Survey of Decision Field Theory

Jerome R. Busemeyer
Indiana University, USA
&
Adele Diederich
Universitat Oldenburg, Germany

September 27, 2000

Send all correspondence to:

Jerome R. Busemeyer
Indiana University
Psychology Department
Bloomington, In 47405
812-855-4882
jbusemey@indiana.edu

Abstract

This article summarizes the cumulative progress of a cognitive - dynamical approach to decision making and preferential choice called decision field theory. This review includes applications to (a) binary decisions among risky and uncertain actions, (b) multi-attribute preferential choice, (c) multi-alternative preferential choice, and (d) certainty equivalents such as prices. The theory provides natural explanations for violations of choice principles including strong stochastic transitivity, independence of irrelevant alternatives, and regularity. The theory also accounts for the relation between choice and decision time, preference reversals between choice and certainty equivalents, and preference reversals under time pressure. Comparisons with other dynamic models of decision-making and other random utility models of preference are discussed.

Survey of Decision Field Theory

Decision field theory (DFT) is a dynamic - cognitive approach to human decision-making based on psychological rather than economic principles. The decision processes posited by DFT have been applied successfully across a broad range of cognitive tasks including sensory detection (Smith, 1993), perceptual discrimination (Link, 1992), memory recognition (Ratcliff, 1978), conceptual categorization (Ashby, 2000; Nosofsky & Palmeri,1995), and preferential choice (Aschenbrenner, Albert, & Schmalhoffer, 1984).

DFT was initially proposed as a deterministic-dynamic model of approach - avoidance conflict behavior (Townsend & Busemeyer, 1989). Soon afterwards, the theory was recast as a stochastic-dynamic model of decision-making behavior (Busemeyer & Townsend, 1992). DFT was originally applied to binary choices for decision-making under uncertainty (Busemeyer & Townsend, 1993). Subsequently, the theory was applied to selling prices, certainty equivalents, and probability equivalents of gambles (Townsend & Busemeyer, 1995). Next, the theory was extended to account for multi-attribute decision-making (Diederich, 1997). Most recently, the theory was generalized to accommodate multiple (more than two) choice options (Roe, Busemeyer, & Townsend, in press). The purpose of the present article is to summarize these developments in one place using a common terminology and notation.

I. Model of the Task

It will be helpful to introduce the theoretical concepts through the use of a simplified decision problem entailing a choice among three options characterized by two attributes and two uncertain states of nature. Note, however, that the theory is generally applicable to an arbitrary number of alternatives, attributes, and states.

The table below represents a hypothetical choice among three medical insurance options (denoted A,B,C) provided by a university to its staff and faculty. These options differ on two major *attributes* (Cost of health care and Quality of health care), and the values of the attributes depend on two states of nature. The cells of the table indicate the hypothetical subjective values of each option on each attribute under each state of nature. For example, C_{BY} represents the Cost of program B under state of nature Y.

<Insert Table 1 here>

Important decisions, such as the one above, usually require time for deliberation. The institution arranges a procedure, called a *stopping rule*, which constrains when and how the decision is made. There are at least two stopping rules that may be used. One is a *fixed stopping time* in which the institution selects the time to make the decision. For example, the decision maker may be assigned a date and time to come into an office and sign a contract. Another procedure is an *optional stopping time* in which the decision maker selects the time to make the decision. For example, the decision maker may mail the contract to the university any time before the end of the year.

II. Model of Preference Evolution

DFT provides a formal description of the dynamic evolution of preferences during deliberation. Deliberation starts at time $t = 0$ with the initial presentation of information about the choice options, and continues during the interval $(0 < t < T_D]$ until the time T_D at which point the final decision is reached. The main task of DFT is to predict which option will be chosen and the time it takes to make the choice.

Dynamic Assumptions. At any time point during deliberation, each option has a strength of preference, denoted $P_i(t)$, corresponding to option i at time t. The row vector, $\boldsymbol{P}(t)' = [P_1(t), P_2(t), P_3(t)]$ represents the *preference state* at time t for the three options in the above example. (Note that $\boldsymbol{P}(t)$ denotes the column vector corresponding to the row vector $\boldsymbol{P}(t)'$) . The preference state is updated from one moment, t, to the next moment, $t+h$, according to the following linear difference equation:

$$\boldsymbol{P}(t+h) = \boldsymbol{S}\,\boldsymbol{P}(t) + \boldsymbol{V}(t+h). \qquad (1)$$

In Equation 1, h is an arbitrarily small time unit, and the preference state, $\boldsymbol{P}(t+h)$, approximates a diffusion process in the limit as $h \rightarrow 0$. The matrix $\boldsymbol{S}$ has the form $\boldsymbol{S} = (\boldsymbol{I} - h\boldsymbol{\Gamma})$, and it is assumed to be symmetric ($\gamma_{ij} = \gamma_{ji}$ for all i,j), and the diagonal values are assumed to be equal ($\gamma_{ii} = \gamma$ for all i). The diagonal elements of $\boldsymbol{S}$ provide memory for previous states of the system. The off-diagonal values of $\boldsymbol{S}$ allow for competitive interactions among competing alternatives. Furthermore, to ensure system stability, the eigenvalues of $\boldsymbol{S}$ are assumed to be less than one in magnitude. The input to this dynamic system, $\boldsymbol{V}(t)$, is called the valence vector, which is described next.

At each time point, the anticipated value of an option on an attribute is compared with the anticipated values of other options on the same attribute. These comparisons produce a valence for each option at each time point, denoted $v_i(t)$, for the valence of option i at time t. The row vector, $\boldsymbol{V}(t)' = [v_1(t), v_2(t), v_3(t)]$ represents the *valence input* at time t for the three options in the above example. (Note that $\boldsymbol{V}(t)$ denotes the column vector corresponding to the row vector $\boldsymbol{V}(t)'$). The valence vector is composed of a product of three matrices (described separately below):

$$\boldsymbol{V}(t) = \boldsymbol{C}\,\boldsymbol{M}\,\boldsymbol{W}(t) \qquad (2)$$

The purpose of the contrast matrix, ***C***, is to compare the weighted evaluations of each option produced by the product ***MW(t)***. More formally, the elements of the matrix ***C*** are defined as $c_{ii} = 1$ and $c_{ij} = -1/(n-1)$ for $i \neq j$ where n = number of options. For example, with three options $\mathbf{C} = \begin{bmatrix} 1 & -.5 & -.5 \\ -.5 & 1 & -.5 \\ -.5 & -.5 & 1 \end{bmatrix}$, and for two options $\mathbf{C} = \begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix}$.

Note that this definition of ***C*** guarantees that ***V(t)*** always sums to zero.

The value matrix ***M*** represents all the possible evaluations of each option on each attribute under each state of nature. Each element, M_{ij}, represents the value of option *i* under a particular state and attribute combination *j*. For the medical insurance example, ***M*** is identified with Table 1, and thus ***M*** is a 3 x 4 matrix. In particular, $M_{23} = C_{BY}$, that is the Cost of program B under state of nature Y.

The weight vector ***W(t)*** contains a weight corresponding to each column of ***M***. Each weight $w_j(t)$ corresponds to the joint effect of the importance of an attribute and the probability of a state. In the medical insurance example, ***W(t)*** is a 4 x 1 vector, where the first two weights correspond to the Cost and Quality under state X, and the last two weights correspond to the Cost and Quality under state Y. The matrix product, ***MW(t)***, can be interpreted as a vector containing the weighted average values of each option.

The cumulative effects of the valences on the evolution of preference over time can be seen more clearly by expanding Equation 1 (setting $t = nh$):

$$\boldsymbol{P}(t) = \boldsymbol{P}(nh) = \boldsymbol{\Sigma}_{j=0,\, n-1}\ \boldsymbol{S}^{j}\ \boldsymbol{V}(nh-jh) + \boldsymbol{S}^{n}\ \boldsymbol{P}(0). \quad (3)$$

This equation shows that the current preference state can be viewed as a weighted sum of the previous input valences. The weight placed on each previous input is determined by the feedback matrix raised to a power, where the power equals the lag between the current state and the previous input. The system is stable when the eigenvalues of the feedback matrix are restricted to be less than one in magnitude. In this case, the effect of the feedback matrix decays toward zero as the lag increases in value.

Distributional Assumptions. An important assumption of DFT is that the weight vector $\boldsymbol{W(t)}$ changes over time according to a stationary stochastic process. Psychologically, the weights represent fluctuations in the decision maker's attention to attributes and states over time. Diederich (1997) used a Markov process to model the stochastic changes in weights over time. Roe et al. (in press) simply assumed that the weights are identically and independently distributed (*iid*) over time. For the time being, we will maintain the *iid* assumption, but later in this article we will examine the Markov assumption.

If the weights are stochastic, then the valence vector and the preference state are also stochastic. The valence vector defined by Equation 2 is a linear transformation of the stochastic weight vector, $\boldsymbol{W(t)}$. The stationarity assumption for the weights implies that the mean vector, $\mathrm{E}[\boldsymbol{W(t)}] = \boldsymbol{w}\cdot h$, and covariance matrix $Cov[\boldsymbol{W(t)}] = \mathrm{E}[\,\boldsymbol{(W(t)-w)(W(t)-w)'}]$ $= h\cdot\boldsymbol{\Psi}$ are constant across time. (The dependence of the moments on the time unit, h, is required for the approximation to the continuous time process as $h \rightarrow 0$).

