RUNNING HEAD: COMPUTATIONAL MODEL OF DECISION WEIGHTS

A computational model of the attention process used to generate decision weights

Joseph G. Johnson

Miami University

Jerome R. Busemeyer

Indiana University

February 22, 2008

Address correspondence to: Joseph G. Johnson; Department of Psychology; Miami University; Oxford, OH 45056. Phone: (513) 529-2400; Fax: (513) 529-2420. Email should be sent to johnsojg@muohio.edu.

Abstract

During the past 50 years, a number of paradoxes have led to the downfall of expected utility theory as a viable description of decision making. Recently, rank-dependent theories (e.g. cumulative prospect theory) have arisen as the "leading" descriptive approach. A key component of these theories is the decision weight that reflects the importance of consequences when evaluating an action. Several researchers have had success in fitting functional forms for decision weights to account for empirical data. However, little is known about the cognitive source of these weights. The current work presents a computational model that derives decision weights from elementary attentional processing mechanisms. It is shown that the attention process model provides a coherent explanation for several robust phenomena including some that cannot be explained by rank-dependent utility theories.

Keywords: Risky decision making, attention, Allais paradox, Branch independence, Stochastic dominance

A computational model of the attention processes used to generate decision weights

When faced with a decision under risk or under uncertainty, a person needs to seriously consider each of the possible consequences or outcomes of the decision. For example, if a person is faced with a decision about a serious medical surgery procedure, he or she is encouraged to spend time thinking about what happens if it succeeds, or what happens if it fails, or to think about possible negative side effects. How much consideration, or weight, or attention does one give to each of these possible outcomes? What determines these weights, and how do they influence our decisions? The purpose of this article is to present a new cognitive model that describes how much attention a decision maker gives to each possible outcome when faced with decisions under risk. First, we describe past research on this question and review the basic facts of decision making that any model needs to describe; second, we present an attention process for decision weights, and show how it accounts for these basic facts; third we present additional experimental tests directed at the assumptions of the attention model; and finally we conclude with a discussion of the advantages produced by a cognitive modeling approach to decision making.

Past Research on Decision Weights

In this section, we present the progression of ideas that has taken place in the field of risky decision theory over the past 50 years. The story starts out simple, and it’s very interesting, but it soon becomes unavoidably complicated. This is just the nature of the field, yet we have tried to make it simple to understand through concrete examples.

EV theory

For the past 50 years, decision theorists have examined questions about risky decision making using simple gambles. For example, suppose you had a choice between the following two options:

Option A_0: Win $1000 for sure,

Option B_0: Winning $10,000 with 1/10 chance, or $500 with 4/10 chances, or nothing with 5/10 chances.

How much weight should be given to each of the three payoffs for option B_0? At first, one might think that the right answer 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: $EV(B_0) = (.10)\cdot(+\$10,000) + (.40)\cdot(\$500) + (.5)\cdot(\$0) = \$1200$, which is better than $1000 for sure, and so this EV rule would lead one to choose the risky option B_0 over A_0. However, many people may be risk averse and fear the possibility of getting nothing at all, and so they may prefer the safe option (A_0) over the risky gamble (option B_0). In fact, this is what we all do when we buy car or house insurance. So generally, the EV rule fails to describe risky decision making. Instead, very early on in the history of the field (Bernoulli, 1738/1954), decision theorists postulated that we assign a subjective worth to each outcome, called a utility, and choose the option that maximizes expected utility.

A common example of a utility function is $u(x) = x^{\alpha}$, where the exponent $\alpha < 1$ is used to produce diminishing marginal utility. This means that a change from a small to medium amount of money is worth more than a change from medium to a large amount even when the difference in dollars remains constant (a Weber's Law for money). This utility function is popular because it accounts for risk aversion, i.e. a person tends to prefer a sure thing over a gamble with equal expected value. This risk aversion parameter may vary across individuals producing different preferences for gambles across people. For example, if we use a common empirical estimate of around $\alpha = 0.75$, then $u(x) = x^{.75}$ in which case $EU(B_0) = (.10)\cdot u(+\$10{,}000) + (.4)\cdot u(\$500) + (.5)\cdot u(\$0) = (.10)\cdot(10{,}000)^{.75} + (.4)\cdot(500)^{.75} = 142.29$ which is now lower than $EU(A_0) = u(\$1000) = 1000^{.75} = 177.82$, and so option A_0 is chosen.

EU Theory

More generally, consider a three outcome gamble that produces outcome x_1 with probability p_1, outcome x_2 with probability p_2, and outcome x_3 with probability p_3. For brevity, we will denote this gamble as $G = (x_1, p_1 ; x_2, p_2 ; x_3, p_3)$. The expected utility of this gamble is defined as $EU(G) = p_1\cdot u(x_1)+ p_2\cdot u(x_2)+ p_3\cdot u(x_3)$. According to this rule, the utility of each outcome, $u(x_i)$, is again weighted by the probability, p_i, of that outcome. In other words, the EU rule *still* recommends that a person should *weight* each outcome by its *probability* of occurrence. The EU rule reduces to the EV rule if $u(x_i) = x_i$. One important lesson from EU theory is that one cannot say that a particular choice is rational or not, because that depends on an individual's utility function $u(x_i)$. Instead, rationality is defined in terms of consistency of choices with the axioms of EU theory described next.

Later in the history of decision theory, von Neuman and Morganstern (1944) discovered that the EU rule requires a decision maker to obey a simple set of axioms. One of the axioms that must be satisfied by the EU rule is the independence axiom. For example, consider a choice between (M stands for million)

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

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

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 is fine with utility theory. However, the independence axiom requires that the choice between A_1 and B_1 above is consistent the choice between A_1* and B_1* below:

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

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

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. Viewing the right hand side, we see that all we did 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 axiom requires 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 of the EU rule. 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

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

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

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

B_2*: ($5M, .16; $0M, .84) = ($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. According to the EU rule, 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). In short, the EU rule systematically fails to describe how people choose between risky options.

SEU theory

To account for the common ratio and common consequence effects, Kahneman and Tversky (1979) proposed prospect theory , which assumes that the objective probabilities, p_i of the EU rule are replaced with *subjective probabilities,* $\pi(p_i)$ in the prospect utility (PU) rule. The subjective transformation $\pi(p_i)$ is assumed to be a monotonic function that ranges from $\pi(0) = 0$ to $\pi(1) = 1$. For example, a power function,

$\pi(p) = p^{\gamma}$, has frequently been used here as well, where the exponent γ determines the shape of the nonlinear transformation (e.g., Kahneman & Tversky, 1979). The fact that $\pi(p_i)$ is nonlinear implies that these subjective probabilities *do not* follow the rules of objective probability theory. For example, the preference for A_1 over B_1 implies

$$SEU(A_1) = u(\$1M) > \pi(.10)\cdot u(\$5M) + \pi(.89)\cdot u(\$1M) = SEU(B_1),$$

and the preference for B_1*over A_1* implies

$$SEU(B_1^*) = \pi(.10)\cdot u(\$5M) > \pi(.11)\cdot u(\$1M) = SEU(A_1^*),$$

and together these two preferences imply that $[1–\pi(.89)] > [\pi(.11)–0]$. For example, if $\gamma = 1.3$ then $[1–(.89)^{1.3}] = .14 > [(.11)^{1.3}–0] = .06$. The latter implies that subjective probabilities are more sensitive to changes downward from the upper extreme (1.0) as compared to changes upward from the lower extreme (0.0), even though objectively these changes are identical.

