An Integrated Model of Discrete Choice and Response Time with Application to Conjoint Analysis

by

Thomas Otter
Fisher College of Business
Ohio State University
otter_2@cob.osu.edu

Greg M. Allenby
Fisher College of Business
Ohio State University
allenby.1@osu.edu

Trish van Zandt
Department of Psychology
Ohio State University
van-zandt.2@osu.edu

March, 2006

This research was funded by NSF grants SES-0437251 and SES-0214574. The authors thank Bryan Orme and Sawtooth Software for helpful comments and making the data available for study. We thank Pradeep Chintagunta, JP Dube, John Liechty, Steve MacEachern, Puneet Manchanda, Mario Perrugia, Shiling Ruan, and Michel Wedel for comments. Previous versions of this paper were presented at the University of Chicago, the ART-Forum, 'Markentag' at WU-Wien, the 70th Annual Meeting of the Psychometric Society, the Joint Statistical meeting, the conference on Bayesian Applications and Methods in Linz, Austria, Notre Dame University, and the University of Michigan.

An Integrated Model of Discrete Choice and Response Time with Application to Conjoint Analysis

March, 2006

Abstract

Response times and discrete choices are jointly modeled using a time varying Poisson process where choice alternatives generate signals in the respondent's mind. A choice is made when a sufficient number of signals are generated, and the response time corresponds to the time taken for the fastest process to reach a threshold. The allocation of limited cognitive resources in evaluating multiple alternatives in a choice set alters the rate at which signals are generated for any one alternative. The integrated model, marginally, nests the standard logit model, and is shown to improve inference in conjoint analysis.

Keywords: Poisson Process; Choice Model; Bayesian Analysis; Human Information Processing

An Integrated Model of Choice and Response Time with Application to Conjoint Analysis

1. Introduction

The proliferation of computer and web-based interviewing tools makes available response times for analysis by choice modelers. Although consumer behavior researchers have long studied response times as a *dependent* variable related to aspects of memory, attitudes and decision making (e.g. Aaker et al. 1980, Fazio and Olson 2003, MacLachlan and Myers 1983, MacLachlan et al. 1979, Park and Hastak 1994, Tyebjee 1979), response times have not been successfully incorporated as dependent variables into the mathematical formulation of choice models in the marketing literature.

Mathematical psychologists have developed formal frameworks for integrating choice and response times (e.g., Townsend and Ashby 1983), but have not investigated its applicability to the types of decisions studied in applied disciplines such as marketing. This paper starts with a model put forth by mathematical psychologists and extends it to study consumer choice within the framework of conjoint analysis.

Choice-based conjoint analysis is an experimental technique frequently used in marketing to infer the relationship between respondents' product choices and their unobserved valuations of product attributes, sometimes called part-worths. Respondents are asked to make choices from a number of sets of varying composition. Subsequently respondents' product choices are regressed on product attributes that are varied across choice sets. The typical data set consists of a limited number of choices per respondent and involves a relatively much larger number of respondents. Therefore, analysis usually relies on hierarchical models.

We show that response times are informative about respondents' valuations of products and their attributes. Quick response times point to easy decisions where at least one of the alternatives is outstanding, and slow response times point to hard decisions where the alternatives are less or equally attractive. The link between response times and these aspects of a choice task is conditional on respondents' diligence and allocation of cognitive capacity, which are latent. We extend the basic modeling framework used in psychology to include measures of both latent variables.

Our model treats response times and choice as dependent variables in an underlying latent process where the choice alternatives are conceived as signal generating processes in a respondent's mind. A choice is made when the cumulative number of signals generated by any one alternative reaches a threshold value. The observed response time is the time it takes the winning alternative to generate a sufficient number of signals. We show that this model, marginalized over response time, reduces to the standard logit model when the threshold value is equal to one signal. An advantage of this formulation is that by introducing the threshold as a model parameter we obtain a measure of respondent diligence, or effort expended, that is conceptually independent of the attractiveness of the alternatives under study. That is respondents A and B could share the same tastes but have different choice probabilities because they expended different amounts of effort. A higher threshold implies a longer race among signal generating processes; such a race is disproportionally less likely to be won by a bad alternative as the attractiveness of other alternatives increases. Intuitively, this makes sense as a consequence of more effortful decision making aimed at identifying the most attractive alternative.

We investigate properties of the model using conjoint data collected by a leading marketing research firm. The choice task involved choice sets of five products and each respondent provided 18 choices. We show that i) the estimated model has threshold values different from one for the majority of respondents; ii) the need to allocate limited cognitive resources among similarly attractive alternatives increases response times; iii) respondents exhibit learning effects associated with quicker responses later in the conjoint task; iv) holdout predictions for the proposed model improve relative to the logit model; and v) inferences are different from those obtained from a standard logit model.

The organization of the paper is as follows. In section 2 we derive the integrated model and show that it, marginally, nests the standard logit model. Model estimation is discussed in section 3, and the data are described in section 4. Parameter estimates are reported in section 5, and section 6 discusses implications of the model with regard to inference in conjoint analysis. Section 7 offers concluding remarks and discusses avenues for future research.

2. The Model

The primitive of the proposed model is a binary zero/one signal, where zero is equivalent to no signal and one is a signal. Each alternative in a choice set is assumed to independently trigger signals in a respondent's mind. Once the number of signals generated by any one alternative surpasses a threshold value, the respondent picks this alternative and the process is terminated. Response time is the time it takes the winning alternative to trigger a threshold number of signals in a respondent's mind. At this time, each of the remaining alternatives in the choice set has triggered fewer signals than the threshold value. The primitive assumptions of the

model only allow for favorable signals. A bad alternative is therefore characterized by the lack of favorable signals.

Alternatives with higher quality are assumed to generate signals at a faster rate, and the threshold value determines the diligence of the decision maker. High threshold values indicate well deliberated choices, where the best alternative is identified with greater certainty. In the limit, as the threshold value approaches infinity, the response time also approaches infinity and the choice probability converges to zero or one. For moderate values of the threshold, the response time and choice probability are less extreme.

A probability law for the number of signals in favor of an alternative at a specific point in time may be derived from the following assumptions (Smith and Van Zandt 2000): i) Each alternative starts with zero signals at time zero, and the number of signals is a weakly monotone function of time. ii) The probability of a signal in the time interval from t to $t + h$ is assumed proportional to the rate at which signals are generated at time t and the length of the time interval. iii) The probability of more than one signal in the time interval from t to $t + h$ converges to zero as the length of this interval approaches zero. Together these assumptions imply a Poisson process with time varying parameter for alternative specific signal counters,

$$P\left(N\left(t\right)=n\right)=\frac{\left[R_a\left(t\right)\right]^n \exp\left(-R_a\left(t\right)\right)}{n!} \tag{1}$$

where $N(t)$ is the number of signals generated in favor of alternative a at time t and $R_a(t)$ is the integrated rate at which alternative a triggers signals in the respondent's mind, i.e.

$R_a(t) = \int_0^t r_a(s)\,ds$ with $r_a(s)$ strictly positive (see Townsend and Ashby 1983 for a general discussion of counting models).