Stationarity of the weights also implies that the valence vector is a stationary stochastic process with mean

$$E[\boldsymbol{V(t)}] = E[\boldsymbol{CMW(t)}] = \boldsymbol{CM}\,\mathrm{E}[\boldsymbol{W(t)}] = \boldsymbol{CM(w\cdot}h\boldsymbol{)} = \boldsymbol{\mu}\cdot h\,,$$

and covariance matrix

$$Cov[\boldsymbol{V(t)}] = Cov\,[\,\boldsymbol{CMW(t)}\,] = \boldsymbol{CM}Cov[\boldsymbol{W(t)}]\,\boldsymbol{M'C'} = h\boldsymbol{CM\Psi M'C'} = h\boldsymbol{\Phi}\,.$$

The mean preference state is then obtained by taking the expectation of Equation 3:

$$\boldsymbol{\xi}(t) = E[\, \boldsymbol{P}(t)\,] = (\boldsymbol{I} - \boldsymbol{S})^{-1}(\boldsymbol{I} - \boldsymbol{S}^{n})\, \boldsymbol{\mu}\cdot h + \boldsymbol{S}^{n}\, \boldsymbol{P}(0)\ .$$

For stable systems, as $t \rightarrow \infty$, then $\boldsymbol{\xi}(t) \rightarrow (\boldsymbol{I} - \boldsymbol{S})^{-1}\, \boldsymbol{\mu}\cdot h = \boldsymbol{\xi}(\infty)$, which is a simple formula for analyzing the effects of the feedback matrix on the asymptotic mean preference state. Assuming *iid* weights, then the covariance matrix of the preference state evolves over time according to:

$$\boldsymbol{\Omega}(t) = Cov[\boldsymbol{P}(t)] = \mathrm{E}\{(\boldsymbol{P}(t) - \mathrm{E}[\boldsymbol{P}(t)])(\boldsymbol{P}(t) - \mathrm{E}[\boldsymbol{P}(t)])'\} = \boldsymbol{\Sigma}_{j = 0,\, n-1}\, \boldsymbol{S}^{j}\, (h\boldsymbol{\Phi})\, \boldsymbol{S}^{j\,'}\ .$$

As $j \rightarrow \infty$, $\boldsymbol{S}^{j} \rightarrow \boldsymbol{0}$, and so $\boldsymbol{\Omega}(t) \rightarrow \boldsymbol{\Omega}(\infty)$, the asymptotic covariance matrix (Anderson, 1971, p. 181).

DFT does not make any specific assumptions about the form of the distribution of the valence vector $\boldsymbol{V}(t)$. Note, however, that if the attentions weights $\boldsymbol{W}(t)$ are *iid*, then the valence $\boldsymbol{V}(t)$ is statistically independent of the past, $\boldsymbol{V}(t-kh)$ for all $k>0$, and therefore the preference state, $\boldsymbol{P}(t)$, is the sum of independent increments. According to the multivariate central limit theorem, the sum of independent increments converges in distribution to the multivariate normal or Gaussian distribution (Anderson, 1971, p. 429). Given that the preference state is a sum of independent increments, we conclude that $\boldsymbol{P}(t)$ converges in distribution to a Gaussian distribution, for any independent valence distribution. To approximate a diffusion process, we assume that h is a relatively small time unit, which implies that the convergence of $\boldsymbol{P}(t)$ to Gaussian occurs fairly rapidly.

Parameters. At this point, the model for preference evolution entails the following parameters. Like standard weighted additive utility models (e.g., Keeney & Raiffa, 1976), DFT requires specifying the value matrix $\boldsymbol{M}$ and the mean of the weight vector $\boldsymbol{w}$. Like the general Thurstone model (see Bockenholt, 1992; De Soute, Feger, & Klauer, 1989), DFT also requires specifying a covariance matrix, and in our case we specify a covariance matrix for the attention weights, $\boldsymbol{\Psi}$. The only new parameters introduced at this point are contained in the symmetric feedback matrix $\boldsymbol{S}$ of for linear dynamic system, and the initial preference state, $\boldsymbol{P}(0)$. But the latter is set to $\boldsymbol{P}(0) = \boldsymbol{0}$ in most applications, leaving only those in $\boldsymbol{S}$ as the new parameters. These new parameters are justified when one considers the increase in power gained by predicting new choice probabilities at each deliberation time point, using the *same* parameters across all time points. If a new set of choice probabilities can be observed at each successive time point, then this provides additional degrees of freedom for testing the model.

III. Choice Model for Fixed Stopping Time Tasks.

For the fixed stopping time task, the stopping time T_D is predetermined, and so the preference state is assumed to evolve in an unconstrained manner until the fixed appointment time T_D is reached. At the appointment time, the option with the greatest preference value is chosen. For a fixed set of choice options, DFT can be interpreted as a generalization of the Thurstone choice model in which the mean and covariance matrix evolve dynamically over time. As noted above, the distribution of $\boldsymbol{P}(t)$ will be approximately Gaussian with mean $\boldsymbol{\xi}(t)$ and variance-covariance matrix $\boldsymbol{\Omega}(t)$.

In general, the probability of choosing option i out of a set of n options at a fixed time T_D equals the $\Pr[P_i(T_D) = \max(P_j(T_D), j = 1,\ldots, n)]$. In the case of a binary choice, the probability of choosing A over B at a fixed time T_D is:

$$\Pr[A \mid \{A,B\} \text{ at } T_D] = \Pr[\, P_1(T_D) - P_2(T_D) > 0 \,]$$

$$= \int_{x>0} exp[-(x_A - \delta_A)^2 / 2\lambda_A \,] / (2\pi\lambda_A)^{.5}\, dx \qquad (4a)$$

where $x_A = P_1(T_D)-P_2(T_D)$, $\delta_A = \xi_1(T_D)-\xi_2(T_D)$, and $\lambda_A = \varphi_{11}(T_D) + \varphi_{22}(T_D) - 2\varphi_{21}(T_D)$. For trinary choice, the probability of choosing A from {A,B,C} at fixed time point T_D is:

$$\Pr[\,A \mid \{A,B,C\} \text{ at } T_D] = \Pr[P_1(T_D) - P_2(T_D) > 0 \;\&\; P_1(T_D) - P_3(T_D) > 0\,]$$

$$= \int_{X>0} exp[-\,(\boldsymbol{X_A}-\boldsymbol{\delta_A})'\,\boldsymbol{\Lambda_A}^{-1}\,(\boldsymbol{X_A}-\boldsymbol{\delta_A})/2\,] \,/\, (2\pi\,|\,\boldsymbol{\Lambda_A}\,|^{.5})\, d\boldsymbol{X_A}\ , \tag{4b}$$

where $\boldsymbol{X_A} = \boldsymbol{L_A P(t)}$, $\boldsymbol{\delta_A} = \boldsymbol{L_A}\xi(T_D)$, $\boldsymbol{\Lambda_A} = \boldsymbol{L_A \Omega}(T_D)\boldsymbol{L_A}'$ and $\mathbf{L}_A = \begin{bmatrix} 1 & -1 & 0 \\ 1 & 0 & -1 \end{bmatrix}$.

One of the main contributions of DFT is to predict how choice probabilities change as a function of deliberation time. Static probabilistic choice models, such as the simple scalable class of models or standard random utility class of models, are unable to do this. However, it is worthwhile to compare the asymptotic predictions of DFT to the predictions generated by the static models. Based on the central limit theorem, as $T_D \rightarrow \infty$, the probability distribution of the preference state converges toward an equilibrium distribution that is Gaussian with mean $\xi(\infty)$ and covariance matrix $\boldsymbol{\Omega}(\infty)$. Two important behavioral properties of the asymptotic choice probabilities produced by DFT are described next.

Independence. According to the simple scalable class of choice models, the probability of choosing A_i from a set $X = \{A_1, \ldots, A_n\}$ is given by

$$\Pr[A_i \mid X] = F_X(i, u(A_1), \ldots, u(A_n)) \tag{5}$$

where the function F_X is increasing in the *i*-th variable and decreasing in each of the remaining variables. A scale value $u(A_i)$ is assigned to each option *i* independent of the choice set in which it appears. For example, the Luce (1959) choice model is a member of this class. The simple scalable class of models implies a behavioral property called *independence from irrelevant alternatives* (Tversky, 1972):

$$\Pr[\,A \mid \{A,B\}\,] \geq \Pr[\,B \mid \{A,B\}] \leftrightarrow \Pr[\,A \mid X\,] \geq \Pr[\,B \mid X].$$

This follows from the fact that the left hand side implies $u(\mathrm{A}) \geq u(\mathrm{B})$, which in turn implies the right hand side. In particular, if Pr[A | {A,B}] = .50, then $u(\mathrm{A}) = u(\mathrm{B})$, which implies Pr[A | X] = Pr[B | X].

DFT does *not* obey the above property. According to DFT, violations are caused by changes in the covariance matrix $\boldsymbol{\Omega}(\infty)$ across choice sets (see Roe et al, in press). For a simple example, suppose $\delta_A = \delta_B$ so that the model predicts Pr[A|{A,B}] = .50 in the binary choice. For the trinary choice set, we still have $\boldsymbol{\delta}_A = \boldsymbol{\delta}_B$, but if $\varphi_{AC} \neq \varphi_{BC}$ or $\varphi_{AA} \neq \varphi_{BB}$ in $\boldsymbol{\Omega}(\infty)$, then $\boldsymbol{\Lambda}_A \neq \boldsymbol{\Lambda}_B$, and therefore Pr[A|{A,B,C}] ≠ Pr[B|{A,B,C}], violating the independence property. Empirical research on human preferential choice has demonstrated systematic and robust violations of the independence property, thus ruling out the simple scalable class of models (Tversky, 1972). See Roe et al. (in press) for more details concerning the application of DFT to violations of independence.

Regularity. According to standard random utility models, the probability of choosing A_i from a set $X = \{A_1,\ldots,A_n\}$ is given by

$$\Pr[A_i \mid X] = \Pr[\ U_i = \max(U_j\,,\ j \varepsilon X)], \tag{6}$$

where $\boldsymbol{U_n}' = [U_1, \ldots, U_n]$ is a vector of random utilities. The standard random utility model assumes that the distribution of a utility vector $\boldsymbol{U}_m$, $m < n$, for a set $Y \subset X$ is derived from the distribution of $\boldsymbol{U}_n$ by integrating out the random utilities of the options in X and not in Y. For example, the Thurstone choice model (Bock & Jones, 1968; De Soute. et al., 1989) is a member of this class. The standard class of random utility models imply a behavioral property called *regularity* (see Fishburn, 1998):

$$\Pr[\ A_i \mid Y] \geq \Pr[\ A_i \mid X\ ].$$

This follows from the fact that $\Pr[U_i = \max(U_j, j \varepsilon Y)] \geq \Pr[U_i = \max(U_j, j \varepsilon Y)] \times$ $\Pr[U_i = \max(U_j, j \varepsilon X\text{-}Y) \mid U_i = \max(U_j, j \varepsilon Y)] = \Pr[U_i = \max(U_j, j \varepsilon X)]$.

