183). For comparison, the CPT model produces $R^2 = 0.43$ (SSE = .246) for a one-parameter weighting function, and only nominal improvement using a two-parameter function.

Finally, it is important to note that the performance of the process model is not simply due to added complexity or an increase in parameters, as evidenced by the applications of the generalization criterion method (Busemeyer & Wang, 2000) reported in the text. This procedure involves fitting model parameters to a subset of empirical conditions and then testing the models with these parameters fixed under the remaining conditions. For the current application, Wu and Gonzalez (1996) obtained data within a given "ladder" or choice pair for eight different probability values of the common outcome: {0, 0.10, 0.20, 0.30, 0.45, 0.60, 0.75, 0.90}. We first fit both the attention process model and the RDU model presented in Wu and Gonzalez (1996) to choices using the lower four probability values and then used the resulting parameter estimates to predict choices among the higher four probability values; we then reversed the role of the choice sets. These tests, reported in the text, provide strong evidence that the performance of the process model is robust across data sets, as it accounts for the same amount of participant variability even when parameters are not estimated from the test data.

Importantly, it seems that the process model, not just the DFT choice mechanism, is respon-

sible for the result. We also examined the use of a logistic function (Equation C2, with utilities calculated via Equation 1 rather than Equation C1) to determine choice probabilities, thus following the procedure of Wu and Gonzalez (1996) in all respects except the weighting function. In this case, the model still outperformed the reported fit of prospect theory, but did not quite match the performance with the DFT choice model. Best-fitting parameters in this case were $z = .05$, $\beta = .79$, and $\alpha = .39$; this produced SSE = .21 and $R^2 = 0.52$.

Application to Data From Gonzalez and Wu (1999)

The data reported by Gonzalez and Wu (1999) consist of certainty equivalents (CE) for 165 individual gambles, for each of 10 individuals. We derived predictions for each of the 10 individuals, as well as for the set of median CEs, as follows. The dwelling probability β was again fit to the data, subject to the constrained interval [0, 1]. Furthermore, although all gambles in this task were binary, for consistency the same binomial distribution method (see Appendix B) as in the previous application was used to determine the distribution over initial states in the weighting model with one parameter, z . For each gamble G , we computed the simple EU(G) Equation 1, with $u(x_i) = x_i^\alpha$ and derived $w(p_i)$ again using Equation 2. Then, we used the inverse of the value function, $u^{-1}(x_j) = x_j^{1/\alpha}$, to transform EU(G) into a predicted certainty equivalent: CE = u^{-1} (EU(G)) = EU(G) $^{1/\alpha}$. We found the values of the free parameters α , β , and z that minimized the SSE between the reported and predicted CEs separately for each individual, and for the median data.

Received September 10, 2014

Revision received June 2, 2015

Accepted November 25, 2015 ■

E-Mail Notification of Your Latest Issue Online!

Would you like to know when the next issue of your favorite APA journal will be available online? This service is now available to you. Sign up at <http://notify.apa.org/> and you will be notified by e-mail when issues of interest to you become available!