The density for the time it takes to accumulate a number of signals equal to the threshold value K is derived using the identity $P(N(t) < K) = P(T > t)$, i.e. the probability that alternative a has triggered fewer signals than the threshold value at time t is equal to the probability that this alternative's finishing time T exceeds t.

$$p(t) = \frac{\left[R_a(t)\right]^{(K-1)} r_a(t) \exp\left(-R_a(t)\right)}{(K-1)!} \tag{2}$$

The observed response time corresponds to the time it takes the chosen alternative to accumulate a number of signals equal to the threshold K. As a consequence of observing which alternative was chosen, it is also revealed that each of the remaining alternatives has accumulated at most $(K - 1)$ signals in the respondent's mind at this point in time. Because all stochastic processes involved are, conditioned on parameters, assumed independent we can simply multiply the corresponding densities together and arrive at the following expression for choosing, say, alternative a, from a set of J alternatives at time t:

$$p(t,a) = \frac{\left[R_a(t)\right]^{(K-1)} r_a(t) \exp\left(-R_a(t)\right)}{(K-1)!} \prod_{j:1,\ldots,J\setminus a} \left(\sum_{k=0}^{K-1} \frac{\left[R_j(t)\right]^k \exp\left(-R_j(t)\right)}{k!} \right) \tag{3}$$

If we are willing to assume (i) that the integrated rates can be factored into a fixed alternative specific component and a function of time $U(t)$ independent of the alternative, e.g.

$R_a(t) = r_a\, U(t)$, and (ii) that $U(t)$ is continuous in $\mathbb{R}^+$ we obtain marginal choice probabilities, e.g. $P(a \mid 1,...,J) = \int_{t=0}^{\infty} p(t,a)\,dt$, that only depend on the alternative specific fixed components and the threshold value (Smith and Van Zandt 2000).

We illustrate this result for the case of a trinary choice set composed of alternatives *{a, b, c}* that trigger signals at the instantaneous rates $\{r_a\, U'(t),\ r_b\, U'(t),\ r_c\, U'(t)\}$, respectively where $U'(t)=dU(t)/dt$. The marginal probability of choosing *a* from the trinary set *{a, b, c}* is:

$$P(a \mid \{a,b,c\}) = \int_{t=0}^{\infty} p(t,a)\,dt = \sum_{k_b=0}^{K-1} \sum_{k_c=0}^{K-1} \frac{(K+k_b+k_c-1)!}{k_b!\,k_c!\,(K-1)!} \frac{r_b^{k_b} r_c^{k_c} r_a^{K}}{(r_a+r_b+r_c)^{K+k_b+k_c}} \qquad (4)$$

For the special case of $K = 1$ (which implies that k_b and k_c are zero in equation (4)) the marginal choice probability reduces to a standard multinomial logit $P(a \mid \{a, b, c\}) = r_a/(r_a+r_b+r_c)$. When $K > 1$, ratios of choice probabilities are no longer independent of the composition of the choice set and do not exhibit the independence of irrelevant alternative (IIA) property associated with the logit model.

We illustrate the differences in the implied choice probabilities with six simulated trinary choice sets. The rates r_a and r_b are constant across sets with r_a smaller than r_b. Rate r_c is increased from set to set and probabilities as well as probability ratios are computed applying equation (4) and setting $K = 1$ and 2 respectively. A comparison of $P(c)$ across the choice sets shows that the probabilities obtained from the model with $K = 2$ are more extreme than those from the logit model, given the same underlying rates (rows 5 and 6 in Table 1). The probability ratio of choosing alternative *b* over alternative *a* illustrates the structural difference between the

two sets of probabilities (rows 7 and 8 in Table 1). This ratio is independent of the probability of choosing alternative *c* in the logit model and therefore constant across the sets (IIA). However as $K > 1$ this ratio is increasing in the choice probability of alternative *c*, i.e. share is more readily stolen from the weaker alternative without penalizing alternatives that are relatively stronger.

	K	Set #1	Set #2	Set #3	Set #4	Set #5	Set #6
P(*a*) ($r_a = 7.39$)	1	.2880	.2492	.1744	.0890	.0348	.0118
	2	.2237	.1959	.1116	.0291	.0040	.0004
P(*b*) ($r_b = 16.44$)	1	.6410	.5547	.3881	.1982	.0774	.0263
	2	.7563	.6719	.4039	.1172	.0181	.0021
P(*c*) (r_c: increasing)	1	.0710	.1961	.4376	.7128	.8878	.9619
	2	.0200	.1321	.4845	.8537	.9778	.9975
P(*b*) / P(*a*)	1	2.2255	2.2255	2.2255	2.2255	2.2255	2.2255
	2	3.3804	3.4294	3.6191	4.0284	4.4904	4.7749
r_c		1.82	5.81	18.54	59.15	188.67	601.85

Table 1: Choice probabilities and ratios of choice probabilities for $K = 1$ and $K = 2$

In a logit model where $K = 1$ the joint density of choosing alternative *a* at time *t* is obtained from equation (3) as

$$p(t,a) = r_a \exp\left(-r_a U(t)\right) \prod_{j:1,\ldots,J\setminus a} \exp\left(-r_j U(t)\right) \tag{5}$$

The logit model therefore implies exponentially distributed response times with the immediate consequence that response time (conditional on parameters) is not informative about choice; i.e. the conditional probability of, say, choosing alternative *a* at time t, $p(a \mid t)$, is equal to the marginal probability $\int_{t=0}^{\infty} p(a,t)\,dt$. If *K > 1* the choice probabilities conditional on response time are different from their marginal counterparts and take the following form:

$$P(a|t) = \frac{p(t,a)}{\sum_{j=1}^{J} p(t,j)} = \frac{\left[r_a U(t)\right]^{(K-1)} r_a U'(t) \exp\left(-r_a U(t)\right)}{(K-1)!} \prod_{j:1,\ldots,J\setminus a} \left(\sum_{k=0}^{K-1} \frac{\left[r_j U(t)\right]^k \exp\left(-r_j U(t)\right)}{k!} \right) \Bigg/ \sum_{j=1}^{J} \left(\frac{\left[r_j U(t)\right]^{(K-1)} r_j U'(t) \exp\left(-r_j U(t)\right)}{(K-1)!} \prod_{l:1,\ldots,J\setminus j} \left(\sum_{k=0}^{K-1} \frac{\left[r_l U(t)\right]^k \exp\left(-r_l U(t)\right)}{k!} \right) \right) \quad (7)$$

Parameterization of Rates in Conjoint Analysis

For the purpose of conjoint analysis we parameterize the instantaneous rates at which alternative *a* in choice set *m* triggers signals in respondent *i's* mind as *exp(*$\mathbf{x}'_{am}\,\boldsymbol{\beta}_i$*) U'(t)* where $\mathbf{x}_{am}$ is a known design vector and $\boldsymbol{\beta}_i$ is a conforming vector of unknown parameters sometimes called part-worths. Thus, we obtain the following joint likelihood for respondent *i's* choice of alternative *a* from set *m* comprised of *J* alternatives at time *t*:

$$p\left(a_m, t_m \middle| \boldsymbol{\beta}_i, K_i, \mathbf{X}, U(t)\right) =$$

$$\frac{\left[\exp\left(\mathbf{x}'_{am}\boldsymbol{\beta}_i\right)U(t)\right]^{K_i-1}\exp\left(\mathbf{x}'_{am}\boldsymbol{\beta}_i\right)U'(t)\exp\left(-\exp\left(\mathbf{x}'_{am}\boldsymbol{\beta}_i\right)U(t)\right)}{(K_i-1)!}\times \tag{8}$$

$$\prod_{j:1,\ldots,J\setminus a}\left(\sum_{k=0}^{K_i-1}\frac{\left[\exp\left(\mathbf{x}'_{jm}\boldsymbol{\beta}_i\right)U(t)\right]^k}{k!}\exp\left(-\exp\left(\mathbf{x}'_{jm}\boldsymbol{\beta}_i\right)U(t)\right)\right)$$