DFT does *not* obey the principle of regularity. According to DFT, violations of regularity are caused by changes in the feedback matrix ***S*** across choice sets. For example, suppose once again that $\mu_A = \mu_B = c$, so that the model predicts Pr[A | {A,B}] = .50 in the binary choice. For the trinary choice, suppose we set

$$\mathbf{S} = \begin{bmatrix} 1-h\gamma_1 & 0 & -h\gamma_2 \\ 0 & 1-h\gamma_1 & 0 \\ -h\gamma_2 & 0 & 1-h\gamma_1 \end{bmatrix} = \begin{bmatrix} a & 0 & -b \\ 0 & a & 0 \\ -b & 0 & a \end{bmatrix},$$

This type of feedback matrix would be appropriate for a choice set in which options A and C are highly similar to each other, but option B is dissimilar to both A and C. The eigenvalues of ***S*** in this case are (a, a-b, a+b), which are required to be less than unity to maintain stability. In addition, setting $\mu_C = -2c$ yields:

$$\xi(\infty) = (\mathbf{I} - \mathbf{S})^{-1}\boldsymbol{\mu} = \begin{bmatrix} \dfrac{(1-a+2b)c}{(1-a-b)(1-a+b)} \\ \dfrac{c}{1-a} \\ \dfrac{(-2+2a-b)c}{(1-a-b)(1-a+b)} \end{bmatrix}$$

Under these conditions, ξ_1 and ξ_2 are both positive, and ξ_3 is negative. Furthermore, the difference in means is positive: $\xi_1 - \xi_2 = \dfrac{cb(2(1-a)+b)}{(1-a)(1-a-b)(1-a+b)}$.

As c increases, the probability of choosing option C goes to zero, while the difference between options A and B increases, driving the probability of choosing option A toward 1.0, which violates regularity. Empirical research on human preferential choice has demonstrated systematic and robust violations of regularity (Huber, Payne, Puto, 1982), thus ruling out the standard random utility models.

It is also interesting to compare the asymptotic predictions of DFT with the well-known eliminations by aspects (EBA) model of choice (Tversky, 1972). The EBA model was originally developed to account for violations of independence from irrelevant alternatives. However, the EBA model obeys the regularity property (Tversky, 1972), and thus fails to account for the findings of Huber et al. (1982). Only DFT is capable of explaining both violations of behavioral properties. See Roe et al. (in press) for more details on the application of decision field theory to violations of regularity.

IV. Choice Model for Optional Stopping Time Tasks.

For the optional stopping time task, the decision maker determines when to stop and make a decision. Thus T_D is a random stopping time rather than a fixed stopping time. According to DFT, in this case the preference state continues to evolve over time until the preference strength for one of the options becomes strong enough to exceed a threshold bound. The first alternative to cross the bound is then chosen.

More formally, we assume that deliberation continues as long as $P_i(t) < \theta$ for all options $i \in X$. The process stops and chooses option i at time T_D if $P_i(T_D) \geq \theta$, while $P_j(t) < \theta$ for all $j \in (X\text{-}i)$ and $t < T_D$. Technically, the deliberation process is represented by a multidimensional diffusion process, $\boldsymbol{P}(t)$. Choice probability is determined from the probability of absorption at the threshold bound, and the distribution of stopping times is determined from the first passage time distribution of the diffusion process to the bound. Below we show how to compute the absorption probabilities and mean stopping times for binary choice problems. Later we extend the model for trinary choice with optional stopping.

First consider a binary choice between two options A and B. In this case, the preference state is represented by a two-dimensional diffusion process:

$$\begin{bmatrix} P_1(t+h) \\ P_2(t+h) \end{bmatrix} = \begin{bmatrix} 1-h\gamma_1 & -h\gamma_2 \\ -h\gamma_2 & 1-h\gamma_1 \end{bmatrix} \cdot \begin{bmatrix} P_1(t) \\ P_2(t) \end{bmatrix} + \begin{bmatrix} V_1(t+h) \\ V_2(t+h) \end{bmatrix}, \quad (7a)$$

where [$P_1(t)$, $P_2(t)$] represent the preference states for options A and B, respectively. Note that if $P_1(0) + P_2(0) = 0$ then $P_1(t) + P_2(t) = 0$ for all t, so that the evolution of the preference state can be reduced to one variable, say $P_1(t)$, representing the preference state for option A:

$$P_1(t+h) = [1-h\cdot(\gamma_1-\gamma_2)]\cdot P_1(t) + V_1(t+h)\,. \quad (7b)$$

So for binary choices, the diffusion process is restricted to a line, and the decision process stops and chooses A if $P_1(t) > \theta$, it stops and chooses B if $P_2(t) = -P_1(t) > \theta$, otherwise it continues deliberating. Equation (7b) can be re-arranged into a stochastic difference equation as follows:

$$dP(t+h) = [P_1(t+h) - P_1(t)] = V_1(t+h) - h\cdot(\gamma_1-\gamma_2)\cdot P_1(t) \quad (7c)$$

The mean of the change in preference favoring option A equals

$$\mu(x)h = \mathrm{E}[dP(t+h) \mid P_1(t) = x] = [\mu - (\gamma_1-\gamma_2)\cdot x]\cdot h \quad (8a)$$

and the variance of the change equals

$$h\phi_{11} = Var[dP(t+h) \mid P_1(t) = x] = h\,\mathrm{E}[\,V_1(t+h) - \mu(x)\,]^2 \quad (8b)$$

As $h \rightarrow 0$, the preference state process converges in distribution to a diffusion process known as the Ornstein-Uhlenbeck (OU) process (Bhattacharya & Waymire, 1990, Ch.

V). Equation 8a defines the drift rate, $\mu(x)$, and Equation 8b defines the diffusion rate of the OU process, ϕ_{11}. (For convenience, define $\phi = \sqrt{\phi_{11}}$ as the standard deviation.)

The solutions for the absorption probabilities and the mean decision times for each alternative can be derived from the Kolmogorov backward equation for the OU process (see Busemeyer & Townsend, 1992, p. 271). An alternative method for computing these statistics is based on Markov chain methods. The latter turns out to be more practical when dealing with generalizations to more complex models (Bhattacharya & Waymire, 1990, p. 389-390). Therefore, the Markov chain method is presented below.

To set up the Markov chain representation for the binary choice problem, we approximate the stochastic change in preference, $dP(t+h)$, by a random increment, $\Delta(t+h)$, which is either a small positive step $\Delta = \alpha\phi\sqrt{h}$ or a small negative step $-\Delta = -\alpha\phi\sqrt{h}$, or zero. The state space of the Markov chain is defined by a set of $m = 2k+1$ equally spaced discrete states: $\chi = \{ i\Delta,\ i = -k,\ldots,k\}$. The two threshold bounds, $-\theta = -k\Delta$ and $+\theta = +k\Delta$, represent the two absorbing states of the Markov chain, and the remaining (m-2) intermediate states represent transient states. The transition probabilities for the transient states are symbolized as

$$\Pr[\ \Delta(t+h) = +\Delta \mid P_1(t) = x] = q_1(x),$$

$$\Pr[\ \Delta(t+h) = -\Delta \mid P_1(t) = x] = q_2(x),$$

$$\Pr[\ \Delta(t+h) = 0 \mid P_1(t) = x] = q_3(x),$$

with the constraint that $q_1(x)+q_2(x)+q_3(x) = 1$ for $x \in \chi$. The transition probabilities are defined by equating the moments of the discrete and continuous state processes:

$$E[\ \Delta(t+h) \mid P(t) = x]/h = [q_1(x) - q_2(x)]\cdot\Delta/h = \mu(x) \tag{9a}$$

$$Var[\Delta(t+h) \mid P(t) = x]/h = [q_1(x) + q_2(x)]\cdot\Delta^2/h = \phi_{11}^2 - \mu(x)^2 h\ . \tag{9b}$$

As $h \rightarrow 0$, then $\mu(x)^2 h \rightarrow 0$, and so the second order term in Equation 9b can be ignored for small h. The solution to these two equations (ignoring the second term) yields the pair of transition probabilities (with $\phi^2 = \phi_{11}$):

$$q_1(x) = [\,1 + \sqrt{h}\cdot \mu(x)/\phi\,] / 2\alpha, \tag{10a}$$

$$q_2(x) = [\,1 - \sqrt{h}\cdot \mu(x)/\phi\,] / 2\alpha\,, \tag{10b}$$

$$q_3(x) = 1 - (1/\alpha)\,. \tag{10c}$$

As $h \rightarrow 0$, the above Markov chain process converges in probability distribution to the OU diffusion process (Bhattacharya & Waymire, 1990, Theorem 4.1, p. 387). The parameter, $\alpha > 1$, is a free parameter that has no effect on the choice probabilities, but it improves the approximation to a continuous distribution of decision times.

The transition probabilities for the m states of the Markov chain can be arranged into an $m \times m$ transition matrix, $\boldsymbol{T}$, with elements $T_{11} = 1$, $T_{mm} = 1$ and for $1 < i < m$:

$$T_{ij} = \begin{cases} q_1(\Delta i) \text{ if } j = i+1 \\ q_2(\Delta i) \text{ if } j = i-1 \\ q_3(\Delta i) \text{ if } j = i\ . \end{cases} \tag{11}$$

The transition matrix can be partitioned into three matrices. Define $\boldsymbol{R_A}$ as a $(m-2)$ column vector formed by rows i ε {2,3,…,(m-1)} in the last column of $\boldsymbol{T}$. This contains the transition probabilities from a transient state to the absorption state for option A. Also define $\boldsymbol{R_B}$ as a $(m-2)$ column vector formed by rows i ε {2,3,…,(m-1)} in the first column of $\boldsymbol{T}$. This contains the transition probabilities from a transient state to the absorption state for option B. Finally define $\boldsymbol{Q}$ as the $(m-2)\times(m-2)$ inner matrix containing rows i ε {2,3,…,(m-1)} and columns j ε {2,3,…,(m-1)} of $\boldsymbol{T}$. This contains the transition probabilities from one transient state to another. The initial probability distribution (at t =

0) over the transient states is represented by a (m-2) column vector denoted $\boldsymbol{Z}$. The probability of choosing option i at time $t = nh$ is given by (Bhattacharya & Waymire, 1990, p. 242):

$$\Pr[\text{ Choose } i \text{ at time } t] = \boldsymbol{Z'Q^n R_i} . \qquad (12a)$$

Note that $\boldsymbol{Q}$ is a tridiagonal matrix that is guaranteed to have m-2 linearly independent eigenvectors and m-2 real eigenvalues (see Bhattacharya & Waymire, 1990, p. 242). Therefore, spectral analytic methods can be used to efficiently compute the matrix product $\boldsymbol{Q^n}$ (see Bhattacharya & Waymire, 1990, p. 242).