More generally, for a three outcome gamble, $G = (x_1, p_1 ; x_2, p_2 ; x_3, p_3)$, the subjective expected utility is defined as: $SEU(G) = \pi(p_1)\cdot u(x_1)+\pi(p_2)\cdot u(x_2)+ \pi(p_3)\cdot u(x_3)$ (Edwards, 1962). According to this rule, the utility of each outcome, $u(x_i)$, is now weighted by a nonlinear transformation of the outcome probability, $\pi(p_i)$. In other words, the SEU rule proposes that people *weight* each outcome by its *subjective* rather than objective probability of occurrence. The EU rule is a special case of the SEU rule in which $\pi(p_i) = p_i$.

It was not long before researchers found problems with prospect theory as well. First, it was discovered that PT wrongly predicted people would be willing to choose a gamble that was obviously dominated by another gamble (see Starmer, 2000, for a review). Second, PT did a poor job when it was applied to gambles with more than two nonzero outcomes (Lopes & Oden, 1999). Third, and perhaps most telling, is the violation of *branch independence* (Birnbaum & McIntosh, 1996; see also Birnbaum & Chavez, 1997; Birnbaum & Veira, 1998). When presented with the choices shown below

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 because A_3 gives only one chance for producing a ‘very poor’ outcome but B_3 provides two chances. However, when presented with the pair:

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

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

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 would cancel out for any SEU rule. 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 for an SEU rule. After cancelling out the common terms, there is no difference between these pair of choices. Thus any SEU rule must satisfy branch independence (select A_3 and A_3*, or select B_3 and B_3*), which fails to explain these findings.

The problem with SEU theories is that 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. So, the next theoretical response of decision theorists to these empirical challenges was to devise an alternative way to transform objective probabilities in a manner that makes the weight of an outcome depend on both its probability and its rank.

RDU Theory

The remedy, called rank dependent utility (RDU) theory, is somewhat complicated. The basic idea is to transform the *decumulative* probabilities (the probability of winning x_i or *more*) rather than the outcome probabilities (the probability of winning x_i *itself*). Specifically, the *weight* given to an outcome x_i, now called the *decision weight* and denoted $w(x_i)$, equals the difference between two transformed *decumulative* probabilities—the transformed probability of getting outcome x_i or better, and the transformed probability of getting something strictly better than outcome x_i (Quiggin, 1982; Yaari, 1987; Tversky & Kahneman, 1992). For example, considering option A_3, the weight given to outcome \$24 equals $w(\$24) = \pi(2/3) - \pi(1/3)$, because the probability of getting \$24 or better is 2/3, and the probability of getting something strictly better than \$24 is 1/3. This is different than the weight $w(\$10) = \pi(1) - \pi(2/3)$ given to outcome \$10 in gamble A_3. Once again, the transformation, $\pi(P)$ is a monotonically increasing function of the *decumulative* probability, P, ranging from $\pi(0) = 0$ to $\pi(1) = 1$.

To see how the RDU model works, let us re-examine the branch independence violation described earlier. The choice of A_3 over B_3 implies

$$\mathrm{RDU}(A_3) = [\pi(2/3)-\pi(1/3)]\cdot u(\$24) + [\pi(1/3)-\pi(0)]\cdot u(\$45) >$$

$$[\pi(2/3)-\pi(1/3)]\cdot u(\$12) + [\pi(1/3)-\pi(0)]\cdot u(\$96) = \mathrm{RDU}(B_3).$$

The choice of A_3* over B_3* implies

RDU(A_3*) = [π(1)–π(2/3)]·*u*($24) +[π(2/3)–π(1/3)]·*u*($45) <

[π(1)–π(2/3)]·*u*($12) +[π(2/3)–π(1/3)]·*u*($96) = RDU($B_3$*).

Together these choices imply $\frac{\pi\left(\frac{2}{3}\right)-\pi\left(\frac{1}{3}\right)}{\pi\left(\frac{1}{3}\right)-\pi(0)} > \frac{\pi(1)-\pi\left(\frac{2}{3}\right)}{\pi\left(\frac{2}{3}\right)-\pi\left(\frac{1}{3}\right)}$. In this case, the middle outcome (with the weight [π(2/3)–π(1/3)]) has a greater decision weight than either extreme outcome even though the probabilities are all equal to 1/3! Note that the rank order of outcomes changes across the two pairs: the middle outcome for the {A_3,B_3} pair favors A_3, but the middle outcome for the {A_3*,B_3*} pair favors B3*. This change in the rank order of the outcomes explains why people violate branch independence.

More generally, for a three outcome gamble G = (x_1, p_1 ; x_2, p_2 ; x_3, p_3), with $0 < x_1 < x_2 < x_3$, the rank dependent utility is defined as:

RDU(G) = $w(x_1) \cdot u(x_1) + w(x_2) \cdot u(x_2) + w(x_3) \cdot u(x_3)$,

$w(x_1) = [\pi(1.0) - \pi(p_2+p_3)]$,

$w(x_2) = [\pi(p_2+p_3) - \pi(p_1)]$,

$w(x_3) = [\pi(p_1) - \pi(0)]$.

Note that according to this rule, the utility of each outcome, $u(x_i)$, is weighted by a *decision weight*, $w(x_i)$, which now depends on both its *probability and its rank order*. This is an important step because the decision weight for an outcome no longer depends solely on the probability of the outcome, and instead it also depends on its rank value. If $\pi(P_i) = P_i$, then $w(P_i) = p_i$, and the RDU rule reduces to the EU rule. Cumulative prospect theory (Tversky & Kahneman, 1992) is perhaps the most popular version of RDU theory.

Beyond RDU

It now 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 & Navarette, 1998; Birnbaum, Patton, & Lott, 1999; Loomes, Starmer, & Sugden, 1992). Consider the following pair of choices:

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

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

People choose A_4 over B_4 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 violate stochastic dominance with the pair {A_4, B_4} but they satisfy it with the pair {A_4*, B_4*} (Birnbaum, 2005; Birnbaum & Navarette, 1998; Birnbaum, Patton, & Lott, 1999; Loomes, Starmer, & Sugden, 1992).

More generally, Option B stochastically dominates option A if and only if the decumulative probability for option B is always greater than or equal to the decumulative probability for option A, i.e., Pr(obtaining an outcome ≥ x | B) ≥ Pr(obtaining an outcome ≥ x | A) for all x, and the inequality is strict for at least one x. The problem for RDU theories is that if B stochastically dominates A (in objective probabilities), then monotonicity of the π function implies that B dominates A using the decision weights, and the latter implies that B has a greater RDU than A. Therefore, RDU theories cannot account for empirically-observed choice of a stochastically-dominated option.

Once again, decision theorists have reacted by positing even more complex forms for the decision weights that apply to each outcome. Birnbaum (2008) argues for a Transfer of Attention Exchange (TAX) rule, which succeeds in explaining all of the above results. A basic idea of the TAX model is that weight is taken away from higher rank (better) outcomes and given to lower rank (worse) outcomes. For a three outcome gamble G = (x_1, p_1 ; x_2, p_2 ; x_3, p_3), with $0< x_1 < x_2 < x_3$, the TAX rule is

$$\text{TAX(G)} = w(x_1)\cdot u(x_1)+w(x_2)\cdot u(x_2)+ w(x_3)\cdot u(x_3),$$

$$w(x_1) = \frac{\pi(p_1)+\frac{\delta}{4}\cdot\pi(p_2)+\frac{\delta}{4}\cdot\pi(p_3)}{\pi(p_1)+\pi(p_2)+\pi(p_3)}$$

$$w(x_2) = \frac{\pi(p_2)-\frac{\delta}{4}\cdot\pi(p_2)+\frac{\delta}{4}\cdot\pi(p_3)}{\pi(p_1)+\pi(p_2)+\pi(p_3)}$$

$$w(x_3) = \frac{\pi(p_3)-\frac{2\cdot\delta}{4}\cdot\pi(p_3)}{\pi(p_1)+\pi(p_2)+\pi(p_3)},$$

where δ>0 is a free parameter (often δ = 1 is used). We do not attempt to justify this form here, except to say that it works empirically (Birnbaum, 2008). At this point the reader may be asking – *where* do these decision weights come from? This is exactly the question that we are trying to answer by postulating a simple cognitive process that generates these complex decision weights.