The unknown parameters that need to be estimated are the vector of part-worths $\boldsymbol{\beta}_i$ the respondent specific threshold value K_i and the functional values $U(t)$ and $U'(t)$. Before we discuss the substantive implications of different choices for $U(t)$ we would like to point out that, for $U(t)=f(\cdot)t$ and $U'(t)=f(\cdot)$ independent of t, the joint likelihood of equation (8) may be expressed as

$$p\left(a_m, t_m \middle| \boldsymbol{\beta}_i, K_i, \mathbf{X}, f(\bullet)\right) =$$

$$\frac{\left[\exp\left(\mathbf{x}'_{am}\boldsymbol{\beta}_i\right)f(\bullet)\right]^{K_i}t_m^{K_i-1}\exp\left(-\exp\left(\mathbf{x}'_{am}\boldsymbol{\beta}_i\right)f(\bullet)t_m\right)}{(K_i-1)!}\times \tag{8a}$$

$$\prod_{j:1,\ldots,J\setminus a}\left(\int_{t_m}^{\infty}\frac{\left[\exp\left(\mathbf{x}'_{jm}\boldsymbol{\beta}_i\right)f(\bullet)\right]^{K_i}s^{K_i-1}\exp\left(-\exp\left(\mathbf{x}'_{jm}\boldsymbol{\beta}_i\right)f(\bullet)s\right)}{(K_i-1)!}ds\right)$$

The first factor is in the form of a two-parameter Gamma density with (integer) shape parameter K_i and scale parameter $exp(\mathbf{x}'_{am}\boldsymbol{\beta}_i)\,f(\cdot)$. The second factor is derived from the second factor in equation (8) using the identity $P(N(t) < K) = P(T > t)$. It is the joint probability that the time it takes the non-chosen alternatives to generate K_i signals exceeds the observed response time t_m. Whenever $f(\cdot)$ is chosen such that it is equal for all alternatives in a choice set m, marginal choice

probabilities are, conditioned on $\boldsymbol{\beta}$ and K, independent of $f(\cdot)$ (see equation (4)). That is, while estimates of $\boldsymbol{\beta}$ and K are informed by the response times through the likelihood, conditioned on these estimates marginal choice probabilities are independent of $f(\cdot)$.

The simplest mapping $U(t) \equiv t$ may not be the best choice in applied settings for three reasons: i) respondents may differ with regard to their speed of thinking and the cognitive capacity allocated to the task; ii) respondents may change the amount of cognitive capacity allocated over the course of repeated measurements and/or become more proficient; iii) respondents' cognitive capacity is limited. We will address each of these points in turn proposing changes to the basic model that we explore in our empirical analysis.

Heterogeneity

With the simple map $U(t) \equiv t$ the rate at which alternative a accumulates signals is equal to $\exp\left(\mathbf{x}'_{am}\boldsymbol{\beta}_i\right)$. For fixed respondent-specific threshold K_i and tastes $\boldsymbol{\beta}_i$, differences in response time distributions across choice tasks are thus a function only of $\mathbf{X}_1 \ldots \mathbf{X}_M$, where $\mathbf{X}_m$ is the choice set specific design matrix $[\mathbf{x}_{1m} \ldots \mathbf{x}_{Jm}]'$. Differences in response time distributions across respondents conditional on some $\mathbf{X}_m$ occur because of taste heterogeneity, i.e. differences in $\boldsymbol{\beta}$, and differential diligence, i.e. different threshold values K. If two hypothetical respondents with the same tastes and the same threshold value operate at different speeds for other reasons we need a more flexible mapping of time into the model. A straightforward way of capturing heterogeneity in the speed of thinking among respondents is to set $U_i\left(t\right) = t\exp\left(\lambda_i\right)$. The rate at which alternative a accumulates signals in respondent i's mind then becomes $\exp\left(\lambda_i + \mathbf{x}'_{am}\boldsymbol{\beta}_i\right)$.

Learning

If the ability of respondents to encode, process, and integrate the descriptions of the choice options changes over the course of the repeated measurements then the response time distributions across choice sets within a respondent will exhibit variation over and above the variation implied by the difference in the designs across choice sets $\mathbf{X}_1 \ldots \mathbf{X}_M$. We will approximate changes in the speed of processing across repeated measurements by setting $U_{i,m}(t) = t\exp(\lambda_{0,i} + \lambda_{1,i}m)$ where m indexes choice sets. This formulation is restrictive in functional form but flexibly accommodates differences between respondents.

Cognitive Capacity

The proposed model conceptualizes the time it takes the latent counting processes in a respondent's mind to reach a threshold. With any map $U_{i,m}(t)$ that is independent of $\mathbf{X}_m$ and conditional on model parameters, the time it takes any one latent counting process to reach a threshold is independent of the composition of choice sets. As a consequence, the expected response time conditional on model parameters is inversely related to the size of a choice set (i.e. bigger choice sets should be associated with quicker responses). Also, since the observed response time equals the time it took the best alternative to trigger K signals, decisions between equally attractive alternatives should have stochastically shorter response times than decisions between an attractive and a very much less attractive alternative. In the latter case the one attractive alternative will almost always finish first. In the former case the two attractive alternatives have the same expected finishing time and the smaller realization of two independent identically distributed random variables will be recorded as the response time.

However, these implications are inconsistent with the notion of limited cognitive resources that have to be allocated across alternatives before a decision can be made. We can characterize limited processing capacity by way of the superposition process defined by the record of all signals triggered by all of the choice options in a set *m*. This superposition process accumulates signals at the rate $U'(t)\sum_{j=1}^{J}\exp\left(\mathbf{x}'_{jm}\boldsymbol{\beta}_i\right)$. Because of capacity constraints, consumer *i* processes the stream of incoming signals only at the fixed rate $\exp(\delta_i)$. Therefore, alternative specific incoming signals, say in favor of alternative *a*, are processed at the rate $\exp(\delta_i)\exp\left(\mathbf{x}'_{am}\boldsymbol{\beta}_i\right)\Big/\left(\sum_{j=1}^{J}\exp\left(\mathbf{x}'_{jm}\boldsymbol{\beta}_i\right)\right)$; i.e. incoming signals on the superposition process compete for limited processing capacity in proportion to their relative rates that are independent of *U'(t)*. This is consistent with the existence of a bottleneck in the cognitive system (Anderson et al. 2004). Upon realizing that $\exp(\delta_i)\left(\sum_{j=1}^{J}\exp\left(\mathbf{x}'_{jm}\boldsymbol{\beta}_i\right)\right)^{-1}$ is constant for all alternatives in a choice set we set

$$U_{m,i}(t) = t\exp(\delta_i)\left(\sum_{j=1}^{J}\exp\left(\mathbf{x}'_{jm}\boldsymbol{\beta}_i\right)\right)^{-1}. \qquad (9)$$

Equation (9) constitutes a major departure from the original formulation of the model where the rate at which signals are generated and processed is independent of the composition of the choice set. Here, the number of signals processed per unit of time is assumed to be constant across choice sets. Differences in response time across choice sets within a respondent only convey information about the relative magnitude of preference within a choice set. Heterogeneity in the cognitive capacity allocated to the task is captured by the respondent-

specific maximum rate at which signals are processed, $\exp(\delta_i)$. We will approximate unobserved changes in the cognitive capacity allocated to a choice task by setting $\exp(\delta_{i,m}) = \exp(\delta_{0,i} + \delta_{1,i} m)$ where m indexes choice sets.

To simultaneously explore the model with and without cognitive constraints we formulate the combined model that nests all the specifications discussed above,

$$U_{m,i}(t) = t\left(\exp(\delta_{0,i} + \delta_{1,i} m)\left(\sum_{j=1}^{J} \exp(\mathbf{x}'_{jm}\boldsymbol{\beta}_i) \right)^{-1} + \exp(\lambda_{0,i} + \lambda_{1,i} m) \right). \quad (10)$$

In summary, observed response time in the proposed model is positively related to respondent diligence quantified by the threshold K_i, and negatively related to the rate at which alternatives in a choice set generate signals. The latter is larger for (relatively) attractive alternatives in a choice set and for respondents that allocate more cognitive capacity and thus are 'quick to think'. Thus, the model structurally motivates different response time distributions both across choice tasks and respondents.