Standard matrix methods can then be used to solve for the absorption probabilities and mean time to absorb at each boundary (see Cox & Miller, 1965, Ch. 3). The overall probability of choosing option i (i = A, B) can be computed from the matrix products

$$\Pr[\text{ Choose } i] = \boldsymbol{Z'(I-Q)^{-1}R_i}. \qquad (12b)$$

The conditional mean time to choose option i is also given by the matrix products:

$$E[\, T_D \mid \text{Choose } i] = \boldsymbol{Z'(I-Q)^{-2}R_i} / \Pr[\text{ Choose } i]. \qquad (12c)$$

The time unit h is chosen to be as small as necessary to achieve accurate approximation to the continuous diffusion process. Note that $m = 2k+1$, $k = \theta/\Delta$, and $\Delta = \alpha\,\phi\sqrt{h}$. Thus, for a fixed θ, m increases as h decreases. Fortunately, convergence occurs very rapidly for choice probabilities and mean decision times (see, e.g., Figure 1, and note that $k > 10$ works well. See Diederich, 1997, for a more comprehensive evaluation of convergence rate). These matrix computations can be performed quickly and accurately using sparse matrix methods (available for example in Matlab or Mathmatica or Gauss). Two important qualitative predictions derived from decision field theory are described next.

<Insert Figure 1 here>

Stochastic Transitivity. Simple scalable choice models obey another behavioral property of choice, called *strong stochastic transitivity*, which states:

If Pr[A | {A,B}] $\geq$ .5, and Pr[B | {B,C}]$\geq$.5, then

Pr[A | {A,C}] $\geq$ Pr[A | {A,B}], and Pr[A | {A,C}] $\geq$ Pr[B | {B,C}].

This follows from the fact that the first antecedent condition implies u(A) > u(B), and the second implies u(B) > u(C), and transitivity of real numbers implies u(A) > u(C).

DFT does *not* obey the strong stochastic transitivity property. Violations result from the fact that the transition probabilities in Equation 10 depend on the diffusion rate ϕ^2, which varies across choice pairs. Busemeyer and Townsend (1992) proved (see Proposition 2, p. 277) that the probability of choosing option A over B in Equation 12b is a strictly increasing function of the ratio [$\mu(x)/\phi$]. Thus increasing the diffusion rate, ϕ^2, holding $\mu(x)$ fixed, decreases Pr[A |{A,B}] in Equation 12b.

According to DFT, violations of transitivity occur under the following conditions. For simplicity, set (γ_1-γ_2) = 0 so that $\mu(x) = \mu$ for a particular choice pair, and also assume that the preference state begins at $P(0) = 0$. Define $\mu = \mu(A,B) = \mu(B,C) = \mu(A,C)/2 > 0$ as the drift rates for choice pairs {A,B}, {B,C}, and {A,C}, respectively. In this case, the mean valences are equally spaced apart for options A, B, and C. Also define $\phi^2 = \phi^2(A,C) = \phi^2(B,C) = \beta^2\phi^2(A,B)$ with $\beta > 2$ as the diffusion rates for choice pairs {A,B}, {B,C}, and {A,C}, respectively. These types of relations among the diffusion rates occur when the valences for options A and B are highly correlated, but the valences for option C are uncorrelated with either A or B. Then $[\mu(A,B)/\phi(A,B)] = \beta(\mu/\phi) > 0$, which implies Pr[A|A,B]>.5, and $[\mu(B,C)/\phi(B,C)] = (\mu/\phi) > 0$, which implies Pr[B|B,C]>.5. However, if $\beta > 2$, then $[\mu(A,B)/\phi(A,B)] = \beta(\mu/\phi) > 2(\mu/\phi) = \mu(A,C)/ \phi(A,C)$ which implies

Pr[A|{A,B}] > Pr[A|{A,C}], violating strong stochastic transitivity. Empirical research on preferential choice has demonstrated systematic violations of strong stochastic transitivity as predicted by the above analysis (Mellers & Biagini, 1994).

Speed Accuracy Trade Off Effects. The amount of time that a decision maker allocates to make a decision affects the probability that an option will be chosen (Vickers, Smith, & Brown, 1985). Important decisions require more time, increasing the likelihood of choosing the best option; whereas unimportant decisions are made more quickly but often result in a bad choice. Some decision makers tend to be more deliberative and take more time to decide, trying to make the right choice; whereas other decision makers tend to be more impulsive and make quick decisions, caring less about mistakes. One of the unique features of DFT is its ability to account for the relation between speed and accuracy in preferential choice and risky decision-making.

The key idea is that the threshold bound, θ, controls the stopping time. When an important decision is to be made, then this threshold is set to a large magnitude, thus requiring a very high level of preference strength before making a decision. When a unimportant decision is to be made, or when the decision maker is under time pressure to make a quick decision, then the threshold is set to a lower magnitude, allowing a weak preference to make the decision. Busemeyer & Townsend (1992) proved that if $P_I(0) = 0$, and $\mu > 0$ (favoring the choice of option A), then increasing θ in Equation 12 increases both the probability of choosing A, Pr[A | {A,B}], and the mean time, $E[T_D]$, to make a choice, in accordance with the speed – accuracy trade off principle. Thus, the threshold parameter provides a simple way to model the empirically observed relation between speed and accuracy of choice.

V. Multi-Attribute Model with Changing Attention Weights.

Previously it was assumed that the weights, $\boldsymbol{W(t)}$, changed from moment to moment according to an identitically and independently distributed process (*iid*). Diederich (1997) developed a *multi-attribute* version of DFT that allows the attention weights to change according to a Markov process over time. To be more specific, consider once again the health insurance decision, which involves two major attributes--Cost and Quality. Previously, we assumed that attention to these two attributes was represented by a single *iid* stochastic process denoted $\boldsymbol{W(t)}$. Now we assume that $\boldsymbol{W(t)}$ is a mixture of two sub-processes $\boldsymbol{W_1(t)}$ and $\boldsymbol{W_2(t)}$, which are individually *iid*. At any particular time during deliberation, the attention process may be operating on the basis of one of these sub-processes, say $\boldsymbol{W_1(t)}$. During the next moment, from time t to $t+h$, attention either continues to operate under process $\boldsymbol{W_1(t)}$ with probability π_{11}, or attention switches with probability $\pi_{12} = 1-\pi_{11}$ and starts operating on the basis of $\boldsymbol{W_2(t)}$. Similarly, if attention is operating on the basis of $\boldsymbol{W_2(t)}$ at time t, then during the next moment, attention may continue operating under $\boldsymbol{W_2(t)}$ with probability π_{22}, or switch to $\boldsymbol{W_1(t)}$ with probability $\pi_{21} = 1-\pi_{22}$. Thus attention switches from one attribute to another according to a Markov chain process.

The Markov attention-switching hypothesis is incorporated into DFT by extending the Markov chain representation of the binary choice model. Define $\boldsymbol{T_1}$ as a transition matrix, constructed as in Equations 10 and 11, but with the mean drift rate and diffusion rate determined from Equation 9 using the Cost attribute attention process $\boldsymbol{W_1(t)}$. Similarly, define $\boldsymbol{T_2}$ as a transition matrix with transition probabilities determined from the Quality attribute attention process $\boldsymbol{W_2(t)}$. Finally, define a $2m \times 2m$ extended transition matrix $\boldsymbol{T}$:

$$\mathbf{T} = \begin{bmatrix} \pi_{11}\mathbf{T}_1 & \pi_{12}\mathbf{I} \\ \pi_{21}\mathbf{I} & \pi_{22}\mathbf{T}_2 \end{bmatrix} \text{, where } \boldsymbol{I} \text{ is an } m \times m \text{ identity matrix.} \quad (13)$$

To compute the choice probabilities and mean response times, we partition the transition matrix into parts as follows. Using the transition matrix $\boldsymbol{T_1}$ for the Cost attribute, define $\boldsymbol{R_{1A}}$ as the as a (m-2) column vector formed by rows i ε {2,3,…,(m-1)} in the last column of $\boldsymbol{T_1}$, define $\boldsymbol{R_{1B}}$ as a (m-2) column vector formed by rows i ε {2,3,…,(m-1)} in the first column of $\boldsymbol{T_1}$, and finally define $\boldsymbol{Q_1}$ as the (m–2)×(m-2) inner matrix containing rows i ε {2,3,…,(m-1)} and columns j ε {2,3,…,(m–1)} of $\boldsymbol{T_1}$. Using the transition matrix $\boldsymbol{T_2}$ for the Quality attribute, similar definitions hold for $\boldsymbol{R_{2A}}$, $\boldsymbol{R_{2B}}$, and $\boldsymbol{Q_2}$. Inserting these component definitions into the extended matrices produces

$$\mathbf{R}_A = \begin{bmatrix} \pi_{11}\mathbf{R}_{1A} \\ \pi_{22}\mathbf{R}_{2A} \end{bmatrix}, \mathbf{R}_B = \begin{bmatrix} \pi_{11}\mathbf{R}_{1B} \\ \pi_{22}\mathbf{R}_{2B} \end{bmatrix}, \mathbf{Q} = \begin{bmatrix} \pi_{11}\mathbf{Q}_1 & \pi_{12}\mathbf{I} \\ \pi_{21}\mathbf{I} & \pi_{22}\mathbf{Q}_2 \end{bmatrix}, \quad (14)$$

where $\boldsymbol{I}$ is an (m-2) × (m-2) identity matrix. Finally, let $\boldsymbol{Z'} = [\, \boldsymbol{Z_C'}, \boldsymbol{Z_Q'}]$ be a 2(m-2) row vector representing the initial distribution over all of the transient states at time t=0. Inserting these definitions into Equation 12 provide the choice probabilities and mean decision times for multi-attribute DFT with Markov attention switching.