An Attention Process Model for Decision Weights

In this section, we describe an attention process that generates the decision weights used for risky decision making. Our idea originated from decision field theory (DFT, Busemeyer & Townsend, 1993), according to which a decision weight for an outcome reflects the proportion of time spent anticipating and thinking about that outcome during deliberation. However, this attention process model is also consistent with many other theoretical frameworks such as decision by sampling theory (Stewart, Chater, & Brown, 2005) or weight utility models (Birnbaum, 2008). Therefore, we present the attention process as a general explanation for decision weights, which could be used in various types of decision rules or choice mechanisms. (Appendix A provides details about how this process would be implemented in DFT).

The basic idea is that the decision maker tries to anticipate or predict what outcomes will be generated by each gamble. The decision weight is identified as the proportion of times (or probability) an outcome is predicted to occur for a gamble. These predictions are based on mental simulations of the random device actually used to generate outcomes (akin to the "simulation heuristic" of Kahneman and Tversky, 1982).

Consider for example, the choice for the pair {A_4, B_4} described earlier. In these experiments, participants are informed that the outcomes of a gamble are randomly sampled from an urn, and a separate urn is used for each gamble. 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_4: ($12,

.10; \$90, .05; \$96, .85). 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 be finished. 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 not predict this but remain considering it, 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. The proportion of times (or probability) that say \$96 is predicted to occur with gamble A_4—i.e., the proportion of times this thought process 'stops' on \$96—is then identified as the decision weight for the \$96 outcome.

Markov model

Formally, we can model this attention process as a Markov chain model (Karlin & Taylor, 1975). Figure 1 shows the Markov model for three outcomes $0 \leq x_1 \leq x_2 \leq x_3$. 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 'looking' at one of these three states, and the probability of starting in state x_i is denoted z_i 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[1].

Once the attention process is located at a particular state, x_i, one of three events can happen: (E1) a prediction could be made for that state, i.e. 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, i.e. 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 adjacent state, i.e. transit one step to the right or left from state x_i.

INSERT FIGURE 1 ABOUT HERE

The probability of event E1 is denoted ϕ_{i*} in Figure 1, which is assumed to be equal to the objective probability, $\phi_{i*} = p_i$. For Gamble A_4, this means $\phi_{1*} = .10$, $\phi_{2*} = .05$, and $\phi_{3*} = .85$. This is empirically justified by the probability matching law for prediction (Estes, 1959). The probability of event E2 is denoted ϕ_{ii} in Figure 1, which is assumed to be equal to $\phi_{ii} = \beta\cdot(1 - \phi_{i*})$. This is based on the joint probability that the person does not predict outcome x_i and yet this person's attention remains focused on outcome x_i for another time step. The parameter, $0 \leq \beta \leq 1$, represents the probability of staying or dwelling on the same state given that a prediction did not occur. This dwell rate parameter determines the rate of attentional drift up and down the outcome scale. 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 $\phi_{12} = (1–\phi_{1*})\cdot(1–\beta)$; similarly, if attention is focused on the highest payoff x_3, then the probability of shifting focus down to x_2 must equal $\phi_{32} = (1–\phi_{3*})\cdot(1–\beta)$; these are both required because the transition probabilities out of a state must sum to unity. If attention is focused on an intermediate state, x_2 in Figure 1, then the probability of transiting to the adjacent state equals $\phi_{21} = \phi_{23} = (1–\phi_{i*})\cdot(1–\beta)/2$. That is, we assume that attention is equally likely to step up or down the outcome rank. In sum, all of the state transitions are computed from the dwell rate β and the objective outcome probabilities (p_1, p_2 , p_3).

Model Predictions

It is straightforward to directly compute the final probabilities of predicting each outcome (i.e., the decision weights) from the Markov model using simple matrix operations. First, we form an initial state probability row vector as **Z'** = [z_1 z_2 z_3]. Second, we form a transient matrix **Q** and an absorbing matrix **R** from the outcome probabilities (p_1, p_2 , p_3) and the dwell rate β:

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

The probabilities of predicting each of the three outcomes (i.e., the decision weights) form a row vector **W** = [w_1 w_2 w_3]. These decision weights are then computed from the Markov model by the simple matrix equation (see Karlin & Taylor, 1975):

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

The matrix **I** is a 3 x 3 identity matrix (ones on the diagonals, zeros on the off-diagonals), and $(\mathbf{I} - \mathbf{Q})^{-1}$ is the inverse of the matrix (**I** –**Q**). Equation 1 can be computed very easily using a mathematical programming language such as Matlab or Mathematica. For example, if we apply this model to gamble A_4 with $z_1 = 1$ and β = .70, we obtain

$$[w(\$12) \quad w(\$90) \quad w(\$96)] = [.40 \quad .16 \quad .44] =$$

$$= [1 \quad 0 \quad 0] \cdot \begin{bmatrix} 1-(.7)(.9) & -(.3)(.9) & 0 \\ -(.5)(.3)(.95) & 1-(.7)(.95) & -(.5)(.3)(.95) \\ 0 & -(.3)(.15) & 1-(.7)(.15) \end{bmatrix}^{-1} \cdot \begin{bmatrix} .1 & 0 & 0 \\ 0 & .05 & 0 \\ 0 & 0 & .85 \end{bmatrix}.$$

It is interesting to note that if we use the typical parameters from Birnbaum's (2008) TAX model (δ=1 and $\pi(p) = p^{.7}$), then the TAX model produces very similar decision weights for gamble A_4: *w*($12) = .37, *w*($90) = .26, *w*($96) = .37.

Equation 1 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 Equation 1 reduces to (see Appendix B):

$$w_2 = \frac{(1-\beta)z_1 p_2{}^2 + (1-p_2\beta)z_2 p_2}{(1-\beta)p_1 + (1-\beta)p_2{}^2 + \beta p_1 p_2} \qquad (2)$$

Note that decision weights sum to one, so the weight for the other 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 **Z** from a single free parameter by assuming that this initial probability distribution has a binomial distribution (see 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.

Finally, note that we can recover the objective probabilities from the attention process model by assuming $z_i = p_i$, and $\beta = 1$. In this special case, the probability of first considering an outcome is equal to the stated probability, and this outcome is dwelt upon until it is selected.

Predictions for Independent versus Dependent Gambles. The attention process model makes some *unique* predictions for choices between gambles, depending on whether their outcomes are independent or dependent. For example, if the outcomes of each gamble are believed to be determined by a separate urn, then they will be treated *independently*. Referring to the pair {A_3, B_3}, any outcome from A_3 can occur with any outcome of B_3, so that for example, $24 could occur with A_3 and $96 could occur with B_3. In this case, it is assumed that an independent prediction is made for each gamble.

If the outcomes of each gamble are determined jointly, such as by a single draw of a colored marble from one urn, then they will be *dependent*. Referring again to the gambles {A_3,B_3}; if a red marble produces $10 for A_3 and $10 for B_3, and a blue marble produces $24 for A_3 and $12 for B_3, and a green marble produces $45 for A_3 and $96 for B_3, then selecting a single marble color determines the outcomes of both gambles. In particular, a $24 from A_3 could not occur with $96 from B_3, and instead $24 from A_3 could only occur with $12 from B_3.