In contrast to models that condition on observed response times as another explanatory variable (e.g., Haaijer, Kamakura and Wedel 2000), our model treats both time and choice as dependent variables. The distribution of response time is endogenous to the model and linked to the parameters that determine marginal choice probabilities. If our model is the data generating mechanism, inference that conditions on response times as another explanatory variable will be biased. Moreover, in the context of our intended application – conjoint analysis – the treatment of response times as independent variable, the density of which is left unspecified, may be

criticized. The ultimate interest is in predictive densities for choices that are not yet observed and therefore response time is missing.

3. Estimation

Equation (8a) defines the joint likelihood for respondent *i's* choice of alternative *a* from set *m* comprised of *J* alternatives at time *t*. The additional unobservables in *U'(t)=f(·)* are $\boldsymbol{\delta}_i$ and $\boldsymbol{\lambda}_i$ (see equation 10). Due to conditional independence, the joint likelihood of respondent *i's* choices, $\mathbf{y}_i$ and response times, $\mathbf{t}_i$ is:

$$p\left(\mathbf{y}_i, \mathbf{t}_i \middle| \boldsymbol{\beta}_i, K_i, \boldsymbol{\delta}_i, \boldsymbol{\lambda}_i\right) = \prod_{m=1}^{M} p\left(y_{im}, t_{im} \middle| \boldsymbol{\beta}_\mathbf{i}, K_i, \boldsymbol{\delta}_i, \boldsymbol{\lambda}_i\right) \tag{11}$$

In practice, response times recorded using non-specialized interviewing software are usually rounded to the nearest second. This is problematic if some decisions take less than .5 seconds and the observed data contain response times rounded to zero. A response time equal to zero has density zero in the present model. A convenient and statistically principled way of addressing this problem is to augment the data with unobserved exact response times $\mathbf{t}^*_\mathbf{i}$ given the rounded observations, $\mathbf{t}_\mathbf{i}$.

Since our application, conjoint analysis, usually results in 'long but narrow' data, i.e. a small number of observations per respondent relative to the number of respondents observed, we introduce hierarchical prior distributions for all respondent-specific parameters. We factor the joint posterior density of all model parameters and the exact response times given the observed choices and response times as follows

$$
\begin{aligned}
p\left(\left\{\mathbf{t}^*_{\mathbf{i}},\, \boldsymbol{\beta}_{\mathbf{i}},\, K_i, \boldsymbol{\delta}_i, \boldsymbol{\lambda}_i\right\}, \bar{\boldsymbol{\beta}}, \mathbf{Q}, \theta, \boldsymbol{\alpha}, \boldsymbol{\psi} \,\middle|\, \left\{\mathbf{t}_{\mathbf{i}},\, \mathbf{y}_{\mathbf{i}}\right\}\right) \propto & \\
\prod_{i=1}^{N} p\left(\mathbf{t}_{\mathbf{i}} \middle| \mathbf{t}^*_{\mathbf{i}}\right) \times & \\
\prod_{i=1}^{N} p\left(\mathbf{y}_{\mathbf{i}}, \mathbf{t}^*_{\mathbf{i}} \middle| \boldsymbol{\beta}_{\mathbf{i}}, K_i, \boldsymbol{\delta}_i, \boldsymbol{\lambda}_i\right) \times & \\
\prod_{i=1}^{N} p\left(\boldsymbol{\delta}_i, \boldsymbol{\lambda}_i \middle| \boldsymbol{\alpha}, \boldsymbol{\psi}\right) p\left(\boldsymbol{\alpha} \middle| \boldsymbol{\alpha}^0,\, \boldsymbol{\Sigma}_{\alpha^0}\right) p\left(\boldsymbol{\Psi} \middle| \upsilon^0_{\Psi}, \mathbf{S}^0_{\Psi}\right) \times & \qquad (12) \\
\prod_{i=1}^{N} p\left(\boldsymbol{\beta}_{\mathbf{i}} \middle| \bar{\boldsymbol{\beta}}, \mathbf{Q}\right) p\left(\bar{\boldsymbol{\beta}} \middle| \boldsymbol{\beta}^0, \boldsymbol{\Sigma}\right)\, p\left(\mathbf{Q} \middle| \upsilon^0, \mathbf{S}^0\right) \times & \\
\prod_{i=1}^{N} p\left(K_i \middle| \theta\right) p\left(\theta \middle| \theta^0, \sigma^2_{\theta^0}\right) &
\end{aligned}
$$

The term $\prod_{i=1}^{N} p\left(\mathbf{t}_{\mathbf{i}} \mid \mathbf{t}^*_{\mathbf{i}}\right)$ is a product of indicator functions, $\prod_{i=1}^{N}\prod_{m=1}^{M} I\left(t_{i,m} = round\left(t^*_{i,m}\right)\right)$, that generate the observed response times by rounding the exact response times to the nearest integer.

In the most general case investigated here, $U_{m,i}(t)$ is equal to equation (10) and $p\left(\boldsymbol{\delta}_i, \boldsymbol{\lambda}_i \middle| \boldsymbol{\alpha}, \boldsymbol{\psi}\right)$ is a multivariate normal distribution with mean $\boldsymbol{\alpha}$ and variance-covariance $\boldsymbol{\Psi}$. We use conditionally conjugate priors for $\boldsymbol{\alpha}$ and $\boldsymbol{\Psi}$, i.e. $\boldsymbol{\alpha} \sim N\left(\boldsymbol{\alpha} \middle| \boldsymbol{\alpha}^0,\, \boldsymbol{\Sigma}_{\alpha^0}\right)$, $\boldsymbol{\Psi} \sim IW\left(\boldsymbol{\Psi} \middle| \upsilon^0_{\Psi}, \mathbf{S}^0_{\Psi}\right)$, where *IW* denotes an inverted Wishart distribution. In our empirical application we use $\boldsymbol{\alpha}^0 = \mathbf{0}$, $\boldsymbol{\Sigma}_{\alpha^0} = 100 \cdot \mathbf{I}$, $\upsilon^0_{\Psi} = .5(d+1)+1$, $\mathbf{S}^0_{\Psi} = \mathbf{I}\left(\upsilon^0_{\Psi} - .5(d+1)\right)$ where d is the dimension of $\boldsymbol{\alpha}^0$ and the choices for υ^0_{Ψ} and $\mathbf{S}^0_{\Psi}$ result in a proper but diffuse *IW* prior with prior expectation equal to the identity matrix. We take the hierarchical prior distribution of the part-worths to be multivariate normal with *NIW* hyper-priors, i.e., $\boldsymbol{\beta}_{\mathbf{i}} \sim N\left(\bar{\boldsymbol{\beta}}, \mathbf{Q}\right)$; $\bar{\boldsymbol{\beta}} \sim N\left(\boldsymbol{\beta}^0, \boldsymbol{\Sigma}\right)$; and $\mathbf{Q} \sim IW\left(\upsilon^0, \mathbf{S}^0\right)$. We set $\boldsymbol{\beta}^0 = \mathbf{0}$, $\boldsymbol{\Sigma} = 100 \cdot \mathbf{I}$, and $\upsilon^0 = 7$, $\mathbf{S}^0 = \frac{1}{10}\mathbf{I}\left(\upsilon^0 - (11+1)/2\right)$. With the

11-dimensional part-worth vector of our empirical application the particular choices for υ^0 and $\mathbf{S}^0$ again result in a proper but diffuse *IW*-distribution with prior expectation equal to the identity matrix divided by 10. We assume that the hierarchical prior distribution of the quantity 'threshold minus one' – the smallest possible threshold is one – is Poisson with rate parameter θ, i.e., $p(K_i - 1|\theta) = Poisson(\theta)$. We specify a log-normal hyper-prior for the rate parameter θ, $\log(\theta) \sim N(\theta^0, \sigma^2_{\theta^0})$. We set $\theta^0 = 0$ and $\sigma^2_{\theta^0} = 10$ which puts considerable prior mass on thresholds $K = 1$ and at the same time readily accommodates even extreme departures from this special case.