Applying Equation 12 directly to the transition matrices shown in Equation 14 doubles the size of the transition matrix, $\boldsymbol{Q}$, which must be inverted. Fortunately, the inverse of $\boldsymbol{Q}$ can be written in terms of the inverses of the component matrices as follows (see Searle, 1966, p. 210):

$$(\mathbf{I}-\mathbf{Q})^{-1} = \begin{bmatrix} \mathbf{H}^{-1} & -\pi_{12}\mathbf{H}^{-1}\mathbf{G}^{-1} \\ -\pi_{21}\mathbf{G}^{-1}\mathbf{H}^{-1} & \mathbf{G}^{-1}[\mathbf{I}+\pi_{12}\pi_{21}\mathbf{H}^{-1}\mathbf{G}^{-1}] \end{bmatrix},$$

$$\boldsymbol{F} = (\boldsymbol{I}\text{-}\boldsymbol{Q}_1),\ \ \boldsymbol{G} = (\boldsymbol{I}\text{-}\boldsymbol{Q}_2), \text{ and } \boldsymbol{H}^{1} = (\boldsymbol{F} - \pi_{12}\boldsymbol{G}^{-1})^{-1}.$$

Thus, the choice probability and mean response time predictions for the mulit-attribute model can be computed almost as efficiently as the predictions for the original DFT. Moreover, the principles described above can be extended to more than two attributes. For example, a three-attribute model entailing Cost, Quality, and Reputation would entail a $3m \times 3m$ extended transition matrix, composed of individual transition matrices $\boldsymbol{T}_1$, $\boldsymbol{T}_2$, and $\boldsymbol{T}_3$ (see Diederich, 1997, for applications to three attributes). The multi-attribute extension of DFT introduces a new attention transition parameter, π_{jj} , for each new attribute *j*. However, the additional parameters are justified by the increase in explanatory power needed to account for the following empirical findings.

Preference Reversals Under Time Pressure. Empirical research on decision making under time pressure has demonstrated that the difference in choice probabilities, Pr[A | {A,B}] – Pr[B | {B,A}], can change sign under short and long deliberation time limits, thus reversing preferences under time pressure (see Diederich, 1997; Svenson & Edlund, 1993). This tends to occur when the most important dimension (e.g., Cost) weakly favors one option, but the less important dimension (e.g., Quality) strongly favors the alternative option.

According to the multi-attribute DFT, attention is initially focused on the most important attribute, and during this time, the drift rate under $\boldsymbol{W}_C(t)$ drives the preference toward one option, say option A. Under short time pressure, the threshold bound, θ, is set to a low magnitude, only allowing time for the first dimension to be processed, which causes the probability of choosing option A to exceed .50. Under little or no time pressure, the threshold bound is set to a higher magnitude, allowing more processing time for the second attribute. During the time that the second attribute is processed, the drift rate under $\boldsymbol{W}_Q(t)$ drives the preference toward the alternative option, say option B. This causes the probability of choosing option B to exceed .50, thus reversing the choice probabilities under the two time pressure conditions. Figure 2 illustrates this preference reversal as a function time pressure. In this figure, choice probability (computed from Equation 12) is plotted as a function of the threshold bound $\theta = k\Delta$. Diederich (1997) reports a more comprehensive response surface analysis of the model predictions for this finding under a wider range of model parameters.

<Insert Figure 2 about here>

Choice Probability and Decision Time. An important empirical law of preferential choice behavior is the fact that the average time required to choose an option is inversely related to the probability of choosing it (Petrusic & Jamieson, 1978). More specifically, when given a choice between two options, A and B, if Pr[A|{A,B}] > Pr[B|{A,B}] then E[T_D | Choose A] < E[T_D | Choose B]. Thus response time as well as response probability provide alternative measures of preference strength. This basic empirical fact rules out all models predicting decision times to be independent of the choice that is made, which includes Marley (1989), Tversky (1972), and McFadden (1978).

Multi-attribute DFT provides a simple explanation for the dependence of decision time on the choice that is made. It is natural to assume that the attention process begins by focusing on the more important of two attributes (e.g., Cost), and later switches to the less important attribute (e.g., Quality). In this case, it is most likely that the preference state will be driven rapidly toward the threshold bound favoring the most important attribute, allowing this alternative to be chosen more quickly and more frequently. Less frequently, the preferences state may initially wander and fail to reach the threshold, and on these less frequent occurrences, attention will have time to switch to the less important attribute, driving the preference state toward the bound of the less frequently chosen alternative. Thus the popular option tends to get chosen during the early stages when attention is focused on the more important attribute, and the less popular option tends to get chosen during the later stages after attention has switched to the less important attribute.

Figure 3 illustrates the effect of the attention switching parameter on the relation between choice probability and choice time. The top curve shown in this figure shows Pr[A|{A,B}] plotted as a function of π_{ii}, the probability of not switching (in this case we set $\pi_{11} = \pi_{22}$). The bottom curve shows the difference in mean decision times, E[T_D|Choose B]-E[T_D|Choose A], also plotted as a function of π_{ii}. Note that when π_{11} = 1, then attention never switches away from the first dimension, and in this case there is no difference in mean time, but as π_{11} slowly decreases below 1.0, the mean difference increases in the direction of longer decision times for the less favored option. Ultimately, when attention is switching back and forth very rapidly, mixing the two processes together in an almost homogenous manner, then the difference in mean decision times disappears again. Diederich (1997) provides a more comprehensive response surface analysis of the model predictions for this finding under a wider range of model parameters.

<Insert Figure 3 about here>

V. Trinary Choice.

Earlier we presented the trinary choice model for a fixed stopping time procedure (see Equation 4b). Now we develop the trinary choice model for the optional stopping time task. Consider a choice from a set {A, B, C}. The preference state is represented by a three dimensional version of the diffusion process shown as Equation 1:

$$\boldsymbol{P}(t+h) = \boldsymbol{SP}(t) + \boldsymbol{V}(t+h) = (\boldsymbol{I} - h\boldsymbol{\Gamma})\boldsymbol{P}(t) + \boldsymbol{V}(t+h)$$

or in difference equation form

$$d\boldsymbol{P}(t+h) = \boldsymbol{P}(t+h)-\boldsymbol{P}(t) = -h\,\boldsymbol{\Gamma}\boldsymbol{P}(t) + \boldsymbol{V}(t+h).$$

The moments of the changes in preference state are given by

$$\boldsymbol{\mu}(\boldsymbol{X})h = \mathrm{E}[d\boldsymbol{P}(t+h) \mid \boldsymbol{P}(t)=\boldsymbol{X}\,] = (-\boldsymbol{\Gamma X}+\boldsymbol{\mu})h\ ,$$

$$h\boldsymbol{\Phi} = Cov[\ d\boldsymbol{P}(t+h) \mid \boldsymbol{P}(t)=\boldsymbol{X}\,]\ .$$

If the columns of the feedback matrix $\boldsymbol{S}$ sum to the same constant, and if the initial state vector, $\boldsymbol{P}(0)$ sums to zero, then $P_3(t) = -[\ P_1(t) + P_2(t)\ ]$. In this case, option A is chosen as soon as $P_1(t) \geq \theta$, option B is chosen as soon as $P_2(t) \geq \theta$, and option C is chosen as soon as $P_3(t) = -[P_1(t)+P_2(t)] \geq \theta$. Thus, the diffusion process is restricted to a two - dimensional simplex (triangular plane). The following developments are limited to diffusions on this simplex.

Markov chains can be used to approximate multidimensional diffusion processes (see Stroock & Varadhan, 1979, Theorem 11.2.3). In the case of the diffusion on the simplex, we approximate the stochastic change in preference, $d\boldsymbol{P}(t+h)$, by a random increment, $[\Delta_1(t+h), \Delta_2(t+h), \Delta_3(t+h)]$, and we require that $\Delta_1(t+h)+\Delta_2(t+h)+\Delta_3(t+h) = 0$, so that only the first two coordinates need to be included in the model. Each coordinate, $\Delta_i(t+h)$, is a small positive step $\Delta = \alpha\phi\sqrt{h}$ or a small negative step $-\Delta = -\alpha\phi\sqrt{h}$, or zero; and we set $\phi = \sqrt{\phi_{11} + \phi_{22}}$ producing a step size of equal magnitude for each coordinate. The state space of the Markov chain is defined by a set of $m = (3k+2)(3k+1)/2$ equally spaced discrete states that form a grid of points on the simplex:

$$\chi = \{ [i\Delta, j\Delta], (i = -(k+j),\ldots,k), (j = -2k,\ldots, k) \}$$

The threshold bound for option A is defined by the right vertical line of points $[\theta = k\Delta, j\Delta]$ for $j = -2k,\ldots,k$; the threshold bound for option B is defined by the top horizontal line of points $[i\Delta, \theta = k\Delta]$ for $i = -2k,\ldots,k$; and the threshold bound for option C is defined by the negative diagonal line of points $[i\Delta, -(k+i)\Delta]$ for $i = -2k, \ldots,k$. The remaining $(m-3)(m-2)/2$ states represent transient states. The transition probabilities for the transient states are symbolized as

$$\Pr[\Delta_1(t+h) = +\Delta, \Delta_2(t+h) = 0 \mid \boldsymbol{P}(t) = \boldsymbol{X}] = q_{+0}(\boldsymbol{X}),$$

$$\Pr[\Delta_1(t+h) = +\Delta, \Delta_2(t+h) = -\Delta \mid \boldsymbol{P}(t) = \boldsymbol{X}] = q_{+-}(\boldsymbol{X}),$$

$$\Pr[\Delta_1(t+h) = 0, \Delta_2(t+h) = +\Delta \mid \boldsymbol{P}(t) = \boldsymbol{X}] = q_{0+}(\boldsymbol{X}),$$

$$\Pr[\Delta_1(t+h) = 0, \Delta_2(t+h) = -\Delta \mid \boldsymbol{P}(t) = \boldsymbol{X}] = q_{0-}(\boldsymbol{X}),$$

$$\Pr[\Delta_1(t+h) = -\Delta, \Delta_2(t+h) = 0 \mid \boldsymbol{P}(t) = \boldsymbol{X}] = q_{-0}(\boldsymbol{X}),$$

$$\Pr[\Delta_1(t+h) = -\Delta, \Delta_2(t+h) = +\Delta \mid \boldsymbol{P}(t) = \boldsymbol{X}] = q_{-+}(\boldsymbol{X}),$$

and we define $q_{00}(\boldsymbol{X}) = 1-[q_{+0}(\boldsymbol{X})+ q_{+-}(\boldsymbol{X})+ q_{0+}(\boldsymbol{X})+ q_{0-}(\boldsymbol{X})+ q_{-0}(\boldsymbol{X})+ q_{-+}(\boldsymbol{X})]$ as the probability of no change in state. Note that we assume $q_{++}(\boldsymbol{X}) = q_{--}(\boldsymbol{X}) = 0$, that is no simultaneous changes in the same direction. This is required to satisfy the assumption that the increments sum to zero across all three coordinates.