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. This unique prediction of the attention process model is empirically tested in one of the applications described later. Note that previous decision weight models mentioned above make no distinction between independent and dependent gambles.

Psychological interpretation of parameters

Equations 1 and 2 are not computed by the decision maker, and instead, these equations are the mathematically expected consequences of the attention process summarized in Figure 1. There are only two free parameters in the attention process model: z determines the initial distribution **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 probability z_i affects 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. 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 (curvature) 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 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).[2]

INSERT FIGURE 2 ABOUT HERE

In sum, our motivation was to provide a dynamic processing mechanism for the generation of decision weights. On the one hand, this mechanism could be used to determine the decision weights that are inserted into weighted utility rules of risky decisions. On the other hand, this mechanism represents a "missing" component used to determine the attention weights of decision field theory or decision by sampling theory for risky decisions.

Applications of the Attention Weight Process Model to Empirical Findings

The introductory section pointed out that the EV, EU, SEU, and RDU models cannot account for all of the qualitative patterns reviewed earlier. To examine very clearly the ability of the attention process model to explain all of 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 ($z = 0$), and we use a moderate level for the dwelling probability ($\beta = 0.70$). Second, we based the choices on a simple weighted utility formula: $\text{WU(G)} = \sum w(x_i) \cdot u(x_i)$, where $w(x_i)$ is now computed from Equation 1.[3] 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 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. 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, 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 the TAX model (Birnbaum, 2008) that the lowest outcome takes attention away from higher outcomes; it is also consistent with the hypothesis from the priority heuristic

(Brandstatter, Gigerenzer, & Hertwig, 2006) that decision maker's start by comparing the lowest payoffs.

As can be seen in Table 1, the process model correctly predicts each of the pairwise choices that give rise to the qualitative effects introduced above. That is, the inconsistent choice pattern (A, B*) is predicted for each pair of choices, producing violations of independence in line with the common ratio and common consequence effects, violations of branch independence, and violations of stochastic dominance (but not transparent stochastic dominance). Table 1 also gives the predicted weights in each situation.

INSERT TABLE 1 ABOUT HERE

Figure 3 shows how sensitive the model predictions are to the exact parameter values used here. Specifically, the figure shows the parameter value combinations that produce each of the four effects considered, 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, and the existing evidence is mixed; early work (Preston & Barrata, 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.

INSERT FIGURE 3 ABOUT HERE

New 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 reviewed above, including Birnbaum's TAX model (2008), provide any mechanisms for explaining these manipulations.

Juxtaposition effects. The juxtaposition effect is obtained simply by manipulating the way two gambles are displayed, but this in turn affects the way people attend to outcomes.[4] Consider the 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, i.e., 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[Head] = Pr[Tail] = 0.5. Consider a choice between the pair

A_5: ($100, H; –$100, T),

B_5: ($99, H; –$101, T).

Obviously, A_5 stochastically dominates B_5, and people almost always choose A_5 (Diederich & Busemeyer, 1999). Next consider a choice between the pair

A_5: ($100, H; –$100, T);

B_5*: (–$101, H; $99, T).

Note that in terms of probabilities and outcomes, B_5* is identical to B_5 and so A_5 still stochastically dominates B_5*. However, people are almost indifferent between these two options, and the probability of choosing A_5 drops down close to .50 (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 predict 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 A_5 and B_5 is identical to the choice between A_5 and B_5*.

According to the attention process model, attention is now focused on the coin flip (refer back to the section on *Independent versus Dependent Gambles*). On the one hand, when faced with a choice between A_5 and B_5, the only two possible predicted outcomes are \$100 for A_5 and \$99 for B_5 (if attention is focused on the event heads); or –\$100 for A_5 and –\$101 for B_5 (if attention is focused on the event tails). A_5 is clearly better in both circumstances, there is no conflict at all, and so A_5 is always chosen. The decision maker would never predict a pair of outcomes such as \$100 for A_5 and –\$101 for B_5, because this combination of outcomes is impossible for this pair of actions. On the other hand, when faced with a choice between A_5 and B_5*, the only two possible predicted outcomes are \$100 for A_5 and –\$101 for B_5* (if attention is focused on the event heads); or –\$99 for A_5 and \$100 for B_5* (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, given the same model parameters as used to account for the effects in Table 1, and using decision field theory to compute probabilities (Appendix A), the process model predicts Pr[A_5 | {A_5, B_5}] > 0.99, and Pr[A_5 | {A_5, B_5*}] = 0.51, in line with the empirical result.

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 outcomes 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). Future work could incorporate process-tracing in tasks such as that of Weber and Kirsner (1997) to directly test these effects. 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 weight generation process, our approach offers an alternative explanation that is more congruent. By simply assuming that one dwells more on emotional (compared to non-emotional) outcomes—by increasing the dwell parameter β—the model predicts exactly the decreased slope for intermediate probabilities found by Rottenstreich and Hsee (2001), as can be seen by comparing across lines within a given panel in Figure 2. Note that increasing the dwell rate parameter also increases mean response time (see Appendix A), so our approach also makes a testable prediction that emotional outcomes should produce relatively longer decision times.

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, $w(p)$, for many values of p from various non-gamble 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 vs. 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 an RDU model. Second, we compare the model predictions to the empirically-determined probability weighting functions derived from ten 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) x 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.

A generalization criterion was used to compare the two models (Busemeyer & Wang, 2000). 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. Only the results for the generalization test are reported here.

Both models were allowed three free parameters: one risk aversion utility parameter and two parameters to compute the decision weights. We first fit both the attention process model and the RDU 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 twenty conditions with relatively high common outcome probabilities. In this generalization test, the attention process model (SSE = .117, R^2 = 0.58) outperformed the RDU model (SSE = .135, R^2 = .51).[5] 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 (R^2 = 0.56, SSE = .068) outperformed the RDU model (SSE = .128, R^2 = 0.16). The R^2 for both models may appear a bit low, but this is a false impression—one needs to recognize that these are *a priori* quantitative predictions for the relative frequencies obtained from twenty *new* conditions using *no* free parameters. These results provide empirical evidence that the attention process model not only explains the qualitative findings better than the RDU model, but also makes better quantitative predictions than a standard RDU model.

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 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.[6]

Gonzalez and Wu (1999) elicited certainty equivalents from ten 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. Model predictions are shown in Figure 4 for three representative individual participants, and for the median data (cf. Gonzalez & Wu, 1999, Figures 7 and 8); 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 β = 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 ten 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).

INSERT TABLE 2 ABOUT HERE

INSERT FIGURE 4 ABOUT HERE

Discussion

During the past 50 years, a great deal of empirical and theoretical progress has been made toward trying to understand how much weight a decision maker places on each of the various possible outcomes of a risky decision . The ideas started out simple by assuming that the weight equals the outcome probability. But the common ratio and common consequence effects forced decision theorists to change the weight into a nonlinear transformation of the outcome probability. Later, the branch independence violation forced decision theorists to change the weight again into a difference between two nonlinear transformed decumulative probabilities. Most recently, violations of stochastic dominance have driven decision theorists to seek even more complex forms. In short, the theory of decision weights has become very complicated, which makes one wonder where these complicated formulas come from? Much of this theory has been developed by decision scientists rather than cognitive scientists. Perhaps it is time to try and understand the cognitive processes that give rise to these decision weights.

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. First, it does a better job of explaining the qualitative findings than most of the previous decision weight theories, and it also makes more accurate quantitative predictions. Second, and more important, it makes new predictions for manipulations of attention (holding outcomes and probabilities fixed) that are supported by experiments yet cannot be explained by utility theories. First, the attention process model explains how decision weights may change depending on the nature of the random device used to generate the outcomes, even when the outcome probabilities remain the same. For example, the attention weight model predicts that decision weights change 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 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). Second, 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 states, but to any other states in the system. 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:

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

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. Considering this alternative procedural formulation and/or parameter settings, the proposed decision weighting 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. However, so far, only the model in Figure 1 has found to be completely successful in accounting for all of the important findings.