Our choice of independent priors for the part-worths $\{\boldsymbol{\beta}_i\}$ and the thresholds $\{K_i\}$ is motivated by the conceptual independence between tastes $\{\boldsymbol{\beta}_i\}$ and diligence or effective effort as quantified by $\{K_i\}$. We also assume prior independence between the parameters $\{\boldsymbol{\delta}_i, \boldsymbol{\lambda}_i\}$ that quantify cognitive capacity / speed of processing and effective effort. We will explore departures from our assumption of prior independence in the posterior.

Model parameters are estimated using Markov chain Monte Carlo (MCMC) methods. Algorithmic details are contained in appendix A. We use burn-in samples of 10,000 iterations and base our inferences on overall 50,000 iterations from 10 approximately independent chains for each model. When we report numerical standard errors these are based on the variation of posterior point estimates across the 10 approximately independent replications of the MCMC chain for each model. Convergence is assessed informally through time-series plots of the parameters that form the hierarchical prior. Additionally, the combined view of the results from 10 approximately independent chains started from different, randomly chosen starting values provides some empirical support for the absence of pronounced local maxima in the posterior.

4. The Data

The data were supplied by a leading research company and contain responses of 422 respondents to 18 *respondent-specific* choice sets comprised of 5 choice alternatives each. Each alternative was described along six attributes. The products were television sets, and the attributes were brand name, screen size, sound system, channel blackout, picture-in-picture, and price. Attribute-levels are listed in table 4. Respondents had to pick one alternative from each set.

The marginal distribution of response times is depicted in figure 1. The mean of the distribution is 13.44 seconds. Around 4.5% of all 7,596 observed response times were recorded as zero due to rounding to the nearest integer second. A total of 64 respondents had at least one response time recorded as zero. The maximum observed time is 341 seconds. Because response times of zero have density of zero in our model we augment the data with exact, unobserved response times (see section 3 and appendix A).

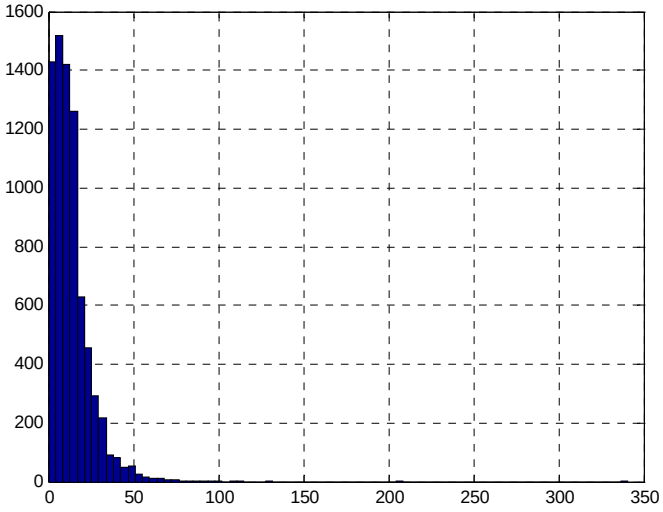


Figure 1 Marginal distribution of response times

Figure 2a shows 18 box-plots of the response time distribution as a function of the choice (the first through the 18th). There is a tendency for responses to get faster as the conjoint task progresses. In estimation we discard the first two choice tasks as training data and use the last two choice tasks as holdout data. Figure 2b presents the empirical distribution of the mean of the respondent-specific response times, where the mean is taken over all 18 choices. Together these graphs motivate a respondent and task-specific map *U(t)*.

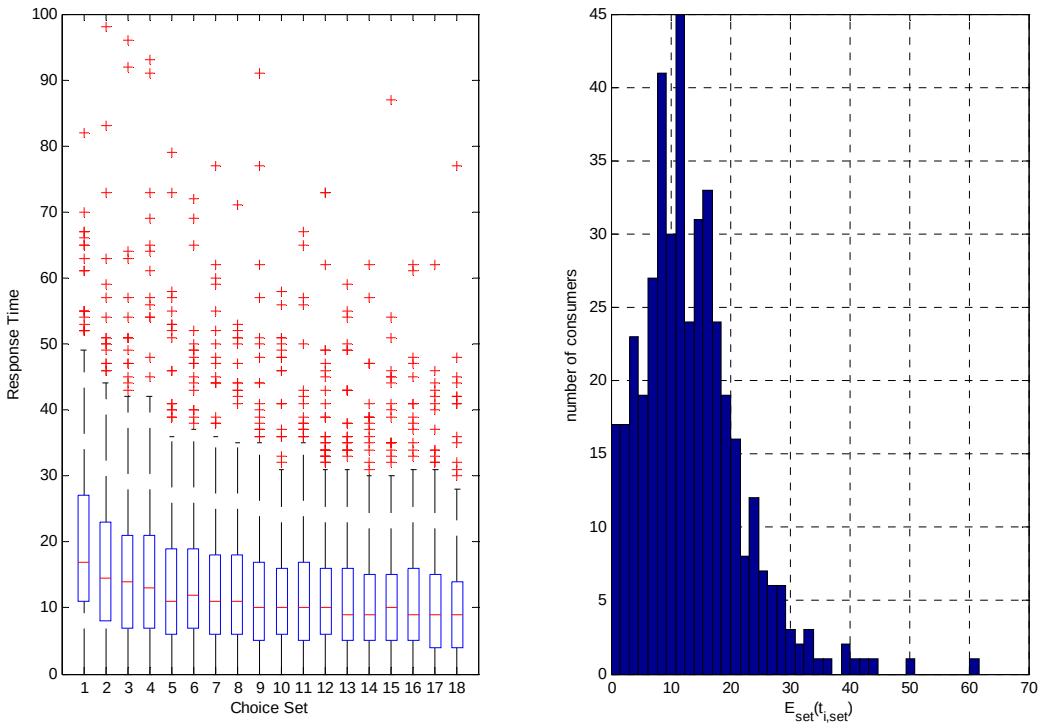


Figure 2a Response time conditional on choice set 2b Respondent heterogeneity in response time

5. Model Fit and Parameter Estimates

Table 2a and table 2b summarize respectively the in-sample fit and the holdout performance of various models applied to data. The in-sample fit statistic, log-marginal density (LMD), is based on the harmonic mean estimator (Newton and Raftery 1994). The columns of table 2a correspond to the LMD of the observed choices and response times given the model, $\log p(\{\mathbf{y}_i, \mathbf{t}_i\} \mid M)$, the LMD of the observed choices given the model, $\log p(\{\mathbf{y}_i\} \mid M)$, and an estimate of the log posterior density of the choice data only, i.e. $\log p\left(\{\mathbf{y}_i\} \middle| \{\mathbf{y}_i, \mathbf{t}_i\}, M\right)$.