The transition probabilities are defined by equating the moments of the discrete and continuous state processes:

$$E[\Delta_1(t+h)|\boldsymbol{P}(t)=\boldsymbol{X}]/h = [q_{+0}(\boldsymbol{X}) + q_{+-}(\boldsymbol{X}) - q_{-0}(\boldsymbol{X}) - q_{-+}(\boldsymbol{X})]\cdot\Delta/h = \mu_1(\boldsymbol{X}), \quad (15a)$$

$$E[\Delta_2(t+h)|\boldsymbol{P}(t)=\boldsymbol{X}]/h = [q_{0+}(\boldsymbol{X}) + q_{-+}(\boldsymbol{X}) - q_{0-}(\boldsymbol{X}) - q_{+-}(\boldsymbol{X})]\cdot\Delta/h = \mu_2(\boldsymbol{X}), \quad (15b)$$

$$Var[\Delta_1(t+h)|\boldsymbol{P}(t)=\boldsymbol{X}]/h = [q_{+0}(\boldsymbol{X})+q_{+-}(\boldsymbol{X})+q_{-0}(\boldsymbol{X})+q_{-+}(\boldsymbol{X})]\cdot\Delta^2/h = \phi_{11}^2 - \mu_1(\boldsymbol{X})^2 h, \quad (15c)$$

$$Var[\Delta_2(t+h)|\boldsymbol{P}(t)=\boldsymbol{X}]/h = [q_{0+}(\boldsymbol{X})+q_{-+}(\boldsymbol{X})+q_{0-}(\boldsymbol{X})+q_{+-}(\boldsymbol{X})]\cdot\Delta^2/h = \phi_{22}^2 - \mu_2(\boldsymbol{X})^2 h, \quad (15d)$$

$$Cov[\Delta_1(t+h), \Delta_2(t+h)|\boldsymbol{P}(t)=\boldsymbol{X}] = -[q_{-+}(\boldsymbol{X})+ q_{+-}(\boldsymbol{X})]\cdot\Delta^2/h = \phi_{21} - \mu_1(\boldsymbol{X})\mu_2(\boldsymbol{X})\, h. \quad (15e)$$

As $h \rightarrow 0$, then $\mu_i(\boldsymbol{X})\mu_j(\boldsymbol{X})\, h \rightarrow 0$, and so they can be ignored for small h. The solution to these 5 equations (conditioned on the state $\boldsymbol{X}$) yields the transition probabilities for each transient state:

$$q_{+0}(\boldsymbol{X}) = h\mu_1/2\Delta + h\phi_{11}/2\Delta^2 + h\phi_{21}/\Delta^2 + \beta$$

$$q_{+-}(\boldsymbol{X}) = -h\phi_{21} / \Delta^2 - \beta$$

$$q_{0+}(\boldsymbol{X}) = h\mu_2 / 2\Delta + h\phi_{22} / 2\Delta^2 - \beta$$

$$q_{0-}(\boldsymbol{X}) = -h\mu_2 / 2\Delta + h\phi_{21} / \Delta^2 + h\phi_{22} / 2\Delta^2 + \beta$$

$$q_{-0}(\boldsymbol{X}) = -h\mu_1 / 2\Delta + h\phi_{11} / 2\Delta^2 - \beta$$

$$q_{-+}(\boldsymbol{X}) = \beta.$$

The parameter β reflects indeterminancy in this linear system of equations. Any value of β will reproduce the moments of the continuous state model. The only constraints on β are that $\beta>0$, it yields transition probabilities between zero and one, and the transition probabilities sum to less than one (thus allowing for $q_{00}(\boldsymbol{X}) > 0$). Constraints produced by the moments of the third coordinate do not provide linearly independent equations. The solutions for $q_{-+}(\boldsymbol{X})$ and $q_{+-}(\boldsymbol{X})$ require $\phi_{21} < 0$, which will generally be true as a consequence of the fact the valences sum to zero.

The transition probabilities defined by Equation 15 are then inserted into the Markov chain model to approximate the continuous state diffusion process. Analogous to the binary choice problem, the transition probabilities defined in Equation (15) are arranged into four matrices: the first is a $(m-3)(m-2)/2 \times 1$ vector $\boldsymbol{R}_1$ containing the probabilities of transiting from a transient state to an absorbing state for option A; the second is a $(m-3)(m-2)/2 \times 1$ vector $\boldsymbol{R}_2$ containing the probabilities of transiting from a transient state to an absorbing state for option B; the third is a $(m-3)(m-2)/2 \times 1$ vector $\boldsymbol{R}_3$ containing the probabilities of transiting from a transient state to an absorbing state for option C; and the fourth is a $(m-3)(m-2)/2 \times (m-3)(m-2)/2$ matrix $\boldsymbol{Q}$ containing the probabilities of transiting from one transient state to another. After making these assignments, Equation 12b and 12c are used to calculate the choice probabilities and mean decision times for each of the three options.

The trinary choice model for optional stopping tasks differs in an important way from the trinary choice model developed earlier for the fixed stopping time task. The current model is designed to predict both the choice probabilities as well as the mean decision times for trinary choices, whereas the latter only predicts choice probabilities at a given time point. It is worth mentioning that no new parameters are introduced by extending the model from binary to trinary choices. The model provides predictions for binary and trinary choices using a common set of parameters. Below we compare the predictions derived from DFT for binary and trinary decision times with another class of models called parallel race utility models.

Comparison of Binary and Trinary Decision Times. According to the class of *horse race* random utility models (Marley & Colonius, 1992), each option within a set generates a random decision time (T_i for option i), and the option producing the minimum decision time is chosen. The probability of choosing A_i from a set $X = \{A_1,\ldots,A_n\}$ is given by

$$\Pr[A_i \mid X] = \Pr[\, T_i = \min(T_j\,,\, j \varepsilon X)], \quad (16)$$

where $\boldsymbol{T_n}' = [T_1, \ldots, T_n]$ is a vector of random decision times. The standard parallel race utility model assumes that the distribution of a random time vector $\boldsymbol{T}_m$, $m < n$, for a set Y $\subset$ X is derived from the distribution of $\boldsymbol{T}_n$ by integrating out the random decision times of the options in X and not in Y. This class of models predicts that mean time to make a choice from the complete set X will be less than the mean time to choose from the restricted set Y $\subset$ X. In other words, adding new options to the choice set can only decrease the mean decision time. This follows from the fact that $\Pr[\min(T_i\,,\, i \varepsilon X) > t] = \Pr[\min(T_i\,,\, i\,\varepsilon\, Y) > t] \times \Pr[\min(T_i\,,\, i \varepsilon X\text{-}Y) > t \mid \min(T_i\,,\, i\,\varepsilon\, Y) > t] < \Pr[\min(T_i\,,\, i\,\varepsilon\, Y) > t]$.

The predictions derived from DFT are more complex -- the mean time to choose an option from the complete set X may be either longer or shorter than the mean time to choose that same option from the restricted set Y$\subset$ X, depending on the composition of the options within each set. Table 2 illustrates these two cases. In the first pair of columns, the mean time to choose from the complete set of three options is longer than the mean time to choose from the restricted set of two options. For the second case, shown in the second pair of columns, the reverse relation is obtained. The first case was produced by adding a relatively inferior third option to a set that already contained two highly competitive options. The second case was produced by adding a highly superior third option to a set that already contained two highly competitive options.

Empirical results generally show that decision time increases as the number of alternatives in the choice set increases, contrary to the horse race utility models (Onken, Hastie, & Revelle, 1985). However, a study by Keisler (1966) indicates that the results are more complex -- for some options, binary choice was faster than four alternative choice, but under other conditions, there was no difference in choice time between binary and four alternative choices. Thus more research on this issue is needed.

Table 2 also shows provides a comparison of the choice probabilities produced by the binary and trinary choices. It is worth noting that DFT produces violations of the independence between alternatives property for the first case shown in Table 2. In this case, $Pr[A \mid \{A,B\}] > Pr[B \mid \{A,B\}]$ but $Pr[A \mid \{A,B,C\}] < Pr[B \mid \{A,B,C\}]$, violating the independence between irrelevant alternatives property. This result arises from the fact that the covariance between options A and C is less negative than the covariances between B and each of the other two options. In other words, C is more similar to A than B, and thus C takes more probability away from A as compared to B. This prediction of the model is consistent with the well-known empirical finding called the "similarity effect" (Tversky, 1972).

VII. Certainty Equivalents and Matching Processes

Certainty equivalents are elicited by asking the decision maker to select a value (e.g., price) that makes the decision maker indifferent between the selected value and an uncertain multi-attribute choice option. The selected value that achieves this indifference is denoted CE_A for the certainty equivalent corresponding to option A.