Future Directions. Additional empirical testing of the attention process models is needed. One important direction is to further validate the attention process using process tracing methods. One prediction that follows from the mechanism outlined in Figure 1 is that transitions should occur back and forth between an outcome and its probability within a gamble. This follows from the idea that decision maker considers whether or not to predict an outcome from a gamble on the basis of its probability. Not all cognitive models of decision making make this prediction. In particular, the priority heuristic (Brandstatter, et. al, 2006) predicts transitions between outcomes from different gambles. A recent process tracing study by Johnson, Schulte-Mecklenbeck, & Willemsen (2008) reports strong evidence for transitions between outcomes and their probabilities within a gamble rather than transitions between outcomes of different gambles.

The attention process model also makes predictions regarding the time it takes to generate a prediction. Specifically, factors that increase the dwell time parameter, β, are predicted to 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.

Just as empirical violations of theoretical properties have led to the current incarnations of utility theory (RDU and TAX), 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). These models account for results that have challenged traditional theories while adding novel predictions, parsimony, and psychological plausibility.

Authors' notes

The authors are extremely grateful to Doug Wedell for first suggesting this application of computational models to decision weighting, and to Michael Birnbaum, Elke Weber, Nathaniel Wilcox and the anonymous reviewers for additional constructive comments. We thank George Wu and Richard Gonzalez for making their data available to us.

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.

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

Bernoulli, D. (1738/1954). Exposition of a new theory on the measurement of risk. *Econometrica, 22*, 23-36.

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.

Birnbaum, M. H. (2008). New paradoxes of risky decision making. *Psychological Review.*

Birnbaum, M. H., & McIntosh, W. R. (1996). Violations of branch independence in choices between gambles. *Organizational Behavior and Human Decision Processes, 67,* 91–110.

Birnbaum, M. H., & Navarette, J. B. (1998). Testing descriptive utility theories: Violations of stochastic dominance and cumulative independence. *Journal of Risk and Uncertainty 17*, 49-78.

Brandstätter, E., Kühberger, A. & Schneider, F. (2002). A cognitive-emotional account of the shape of the probability weighting function. *Journal of Behavioral Decision Making, 15,* 79-100.

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*. Oxford, UK: Blackwell Publishing Co. 133-154

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

Busemeyer, J. R., & Townsend, J. T. (1992). Fundamental derivations for decision field theory. *Mathematical Social Sciences, 23*, 255-282.

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

Diederich, A. (2003). MDFT account of decision making under time pressure. *Psychonomic Bulletin and Review, 10*, 157-166.

Diederich, A. & Busemeyer, J. R. (1999). Conflict and the stochastic dominance principle of decision making. *Psychological Science,10*, 353-359.

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 Psychology, 47,* 304-322.

Dror, I. E., Busemeyer, J. R., & Basola, B. (1999). Decision making under time pressure: An independent test of sequential sampling models. *Memory and Cognition, 27*, 713-725.

Edwards, W. (1955). The prediction of decisions among bets. *Journal of Experimental Psychology, 50,* 201–214.

Edwards, W. (1962). Subjective probabilities inferred from decisions. *Psychological Review 69*, 109–135.

Estes, W. K. (1959). Component and pattern models with Markovian interpretations. In R. R. Bush and W. K. Estes, *Studies in Mathematical Learning Theory,* 9-53. Stanford, Stanford University Press.

Gonzalez, R., & Wu, G. (1999). On the shape of the probability weighting function. *Cognitive Psychology 38*, 129-166.

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).

Kahneman, D., & Tversky, A. (1979). Prospect theory: An analysis of decision under risk. *Econometrica, 47,* 263-291.

Kahneman, D. and Tversky, A. (1982). The simulation heuristic, In D. Kahneman, P. Slovic and A. Tversky (eds.) *Judgment under uncertainty: Heuristics and biases* (pp. 201-208). New York: Cambridge University Press.

Karlin, S., & Taylor, H. M. (1975). *A first course in stochastic processes, Vol. I.* New York: 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.

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: Academic Press.

Luce, R. D. (2000). Utility of gains and losses. Lawrence Erlbaum.

Preston, M., & Baratta, P. (1948). An experimental study of the auction-value of an uncertain outcome. *American Journal of Psychology, 61*, 183–193.

Quiggin, J.(1982). A theory of anticipated utility. *Journal of Economic Behavior and Organization, 3,* 323–343.

Rottenstreich, Y. & Hsee, C. K. (2001). Money, kisses, and electric shocks: An affective psychology of risk. *Psychological Science, 12,* 185-190.

Shah, A. K., & Oppenheimer, D. M. (2007). Easy does it: The role of fluency in cue weighting. *Judgment and Decision Making, 2(6)*, 371-379.

Starmer, C. (2000). Developments in non-expected utility theory: The hunt for a descriptive theory of choice under risk. *Journal of Economic Literature, 38,* 332-382.

Townsend, J. T., & Wenger, M. J. (2004). The serial-parallel dilemma: A case study in a linkage of theory and method. *Psychonomic Bulletin and Review, 11*, 391-418.

Tversky, A. (1967). Utility theory and additivity analysis of risky choices. *Journal of Experimental Psychology, 75*, 27−36.

Tversky, A., & Kahneman, D. (1992). Advances in prospect theory: Cumulative representations of uncertainty. *Journal of Risk and Uncertainty, 5*, 297–323.

Tversky, A., & Wakker, P. (1995). Risk attitudes and decision weights. *Econometrica, 63*, 1255–1280.

Weber, E.U., & Kirsner, B. (1997). Reasons for rank-dependent utility evaluation. *Journal of Risk and Uncertainty, 14*, 41-61.

Wedell, D. H., & Senter, S. M. (1997). Looking and weighting in judgment and choice. *Organizational Behavior and Human Decision Processes*, 70, 41-64.

Wu, G., & Gonzalez, R. (1996). Curvature of the probability weighting function. *Management Science, 42*, 1676–1690.

Yaari, M. E. (1987). The dual theory of choice under risk. *Econometrica, 55*, 95–115.

Appendix A: Integrating the attention weight model into decision field theory

Decision field theory (DFT) hypothesizes that each choice between a pair of gambles is determined by the following dynamic process. During deliberation, preferences are continuously updated on the basis of momentary shifts in attention across various outcomes of each gamble. At one moment, a decision maker may focus on one possible outcome of a gamble (e.g. potential gain) producing a momentary advantage for that option, but at the next moment, perhaps attention shifts to a different outcome (e.g., potential loss), producing a momentary disadvantage for that option. Across time, these momentary evaluations are integrated into a preference state that describes the decision maker's preference for each option at any given moment. Once the preference for a particular option exceeds a preset threshold criterion, the corresponding option is determined to be "good enough" or satisfactory and is chosen.

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 pair {A_4 , B_4}. At each time step, a valence is generated by considering a possible outcome from A_4, and also considering a possible outcome from B_4, 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_4, and using the parameters $z = 0$ and $\beta = .70$, then according to Equation 1, 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:

$$\mathrm{E}[V(t)] = \mu = \mathrm{E}[U_A(t)] - \mathrm{E}[U_B(t)], \qquad \text{(A1)}$$

$$\mathrm{Var}[V(t)] = \mathrm{Var}[U_A(t)] + \mathrm{Var}[U_B(t)] \text{ and } \sigma = \sqrt{\mathrm{Var}[V(t)]}. \qquad \text{(A2)}$$

The individual means and variances are determined as follows, for $j = \{\mathrm{A, B}\}$:

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

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

Here we have inserted the decision weights, $w(x_{kj})$, given by Equation 1 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): $\Pr[A \mid \{A, B\}] = \dfrac{1}{1 + \exp(-2 \cdot d \cdot \theta)}$

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, Busemeyer, Basola, 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 randomly 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):

$$E[\mathbf{t_i}^* \mid \mathbf{W}] = \mathbf{Z}'(\mathbf{I}-\mathbf{Q})^{-2}\mathbf{\Pi}^* h./\mathbf{W} \qquad \text{(A5)}$$

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.