Note that by definition $\log p(\{\mathbf{y}_i\} \mid M)$, i.e. the LMD of the choice data conditional on the model, is independent of the way time enters the model assuming prior independence between parameters in $p\left(\{\boldsymbol{\beta}_i, K_i\}, \bar{\boldsymbol{\beta}}, \mathbf{Q}, \theta\right)$ and $p\left(\{\boldsymbol{\delta}_i, \boldsymbol{\lambda}_i\}, \boldsymbol{\alpha}, \boldsymbol{\Psi}\right)$. This is because marginal choice probabilities depend on $\{\boldsymbol{\beta}_i, K_i\}$, only (see equation (4)). However, observed response times and the way these enter the model inform the marginal posterior $p\left(\{\boldsymbol{\beta}_i, K_i\}, \bar{\boldsymbol{\beta}}, \mathbf{Q}, \theta \middle| \{\mathbf{y}_i, \mathbf{t}_i\}\right)$ via the likelihood. To the extent that the likelihood captures information about $\{\boldsymbol{\beta}_i, K_i\}, \bar{\boldsymbol{\beta}}, \mathbf{Q}, \theta$ contained in the response times, $\log p\left(\{\mathbf{y}_i\} \middle| \{\mathbf{y}_i, \mathbf{t}_i\}, M\right)$ will dominate $\log p\left(\{\mathbf{y}_i\} \middle| \{\mathbf{y}_i\}, M\right)$, where response time is not used in estimation.

The first model is a hierarchical Bayes logit model (HB logit), i.e. the special case where all K_i are equal to one. The second model generalizes the logit to the case where the thresholds K_i may be greater than 1. Neither model attempts to integrate observed response times and both use only the choice data to estimate part-worths and (in the case of the second model) thresholds.

The third and the fourth model employ a time map *U(t)* that is independent of choice set composition where respondents' cognitive capacity is assumed to be unlimited. Model three

accounts for differential speed of processing among respondents and model four additionally allows for respondent specific log-linear changes in the speed of processing from choice set to choice set.

Models five and six employ a time map *U(t)* that is a function of choice set composition and that assumes respondents have limited cognitive resources to be allocated across alternatives in a set. Model five accounts for heterogeneity in the number of signals that can be processed per unit of time. Model six additionally allows for respondent specific log-linear changes in the maximum number of signals that can be processed.

Models seven through ten represent all possible combinations of models three to six. Model 10 is the most general formulation that combines models four and eight.

This collection of models allows us to explore – of course within the boundaries of the set of models investigated – the 'net contribution' of response times to posterior knowledge about the probability surface that generated the choices (model 2 versus models 3 to 10). It also allows us to test whether respondents' limited cognitive capacity becomes relevant in these choice tasks or processing with unlimited capacity is a tenable assumption given the task at hand (model 5 versus model 3). The comparisons between models 3 and 4 and models 5 and 6 shed light on training effects that may have occurred over the course of the repeated measurements. Finally, the comparison between the HB-logit model and model two (which relaxes the assumption that all K_i are equal to one but does not fit response times), illustrates that the two models may be empirically distinguished with choice data alone.

The difference in the LMD of the observed choices between model one and model two provides evidence against the logit model and in favor of the generalization with at least some K_i

different from one as data generating mechanism.[1] Comparing LMDs of the observed choices and response times across models 3 to 10 we see that limited cognitive capacity is supported over unlimited capacity (model 5 versus 3; model 6 versus 4). We also find evidence for change in the speed of processing across the repeated measurements (model 4 versus 3; model 6 versus 5; and more generally model 7 versus models 6, 8, 9, and 10).

	Choice & Time $\log p(\mathbf{y}, \mathbf{t} \mid \boldsymbol{M})$	Choice $\log p(\mathbf{y} \mid \boldsymbol{M})$	Choice $\log p\left(\{\mathbf{y_i}\} \middle\vert \{\mathbf{y_i}, \mathbf{t_i}\}, M\right)$
Fit Choice Data Only			
1. *HB Logit* $(K \equiv 1)$	NA	-4,825 (8)	-4,428 (8)
2. K estimated w/o response times	NA	-4,743 (9)	-4,362 (9)
Jointly Fit Choice and Response Times			
3. K estimated, with $U_{im}(t) = exp(\lambda_{0,i})\ t$	-21,996 (6)	-4,743	-5,002 (8)
4. K estimated, with $U_{im}(t) = exp(\lambda_{0,i}+\lambda_{1,i}m)\ t$	-21,702 (14)	-4,743	-5,003 (4)
5. K estimated, with $U_{im}(t) = exp(\delta_{0,i})\ (\Sigma_j\ exp(\mathbf{x}'_{jm}\boldsymbol{\beta}_i))^{-1}\ t$	-21,425 (13)	-4,743	-4,336 (8)
6. K estimated, with $U_{im}(t) = exp(\delta_{0,i} + \delta_{1,i}m)\ (\Sigma_j\ exp(\mathbf{x}'_{jm}\boldsymbol{\beta}_i))^{-1}\ t$	-21,155 (11)	-4,743	-4,364 (9)
7. K estimated, with $U_{im}(t) = exp(\delta_{0,i})\ (\Sigma_j\ exp(\mathbf{x}'_{jm}\boldsymbol{\beta}_i))^{-1}\ t + exp(\lambda_{0,i})\ t$	-21,409 (11)	-4,743	-4,368 (6)
8. K estimated, with $U_{im}(t) = exp(\delta_{0,i})\ (\Sigma_j\ exp(\mathbf{x}'_{jm}\boldsymbol{\beta}_i))^{-1}\ t + exp(\lambda_{0,i}+\lambda_{1,i}m)\ t$	-21,142 (13)	-4,743	-4,450 (6)
9. K estimated, with $U_{im}(t) = exp(\delta_{0,i} + \delta_{1,i}m)\ (\Sigma_j\ exp(\mathbf{x}'_{jm}\boldsymbol{\beta}_i))^{-1}\ t + exp(\lambda_{0,i})\ t$	-21,102 (10)	-4,743	-4,356 (7)
10. K estimated, with $U_{im}(t) = exp(\delta_{0,i} + \delta_{1,i}m)\ (\Sigma_j\ exp(\mathbf{x}'_{jm}\boldsymbol{\beta}_i))^{-1}\ t + exp(\lambda_{0,i}+\lambda_{1,i}m)\ t$	-21,069 (10)	-4,743	-4,317 (11)

Table 2a In sample fit statistics; numerical standard errors in parentheses

[1] As noted earlier the LMD of the observed choices is the same for models 2 to 10 investigated here. Since the performance of the harmonic mean estimator heavily depends on the mixing properties of the MCMC chain used to sample from the posterior, we decided to base our numerical estimate of $\log p(\mathbf{y} \mid \boldsymbol{M})$ for models 2 to 10 on the posterior from the chain of smallest dimensionality in terms of the number of unobservables. We therefore compute only one numerical estimate of $\log p(\mathbf{y} \mid \boldsymbol{M})$ using the posterior from model 2.

The model that fits the data best is model 10. This model allows processing to be either limited or unlimited capacity and accounts for change in the speed of processing over the course of the repeated measurements. We will return to this point later when we discuss estimates of parameters in the various time maps *U(t)*. The last column of table 2a reports numerical estimates of the log posterior density of the choice data. Models 5, 6, 7, 9, and 10 fit the observed choices at least as well or better than the corresponding model 2 that did not fit response times, which is strong evidence for information in response times about respondent tastes $\boldsymbol{\beta}_i$ and respondent diligence K_i. Models 3 and 4 that assume processing with unlimited capacity fit the choice considerably worse than model 2.