According to DFT, certainty equivalents are found by a matching process composed of a sequence of implicit binary comparisons between the uncertain multi-attribute option, option A, and test values for the certainty equivalent CE. Initially, the decision maker tries out a starting value, say CE(0), and compares this with option A. This comparison process can lead to one of three states: option A is chosen over CE(0), or CE(0) is chosen over option A, or the decision maker is indifferent. If the decision maker is indifferent, then CE(0) is reported as the certainty equivalent for option A, and the matching process stops. If the A is chosen over CE(0), then the test value is increased a fixed amount, $CE(1) = CE(0) + \Delta$ and the process is repeated; if CE(0) is chosen over A, then the test value is decreased a fixed amount, $CE(1) = CE(0)-\Delta$ and the process is repeated. This comparison process continues, generating a sequence of test values CE(0), CE(1), …, CE(k), until the indifference state is eventually reached, say at value CE(K), and the matching process stops. The certainty equivalent, CE_A, reported by the subject is the last test value that happened to stop the matching process, so that $CE_A = CE(K)$. Thus the matching process is represented by a Markov chain defined on the set of possible certainty equivalence values. The transition probabilities for this Markov chain are determined by the comparison process described below.

According to DFT, the comparison process is based on the binary choice model for optional stopping with one minor modification. The original binary choice model allows only two responses, but the modified binary choice model includes an indifference response. A transition to the indifference response occurs with a fixed probability, ω, every time the preference state $P_1(t)$ from the binary choice enters the neutral state $P_1 = 0$. After making this small modification to the transition matrix, the Markov chain method

(Equation 12) is used again to compute the choice probabilities across the three responses. These choice probabilities then drive the higher-level Markov chain for the matching process. See Busemeyer & Townsend (1992) for a detailed presentation of the two stage Markov chain model used to compute the final distribution of certainty equivalents produced by the matching process. This simple modification of DFT introduces only two free parameter, ω and Δ. At the same time, it provides a simple explanation for the complex relations found between choice probabilities and certainty equivalents.

Relations Between Certainty Equivalents and Choices. According to standard deterministic models of preference (see Luce, 2000), the behavioral indifference between CE_A and option A is represented by the indifference equation: $u(CE_A) = u(A)$, where u is a real valued utility function defined on the set of options, A ε X, and it is a monotonically increasing function of the certain value x. According to these standard theories, if A is preferred to B then $u(A) = u(CE_A) > u(CE_B) = u(B)$, and by monotonicity, this in turn implies $CE_A > CE_B$. Surprisingly, empirical research comparing choice and certainty equivalents has found systematic preference reversals such that $\Pr[A|\{A,B\}] > \Pr[B|\{A,B\}]$ but $E[CE_B] > E[CE_A]$.

One common explanation for these reversals is that decision makers change their utility function when performing choice versus certainty equivalence tasks (cf. Tversky, Sattath, & Slovic, 1988). In contrast, DFT assumes that a common utility scale underlies both tasks, and instead reversals are caused directly from the matching process. DFT provides an accurate account of the both the qualitative and quantitative relations among choice probabilities and certainty equivalences (Busemeyer and Goldstein, 1992;

Townsend and Busemeyer, 1995). Furthermore, deterministic theories (e.g., Tversky, et al., 1988) only predict the ordering of mean prices, and fail to account for price distributions. The matching process of DFT is capable of predicting not only the mean price, but also the entire distribution of prices (Busemeyer and Goldstein, 1992).

VIII. Discussion

Comparison with other Models. Within each of the above sections, we described fundamental differences between the predictions of DFT and alternative models of decision-making. In particular, we compared DFT to simple scalability models, standard random utility models, horse race random utility models, and the elimination-by-aspects model. In each case we showed that DFT was able to provide a better account of the basic findings than the competitor. For example, random utility models and the elimination-by-aspects model fail to account for violations of regularity that have been empirically reported, whereas DFT provides a simple explanation for this finding.

A dynamic model of preference that we have not discussed thus far is the *token* model developed by Regenwetter, Falmagne, and Grofman (1999). It is difficult to directly compare this model with DFT because they are designed for different purposes. First, DFT is designed to explain the evolution of preference that occurs during deliberation for within a single decision, whereas token theory is designed to explain changes in preferences across a sequence of choices. Second, DFT is designed to predict the relations between choice probability and decision time, whereas token theory is designed to predict the rate of change in preference orders over time. Third DFT was designed to account for findings from risky decision-making, multi-attribute decision-making, and multi-alternative choice behavior, whereas token theory has mainly been applied to voting behavior. Formally, DFT is a Markov process based on a quantitative preference state, and choice probabilities are derived from absorption probabilities (for the optional stopping task), whereas token theory is a Markov process based on a qualitative preference state, and the choice probabilities are derived from ergodic state probabilities. Competitive empirical tests of these two models are challenges for future researchers.

Testability. Altogether, DFT entails the following models and parameters: First we presented the fixed stopping time model, which required a value matrix $\boldsymbol{M}$, the mean and covariance matrix of the weights, $\boldsymbol{w}$ and $\boldsymbol{\Psi}$, the initial preference state $\boldsymbol{P}(0)$, and a feedback matrix $\boldsymbol{S}$. Second, we presented the optional stopping model, which required an additional threshold bound parameter, θ, and the multi-attribute version of the model required additional attention switching probabilities, π_{ij}. Finally, we presented a model for certainty equivalents, which required a step size parameter Δ and an indifference response probability ω. These parameters allow the theory to predict (a) choice probability as a function of deliberation time, (b) relations between choice probability and decision time, (c) relations between binary and trinary choices, and (d) relations between choices and certainty equivalents. The use of a common theory with a common set of parameter for all of these alternative measures of preference provide opportunities for what we call generalization tests of the theory (Busemeyer & Wang, 2000). The basic idea is that we estimate the parameters from one dependent variable under a broad set of conditions (binary choice probabilities for pairs of options), and then we use these same parameters to test the model against a variety of other dependent variables (choice probabilities under time pressure, trinary choice probabilities, decision times, selling prices). For examples of such tests, see Dror, Busemeyer, and Basola (1999) for choice probability and decision time; Busemeyer and Goldstein (1992) for choice probability and selling prices; Diederich (2000) for choice probability under various levels of time pressure; Roe, et al. (in press) for binary and trinary choice probabilities.

References

Anderson, T. W. (1971) *The statistical analysis of ime series*. NY: Wiley.

Aschenbrenner, K. M., Albert, D., & Schmalhofer, F. (1984). Stochastic choice heuristics. *Acta Psychologica*, 56(1-3), 153-166.

Ashby, F. G. (2000) A stochastic version of general recognition theory. *Journal of Mathematical Psychology, 44*, 310-329.

Bhattacharya, R. N., & Waymire, E. C. (1990). Stochastic processes with applications. New York: Wiley.

Bock, R. D., & Jones, L. V. (1968). The measurement and prediction of judgment and choice. San Francisco, CA: Holden-Day.

Bockenholt, U. (1992). Multivariate models of preference and choice. In F. G. Ashby (Ed.). Hillsdale, NJ: Earlbaum Associates.

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

Busemeyer, J. R. & Townsend, J. T. (1993). Decision Field Theory: A dynamic cognition approach to decision making. *Psychological Review*, 100, 432-459.

Busemeyer, J. R. & Goldstein, W. M. (1992) Linking together different measures of preference: A dynamic model of matching derived from decision field theory. *Organizational Behavior and Human Decision Processes, 52*, 370-396.

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

Cox, D.R., & Miller, H. D. (1965). *The theory of stochastic processes*. London: Chapman Hall

De Soute, G. Feger, H. & Klauer, K. C. (Eds.) (1989). New developments in probabilistic choice modeling. Amsterdam: North Holland.

Diederich, A. (1997). Dynamic stochastic models for decision making under time constraints. *Journal of Mathematical Psychology*, 41(3), 260-274.

Diederich, A. (2000). Explaining experimental decisions under time pressure using multi-attribute decision field theory. Manuscript under review.

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.

Fishburn, P. C. (1998) Stochastic Utility. In S. Barbera, P. J. Hammond, C. Seidl (Eds.) *Handbook of utility theory, Vol. 1: Principles*. (pp. 273-318). Dordrecht: Kluwer.

Huber, J., Payne, J. W., & Puto, C. (1982). Adding asymmetrically dominated alternatives: Violations of regularity and the similarity hypothesis. *Journal of Consumer Research*, 9(1), 90-98.

Keeney, R. L. & Raiffa, H. (1976). *Decisions with multiple objectives: Preference and value tradeoffs.* NY: Wiley.

Keisler, C. A. (1966) Conflict and number of choice alternatives. *Psychological Reports, 18*, 603-610.

Link, S.W. (1992). *The wave theory of difference and similarity.* Hillsdale, NJ: Erlbaum

Luce, R. D. (1959). *Individual choice behavior: A theoretical analysis*. New York: Wiley, 1959.

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

Marley, A. A. J. (1989). Random utility family that includes many of the ‘classical’ models and has closed form choice probabilities and choice reaction times. *British Journal of Mathematical and Statistical Psychology*, 42, 13-36.

Marley, A. A. & Colonius, H. (1992) The "horse race" random utility model for choice probabilities and reaction times, and its competing risks interpretation. *Journal of Mathematical Psychology, 36*, 1-20.

McFadden, D. (1978) Modeling the choice of residential location. In A. Karlquist et al. (Eds.) Spatial interaction theory and planning models.(pp. 75-96). Amsterdam: North-Holland.

Mellers, B. A. & Biagini, K. (1994) Similarity and choice. *Psychological Review 101*, 505-518.

Nosofsky, R. M. &Palmari, T. J. (1997). Comparing exemplar-retrieval and decision-bound models of speeded perceptual classification. *Perception & Psychophysics, 59*, 1027-1048.

Onken, J. Hastie, R., & Revelle, W. (1985) Individual differences in the use of simpification strategies in a complex decision task. *Journal of Experimental Psychology: Human Perception and Performance, 11*, 14-27.

Petrusic, W. F. & Jamieson, D. G. (1978) Relation between probability of preferential choice and time to choose changes with practice. *Journal of Experimental Psychology: Human Perception and Performance, 4,* 471-482.

Ratcliff, R. (1978). A theory of memory retrieval. *Psychological Review*, 85, 59-108.

Regenwetter, M., Falmagne, J.-C. & Grofman, B. (1999). A Stochastic Model of Preference Change and its Application to 1992 Presidential Election Panel Data. *Psychological Review*, **106**, 362-384.

Roe, R., Busemeyer, J. R., & Townsend, J. T. (in press) Multi-Alternative Decision Field Theory: A Dynamic Connectionist Model of Decision-Making. *Psychological Review*.

Searle, S. R. (1982) *Matrix algebra useful for statistics*. NY: Wiley.