Appendix B: Relevant derivations and solutions

Suppose a gamble has n outcomes ordered ($x_n > x_{n\text{-}1} > \ldots > x_{i+1} > x_i > x_1$); 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. The distribution over initial states, z_i contained in **Z** is defined using a binomial distribution that depends on the parameter z:

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

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, $\varphi_{i*} = p_i$.

The joint probability of not predicting state i but continuing to dwell on this state is then equal to $\varphi_{ii} = (1 - \varphi_{i*})\cdot\beta$, 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 $(1–\varphi_{i*})\cdot(1–\beta)$. We adopt the simple random walk assumption that transitions occur only to neighboring states (adjacent rank-ordered outcomes); then, for n outcomes, we obtain:

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

Derivation of process weighting model solution for binary case (Equation 2)

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

$$\mathbf{Q}_2 = \begin{bmatrix} (1-p_1)\beta & (1-p_1)(1-\beta) \\ (1-p_2)(1-\beta) & (1-p_2)\beta \end{bmatrix} = \begin{bmatrix} p_2\beta & p_2(1-\beta) \\ p_1(1-\beta) & p_1\beta \end{bmatrix}.$$

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 1 by subtraction from a 2 x 2 identity matrix and taking the inverse:

$$(\mathbf{I}-\mathbf{Q}_2)^{-1} = \begin{bmatrix} \dfrac{1+(p_2-1)\beta}{p_1(1-\beta)+p_2(1-\beta)+p_1p_2(2\beta-1)} & \dfrac{(p_1-1)(\beta-1)}{p_1(1-\beta)+p_2(1-\beta)+p_1p_2(2\beta-1)} \\ \dfrac{(p_2-1)(\beta-1)}{p_1(1-\beta)+p_2(1-\beta)+p_1p_2(2\beta-1)} & \dfrac{1+(p_1-1)\beta}{p_1(1-\beta)+p_2(1-\beta)+p_1p_2(2\beta-1)} \end{bmatrix}$$

We define the initial state vector generally, $\mathbf{Z}_2 = [z_1\ z_2]'$, and assume that the response or absorbing probabilities in **R** are equal to the objective probabilities (see text):

$$\mathbf{R}_2^* = \begin{bmatrix} p_1 & 0 \\ 0 & p_2 \end{bmatrix}$$

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

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

Appendix C: Quantitative fits of the process model to empirical data

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 several different “risky” and “safe” gamble pairs. For each pair, a “ladder” 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 to 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 $w(x_i)$ using Equation 1. 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.

First we fit one set of parameters to all five choice ladders simultaneously. The best fitting parameters for the attention process were $\alpha = 0.19$, $\beta < .01$, and $z = .02$. 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 RDU model presented in Wu and Gonzalez (1996) produces $R^2 = 0.43$ (SSE = .246) for a one-parameter prospect theory weighting function, and only nominal improvement using 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 two applications of the generalization criterion method (Busemeyer & Wang, 2000). 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. In this application, the attention process model (SSE = .117, $R^2 = 0.58$) outperformed the RDU model (SSE = .135, $R^2 = .51$). Next, we reversed the procedure and fit the models to choices involving the higher four probability values, then predicted choices for the lower four values. Once again, the performance of the process model paralleled its performance on the entire data set, obtaining $R^2 = 0.56$ (SSE = .068). The performance of the RDU model in this application suffered, however, obtaining $R^2 = 0.16$ (SSE = .128). These tests 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.

In summary, the process model provides at least an equally good fit as the algebraic weighting functions to the extensive choice data of Wu and Gonzalez (1996); it seems the process model even produces somewhat better fits and accounts for a larger proportion of variance. This does not simply seem to be the result of additional complexity, as evidenced by the generalization criterion method. Importantly, it seems that the process model, not just the DFT choice mechanism, is responsible for the result. We also examined the use of a logistic function 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 $R^2 = 0.52$.

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 ten individuals. We derived predictions for each of the ten 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 (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) using the equation given in the

Introduction, with $u(x_i) = x_i^{\alpha}$ and derived $w(p_i)$ again using Equation 1. Then, we used the inverse of the value function, $u^{-1}(x_j) = x_i^{1/\alpha}$, 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.

Tables

Table 1. Process model predictions for pairwise choices given in text.

Gamble	WU	w_1	w_2	w_3	w_4
Common consequence					
A_1	1.00	0.00	1.00		
B_1	0.99	0.03	0.96	0.00	
A_1*	0.01	0.99	0.01		
B_1*	0.04	0.99	0.01		
Common ratio					
A_2	2.90	0.03	0.97		
B_2	2.63	0.47	0.53		
A_2*	0.10	0.97	0.03		
B_2*	0.11	0.98	0.02		
Branch independence					
A_3	13.70	0.68	0.27	0.05	
B_3	10.89	0.68	0.27	0.05	
A_3*	39.29	0.68	0.27	0.05	
B_3*	40.64	0.68	0.27	0.05	
Stochastic dominance					
A_4	62.66	0.40	0.16	0.44	
B_4	60.64	0.24	0.20	0.56	
A_4*	49.85	0.27	0.28	0.12	0.33
B_4*	51.38	0.27	0.28	0.12	0.33

Notes: 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*, which are predicted by the process model weights with a linear utility function under "WU," where WU(A) > WU(B) and WU(B*) > WU(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 due to rounding.

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	0.30	0.97	0.68	98891	0.88
2	0.52	0.87	0.35	129560	0.91
3	0.43	0.92	1.00	116550	0.94
4	0.17	1.00	0.59	24388	0.93
5	0.84	0.98	0.14	238060	0.76
6	0.63	0.63	0.67	15817	0.99
7	0.30	1.00	0.57	59718	0.86
8	0.31	0.99	0.41	88122	0.84
9	0.36	0.61	0.58	48739	0.97
10	0.41	0.88	0.51	150450	0.88
Median	0.44	0.93	0.48	58309	0.94