Table 2b summarizes predictive performance. The first column reports hit-rates, i.e. $\mathrm{E}_{i,m}\ I\left(\mathbf{x}_{\left(y_{im}^h\right)}'\boldsymbol{\beta}_i = \max\left\{\mathbf{x}_{jm}'\boldsymbol{\beta}_i\right\}_{j=1}^{J}\middle|\mathbf{D},M\right)$ where *i* indexes respondents, *m* holdout sets, and $\mathbf{x}_{\left(y_{im}^h\right)}$ is the design vector of y_{im}^h, the alternative chosen by the *i*-th respondent in the *m*-th holdout set. **D** stands for the data used to calibrate model *M*. The second column summarizes hit-probabilities defined as $\mathrm{E}_{i,m}\ P\left(y_{im}^h\middle|\mathbf{D},M\right)$. The third and the fourth column of table 2b contain estimates of the log predictive densities $\log p\left(\left\{\mathbf{y}_\mathbf{i}^h\right\}\middle|\mathbf{D},M\right)$ and $\log p\left(\left\{\mathbf{y}_\mathbf{i}^h,\mathbf{t}_\mathbf{i}^h\right\}\middle|\mathbf{D},M\right)$, respectively.

	Hit-rate	Hit-probability	Log Predictive Density (Choice)	Log Predictive Density (Choice & Time)
Fit Choice Data Only				
1. *HB Logit* *(K ≡ 1)*	.5410 (.0004)	.4761 (.0003)	-984 (3)	NA
2. K estimated w/o response times	.5398 (.0012)	.4774 (.0008)	-999 (4)	NA
Jointly Fit Choice and Response Times				
3. K estimated, with $U_{im}(t) = exp(\lambda_{0,i})\ t$	.5240 (.0004)	.4351 (.0004)	-939 (2)	-3,524 (2)
4. K estimated, with $U_{im}(t) = exp(\lambda_{0,i}+\lambda_{1,i}m)\ t$	.5260 (.0004)	.4381 (.0003)	-934 (3)	-3,551 (11)
5. K estimated, with $U_{im}(t) = exp(\delta_{0,i})\ (\Sigma_j\ exp(\mathbf{x}'_{jm}\boldsymbol{\beta}_i))^{-1}\ t$	.5427 (.0005)	.4848 (.0008)	-964 (3)	-3,472 (6)
6. K estimated, with $U_{im}(t) = exp(\delta_{0,i} + \delta_{1,i}m)\ (\Sigma_j\ exp(\mathbf{x}'_{jm}\boldsymbol{\beta}_i))^{-1}\ t$	.5430 (.0006)	.4841 (.0006)	-961 (3)	-3,477 (9)
7. K estimated, with $U_{im}(t) = exp(\delta_{0,i})\ (\Sigma_j\ exp(\mathbf{x}'_{jm}\boldsymbol{\beta}_i))^{-1}\ t + exp(\lambda_{0,i})\ t$	.5448 (.0006)	.4854 (.0005)	-971 (2)	-3,488 (4)
8. K estimated, with $U_{im}(t) = exp(\delta_{0,i})\ (\Sigma_j\ exp(\mathbf{x}'_{jm}\boldsymbol{\beta}_i))^{-1}\ t + exp(\lambda_{0,i}+\lambda_{1,i}m)\ t$	.5408 (.0007)	.4783 (.0005)	-945 (5)	-3,509 (11)
9. K estimated, with $U_{im}(t) = exp(\delta_{0,i} + \delta_{1,i}m)\ (\Sigma_j\ exp(\mathbf{x}'_{jm}\boldsymbol{\beta}_i))^{-1}\ t + exp(\lambda_{0,i})\ t$	.5426 (.0005)	.4841 (.0006)	-968 (3)	-3,508 (4)
10. K estimated, with $U_{im}(t) = exp(\delta_{0,i} + \delta_{1,i}m)\ (\Sigma_j\ exp(\mathbf{x}'_{jm}\boldsymbol{\beta}_i))^{-1}\ t + exp(\lambda_{0,i}+\lambda_{1,i}m)\ t$	.5443 (.0007)	.4851 (.0008)	-965 (4)	-3,474 (8)

Table 2b Performance on holdout data; numerical standard errors in parentheses

The model chosen based on the LMD of choices and response times, model 10 fits the holdout data well, too. Specifically, it outperforms both the logit model and model 2 (which generalizes K but did not fit response times).

The Distribution of the Threshold K_i and of Parameters in U(t)

Table 3 summarizes information about the distribution of the thresholds K_i and of the parameters in the various maps *U(t)* across respondents. The first column reports the posterior mean of the quantity θ + 1, the posterior mean of the thresholds across respondents. The distribution of the threshold is centered away from 1 in all cases where we estimate *K* which speaks against the logit model as a data generating mechanism.

	$E_i(K_i) = \theta + 1$ (posterior std.dev.)	λ_0 (ψ)	λ_1 (ψ)	δ_0 (ψ)	δ_1 (ψ)
1. *HB-Logit (K ≡ 1)*	1*	NA	NA	NA	NA
2. K estimated w/o response times	4.15 (.53)	NA	NA	NA	NA
3. K estimated, with $U_{im}(t) = exp(\lambda_{0,i})\ t$	4.82 (.13)	-1.43 (.67)	0*	-inf*	-inf*
4. K estimated, with $U_{im}(t) = exp(\lambda_{0,i}+\lambda_{1,i}m)\ t$	5.32 (.13)	-1.53 (.76)	2.42 (5.83)	-inf*	-inf*
5. K estimated, with $U_{im}(t) = exp(\delta_{0,i})\ (\Sigma_j\ exp(\mathbf{x}'_{jm}\boldsymbol{\beta}_i))^{-1}\ t$	4.55 (.13)	-inf*	-inf*	-.37 (.61)	0*
6. K estimated, with $U_{im}(t) = exp(\delta_{0,i} + \delta_{1,i}m)\ (\Sigma_j\ exp(\mathbf{x}'_{jm}\boldsymbol{\beta}_i))^{-1}\ t$	4.97 (.14)	-inf*	-inf*	-.48 (.71)	2.62 (6.22)
7. K estimated, with $U_{im}(t) = exp(\delta_{0,i})\ (\Sigma_j\ exp(\mathbf{x}'_{jm}\boldsymbol{\beta}_i))^{-1}\ t + exp(\lambda_{0,i})\ t$	4.56 (.13)	-5.28 (.52)	0*	-.42 (.64)	0*
8. K estimated, with $U_{im}(t) = exp(\delta_{0,i})\ (\Sigma_j\ exp(\mathbf{x}'_{jm}\boldsymbol{\beta}_i))^{-1}\ t + exp(\lambda_{0,i}+\lambda_{1,i}m)\ t$	5.01 (.14)	-4.40 (2.20)	5.90 (14.79)	-.55 (.80)	0*
9. K estimated, with $U_{im}(t) = exp(\delta_{0,i} + \delta_{1,i}m)\ (\Sigma_j\ exp(\mathbf{x}'_{jm}\boldsymbol{\beta}_i))^{-1}\ t + exp(\lambda_{0,i})\ t$	4.97 (.13)	-5.22 (.51)	0*	-.54 (.77)	2.70 (5.04)
10. K estimated, with $U_{im}(t) = exp(\delta_{0,i} + \delta_{1,i}m)\ (\Sigma_j\ exp(\mathbf{x}'_{jm}\boldsymbol{\beta}_i))^{-1}\ t + exp(\lambda_{0,i}+\lambda_{1,i}m)\ t$	4.95 (.14)	-6.17 (2.27)	-1.24 (1.05)	-.54 (.68)	2.71 (4.97)

Table 3 Hierarchical prior distribution of the threshold *K;* Parameters in *U(t)*, posterior means, variance across respondents in parentheses;
* fixed parameters

We report point estimates of parameters in the hierarchical prior that correspond to means and variances of respondent specific $\{\boldsymbol{\lambda}_i, \boldsymbol{\delta}_i\}$ in columns 2 through 5 of table 3. All specifications with free parameters show that signals are, on average, accumulated at a smaller rate than implied by the simple map $U(t) = t$, the average λ_0 and δ_0 are negative. Heterogeneity in λ_0 and δ_0, however, indicates that some respondents come close to the rate implied by the identity map $U(t) = t$ and a small proportion of extreme respondents accumulates signals at an even faster rate. When λ_1 and δ_1 are estimated their mean values across respondents are positive with the exception of λ_1 in model 10. Choice sets were indexed by $m = .01, \ldots, .18$ and respondents, on average, pick up processing speed over the course of the conjoint task. There is, however, considerable heterogeneity in the slope parameters λ_1 and δ_1. Some respondents actually became slower over the course of the conjoint task, possibly because of a change in the amount of cognitive capacity allocated to the task.