Smith, P. L. (1995). Psychophysically principled models of visual simple reaction time. *Psychological Review*, 102(3), 567-593.

Stroock, D. W. & Varadhan, S.R.S. (1979) *Multidimensional diffusion processes*. Berlin: Springer – Verlag.

Svenson, O. & Edland, A. (1987). Change of preferences under time pressure: Choices and judgments. *Scandinavian Journal of Psychology, 28*, 322-330.

Townsend, J. T. & Busemeyer, J. R. (1989) Approach-avoidance: Return to dynamic decision behavior. In C. Izawa (Ed.)Current issues in cognitive processes: Tulane Flowerree Symposium on cognition (pp. 107-133). Hillsdale, NJ: Erlbaum.

Townsend, J. T., & Busemeyer, J. R. (1995). Dynamic representation of decision-making. In R. F. Port and T. van Gelder (Eds.) Mind as Motion. MIT press.

Tversky, A. (1972). Elimination by aspects: A theory of choice. *Psychological Review*, 79(4), 281-299.

Tverksy, A., Sattath, S., & Slovic, P. (1988) Contingent weighting in judgment and choice. *Psychological Review, 95*, 371-384.

Vickers, D., B, J., Smith, P., & Brown, M. (1985). Experimental paradigms emphasizing state or process limitations: I. Effects of speed-accuracy tradeoffs. *Acta Psychologica,* 59(2), 129-161.

Table 1: Hypothetical Choice

	State X		State Y	
Options	Cost	Quality	Cost	Quality
A	C_{AX}	Q_{AX}	C_{AY}	Q_{AY}
B	C_{BX}	Q_{BX}	C_{BY}	Q_{BY}
C	C_{CX}	Q_{CX}	C_{CY}	Q_{CY}

Table 2

Comparison of Predictions Computed from Binary and Trinary Choice Models

	Case 1			Case 2	
Option	Prob.	Time	Option	Prob.	Time
A(2)	.52	9.99	A(2)	.50	10.0
B(2)	.48	9.99	B(2)	.50	10.0
A(3)	.34	15.65	A(3)	.065	8.40
B(3)	.43	13.87	B(3)	.065	8.40
C(3)	.23	16.17	C(3)	.870	9.35

Note: A(2) denotes choosing option A in a binary choice, B(2) denotes choosing option B in a binary choice, A(3) denotes choosing option A in a trinary choice, ect.

Predictions for Case 1 were computed using $\mu_1 = .0548$, $\mu_2 = -.0052$, $\mu_3 = -.0495$, $\phi_{11} = 4.04$, $\phi_{22} = 6.22$, $\phi_{33} = 3.599$, $\phi_{21} = -3.33$, $\phi_{31} = -.711$, $\phi_{32} = -2.89$, $k = 10$, h=.10, $\boldsymbol{\Gamma} = \mathbf{0}$.

Predictions for Case 2 were computed using $\mu_1 = -.5$, $\mu_2 = -.5$, $\mu_3 = 1.0$, $\phi_{11} = 5.25$, $\phi_{22} = 5.25$, $\phi_{33} = 3.0$, $\phi_{21} = -3.75$, $\phi_{31} = -1.50$, $\phi_{32} = -1.50$, $k = 10$, h=.10, $\boldsymbol{\Gamma} = \mathbf{0}$.

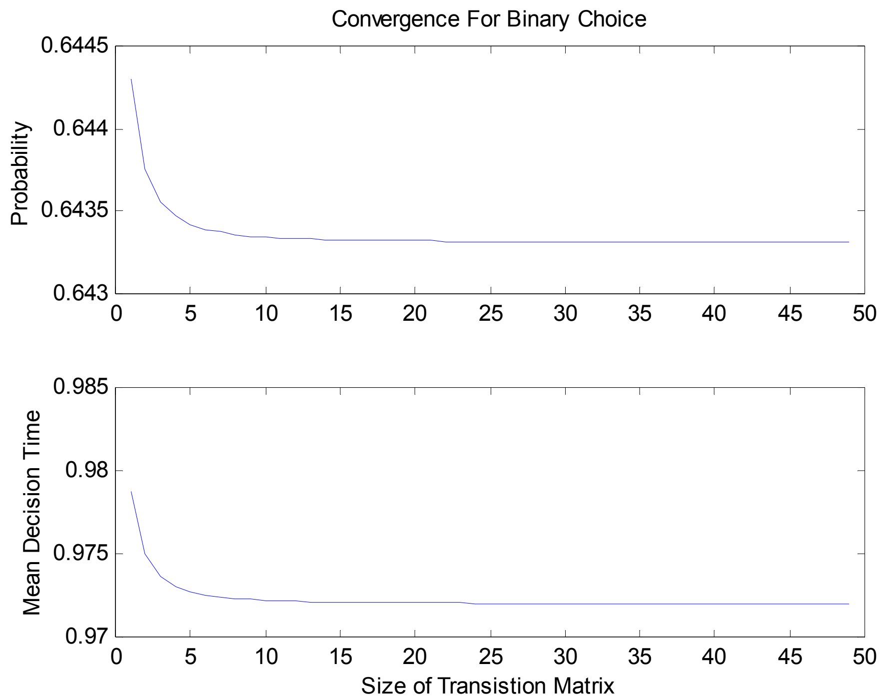
Figures

Figure 1: Convergence of the Markov chain approximation to the results of a diffusion process as the time unit (h) approaches zero and the number of steps to reach the bound becomes large. The top panel shows the convergence for choice probability and the bottom panel shows the convergence for mean decision time.

Figure 2: Effects of the increasing the time limit (thus increasing the threshold bound) on choice probability for the multi-attribute version of DFT with Markov attention switching. Note that the probability starts out above .50 under short time limits and falls below .50 under longer time limits.

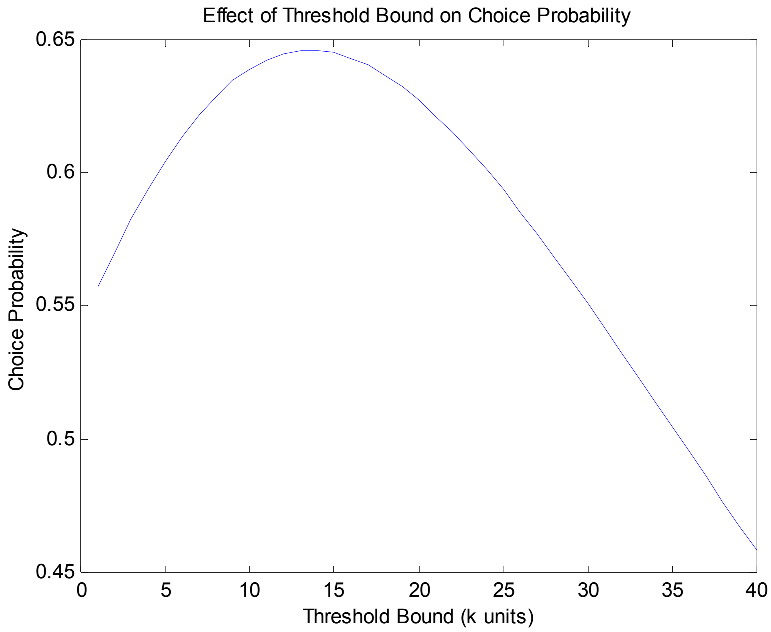
Figure 3: Attention switching moderates the relation between choice probability and mean decision time. The horizontal axis represents the probability of switching attention away from the most important attribute. The top curve represents the probability of choosing option A, and the bottom curve represents the mean time to choose B minus the mean time to choose A. Note that as long as there is some probability of switching, then the more frequently chosen option (A) is also chosen more quickly.

f1


Convergence For Binary Choice
Probability
0.6445
0.644
0.6435
0.643
0 5 10 15 20 25 30 35 40 45 50
Mean Decision Time
0.985
0.98
0.975
0.97
0 5 10 15 20 25 30 35 40 45 50
Size of Transistion Matrix

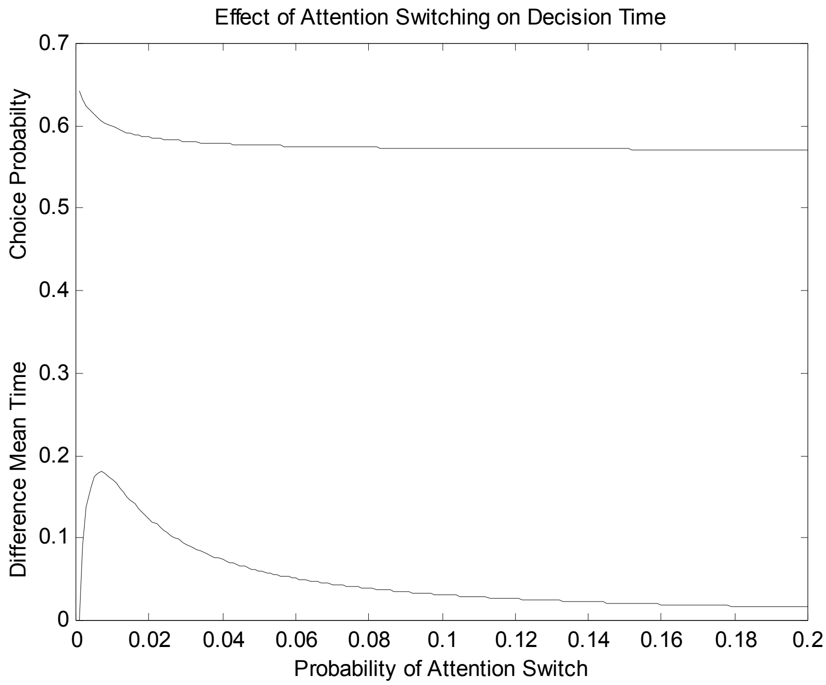
f1

f2


Effect of Threshold Bound on Choice Probability
Choice Probability
0.65
0.6
0.55
0.5
0.45
0
5
10
15
20
25
30
35
40
Threshold Bound (k units)

f2

f3


Effect of Attention Switching on Decision Time
Choice Probability
Difference Mean Time
Probability of Attention Switch
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
0.02
0.04
0.06
0.08
0.1
0.12
0.14
0.16
0.18
0.2

f3