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

Figure Captions

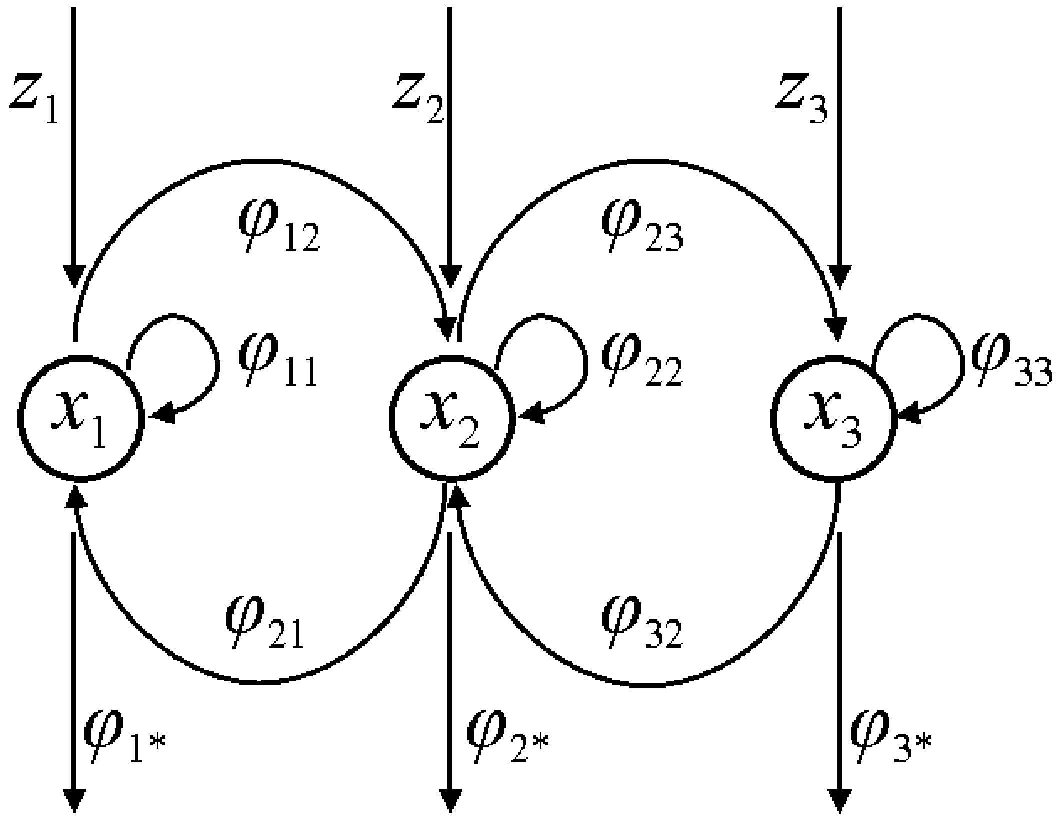
Figure 1. Schematic of the process weighting model. Circular nodes represent outcomes of a single gamble, and arrows represent possible probabilistic transitions in attention during a single "mental play." The probability z_i of first considering each outcome x_i is input to the model, along with the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

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, $z_i = 1/n = ½$, for $i = 1, 2$; and (d) Starting probabilities equal to objective probabilities, $z_i = p_i = \Pr[x_i]$, for for $i = 1, 2$.

Figure 3. Parameter space map producing effects for corresponding stimuli in Table 1. Points indicate the (z, β) parameter combinations, to a precision of 0.1, that produce (a) the common ratio effect, (b) the common consequence effect, (c) violations of branch independence, and (d) violations of stochastic dominance.

Figure 4. Model predictions for best-fitting weighting functions for (a) median data and (b) three participants from Gonzalez and Wu (1999). Compare to Figures 7 and 8 from Gonzalez and Wu (1999). Note the model's ability to account for qualitatively different shapes in the weighting function.

Figure 1



$$\varphi_{i*} = \Pr[x_i] = p_i$$

$$\varphi_{ii} = (1 - \varphi_{i*}) \cdot \beta$$

$$\varphi_{ii} + \varphi_{i(i+1)} + \varphi_{i(i-1)} = 1$$

Figure 2

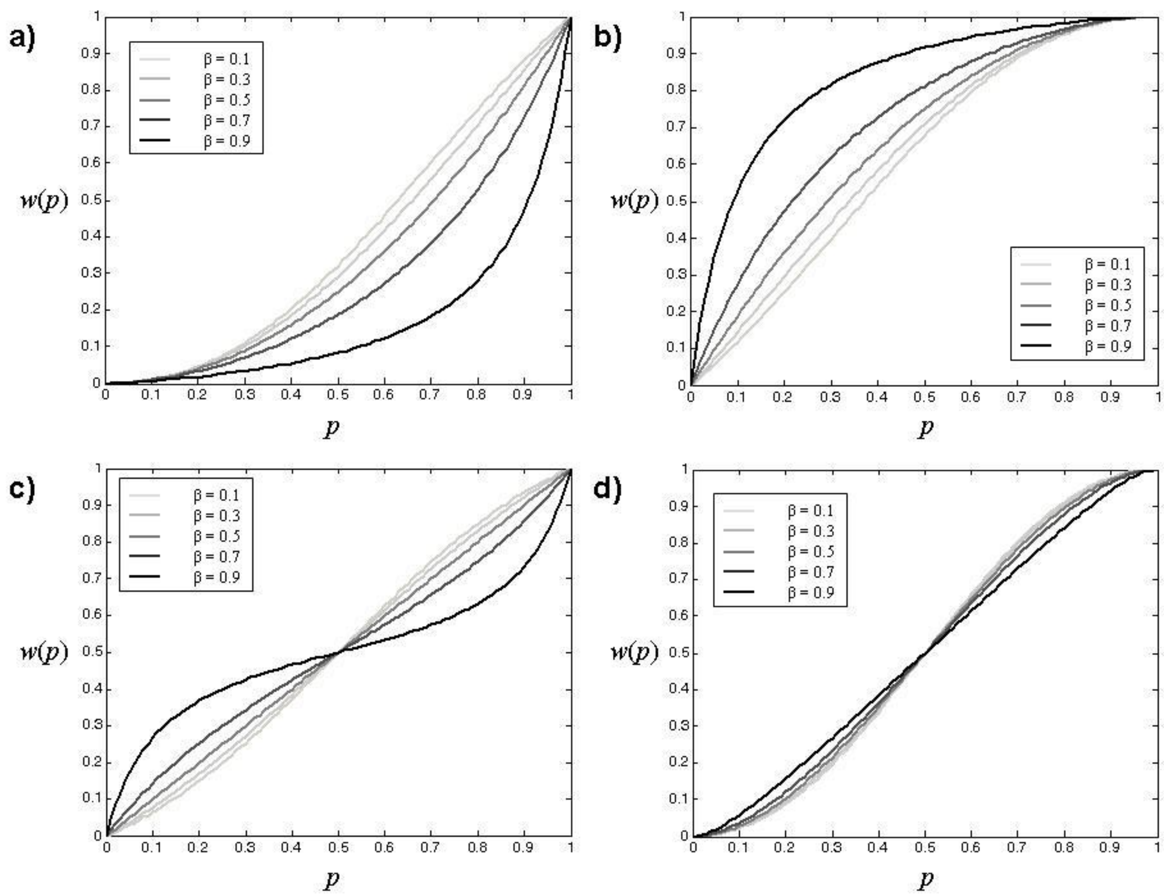


Figure 3

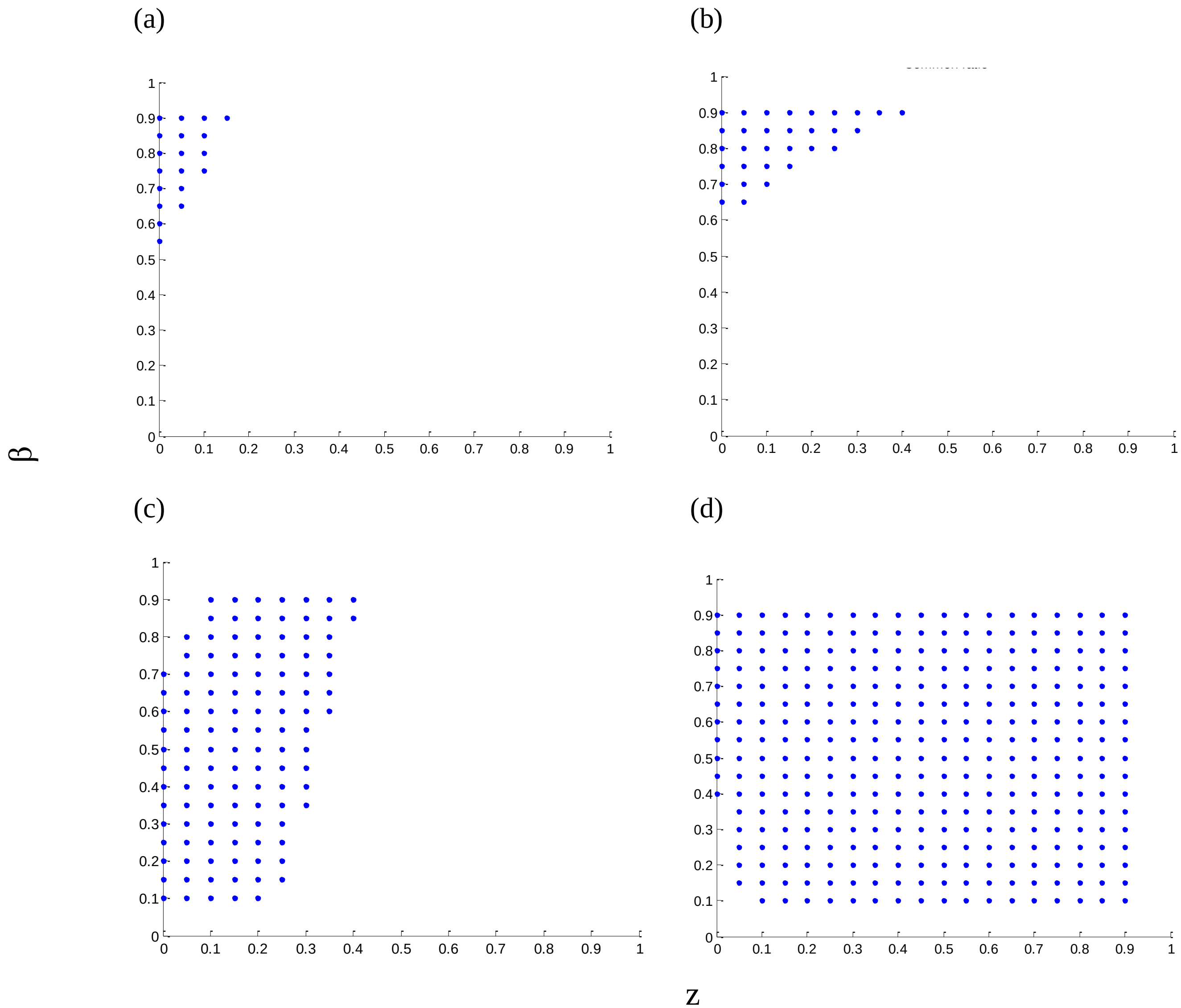
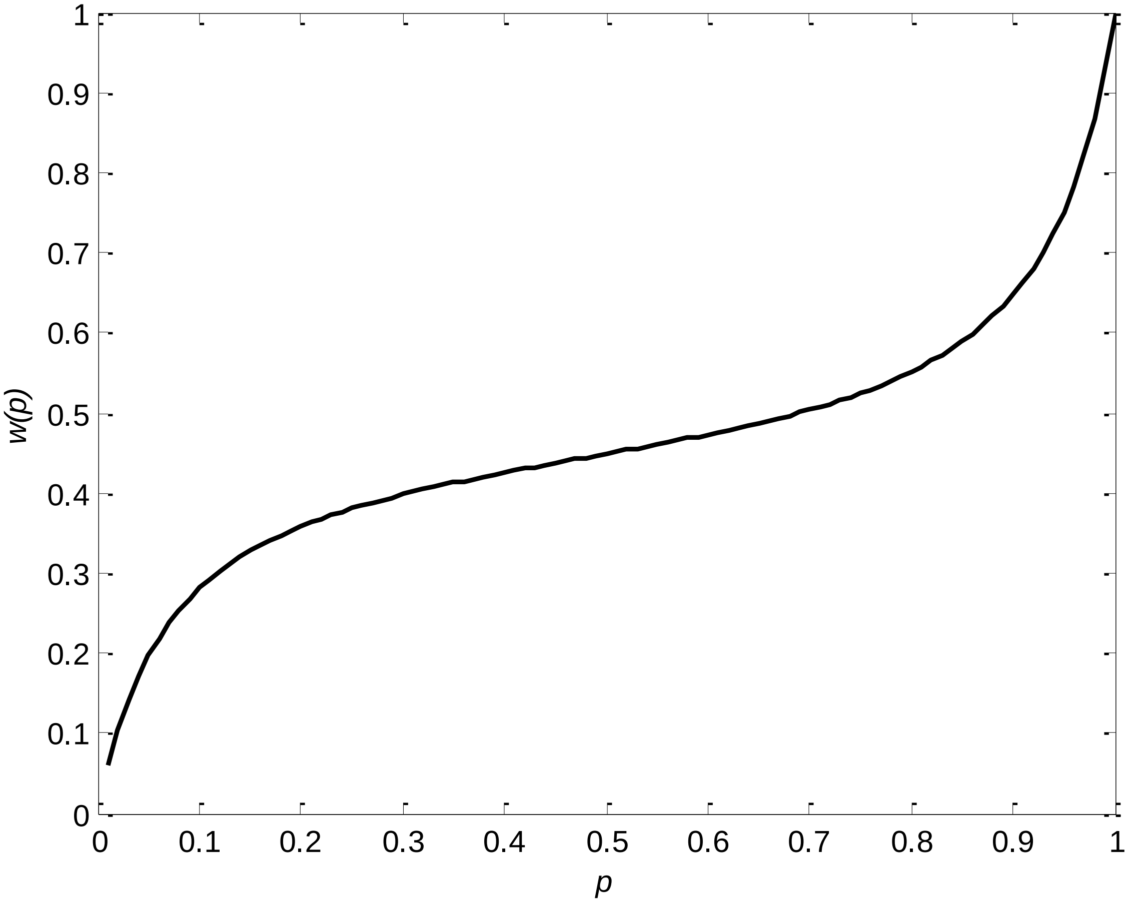
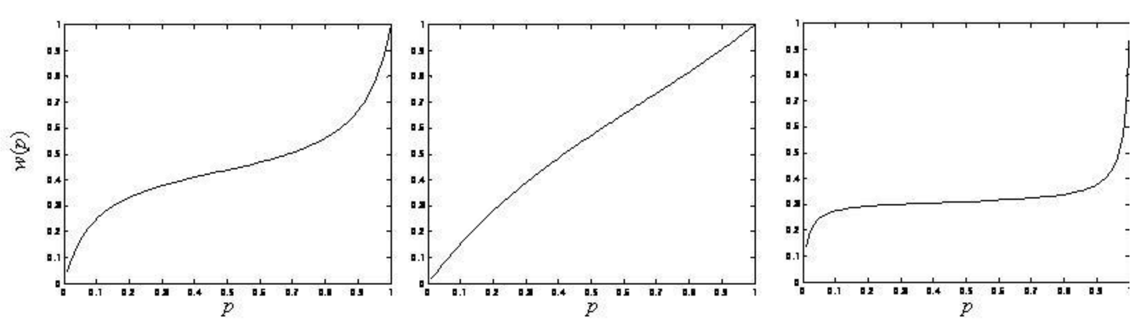


Figure 4

(a)



(b)



Notes

[1] To minimize the number of free parameters, we require the initial state probabilities to be 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 middle outcome. A wide variety of initial distributions can then be generated using a single initial state parameter, z.

[2] 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 $p = 0.5$ for two-outcome gambles. However, this does not necessarily require the "fixed point" of the model to equal 0.5 (all $p < 0.5$ are overweighted and all $p > 0.5$ are underweighted).

[3] 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).

[4] Starmer and Sugden (1993) also studied juxtaposition effects, however these studies also involved event splitting, and so the results are more closely related to the violations of stochastic dominance described earlier.

[5] 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.

[6] 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; β = 0.38; and α = 0.15.