From table 2a the best fitting model overall is model 10. A positive δ_1 in this model shows that respondents, on average, pick up speed related to processing under a capacity constraint, i.e. they allocate more cognitive capacity and/or become more proficient. However, the speed at which they process without constraint on average decreases over the course of the repeated measurements, λ_1 is negative.

We can interpret the limited and unlimited capacity components of the model by speculating that respondents can process choice alternatives in two ways. Processing under a capacity constraint is similar to what a respondent does when she evaluates the strictly local relative attractiveness in one choice task. Processing without constraint is similar to what a respondent does when she simply matches the alternatives to templates of good alternatives stored in memory. Memory-based processing as represented here is independent of the strictly

local context of a particular choice task. Over the course of the repeated measurements the processing component related to producing evidence in favor of the strictly locally best alternative in a particular choice set gains importance over the memory based component.

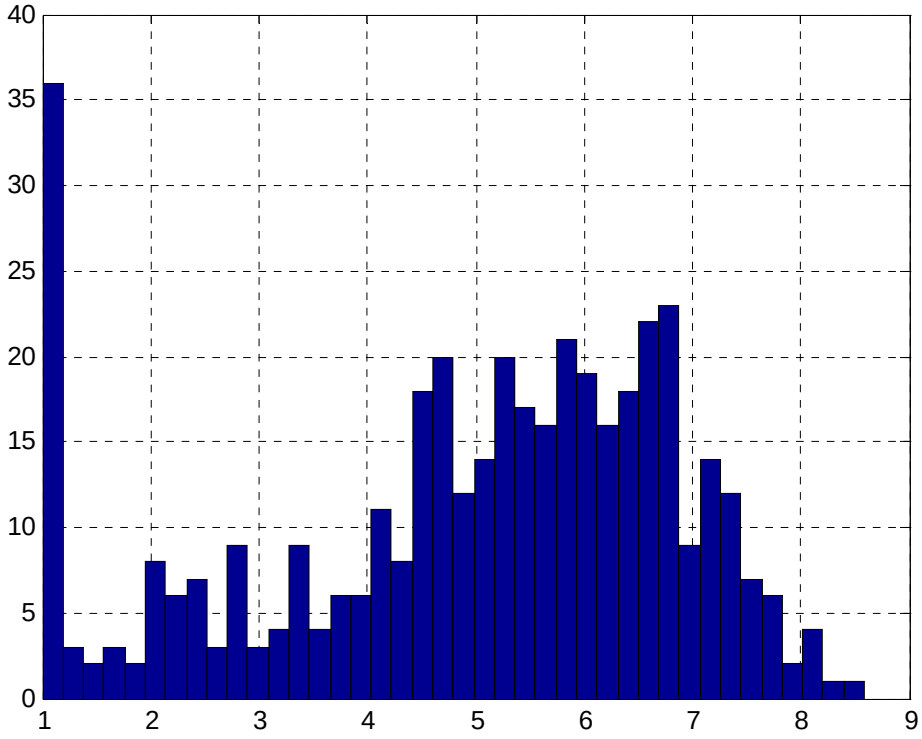


Figure 4 Posterior distribution of expected thresholds K_i in the sample, model 10

Figure 4 depicts the posterior distribution of the threshold K_i obtained from model 10. The majority of respondents in the sample (88.5 %) are characterized by thresholds greater than 1, i.e. the 90 % posterior credible intervals exclude the value 1. The percentages of 'logit-respondents', who are characterized by 90 % posterior credible intervals for K_i that include 1 range from 11 % (model 10) to 45 % (model 2). Response times contribute significantly to the information in any individual respondent's marginal posterior distribution of K_i. However, in

combination with the results reported earlier (see column 1 of table 3) it is evident that response times are not the primary reason why the average threshold is substantively larger than 1.

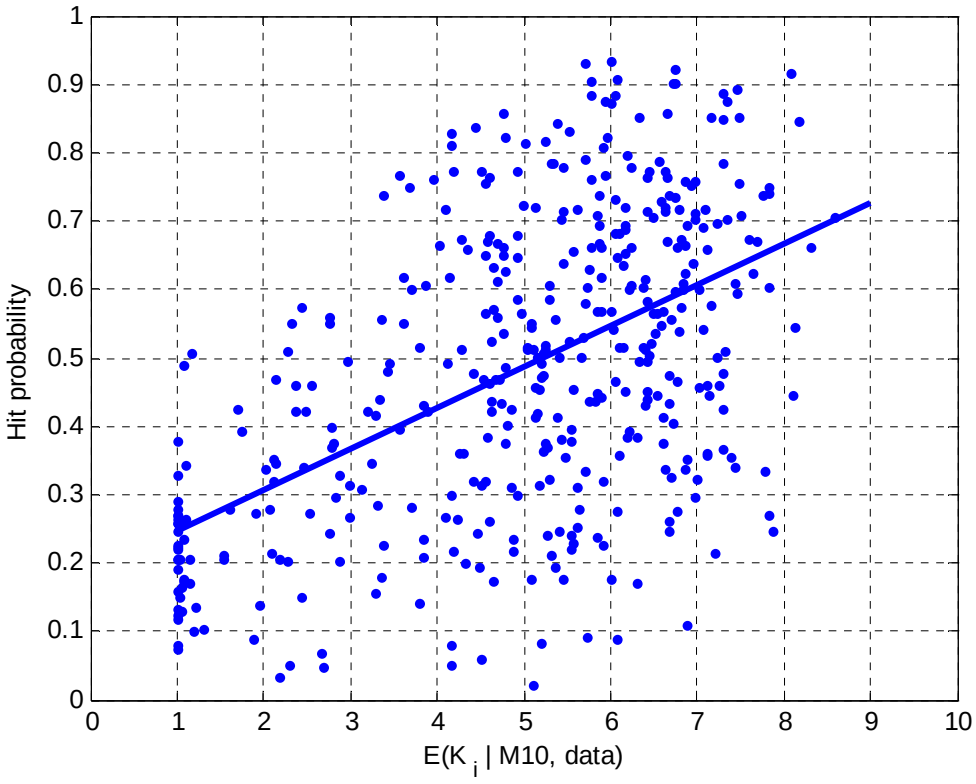


Figure 5 K_i versus Hit-probability, model 10

In section 2 we characterized choices made under a large threshold K_i as well deliberated, in the sense that respondents operating with a large threshold K_i collect more evidence before making a choice than respondents that operate with a smaller threshold. Our empirical analysis provides evidence for this interpretation, where we find that respondents with small thresholds behave less consistently. In the case of model 10, for example, respondent-specific thresholds K_i and holdout hit-probabilities are positively correlated with a Pearson correlation of .53 (see figure 5; regressing the hit-probability on the threshold K_i we expect an increase of 6% in hit-

probability for each additional signal required to make a decision; $R^2 = 27.6$ %). Note that the HB-logit constrains K to be equal to 1 and therefore does not measure respondents' diligence independently of their preference structure.

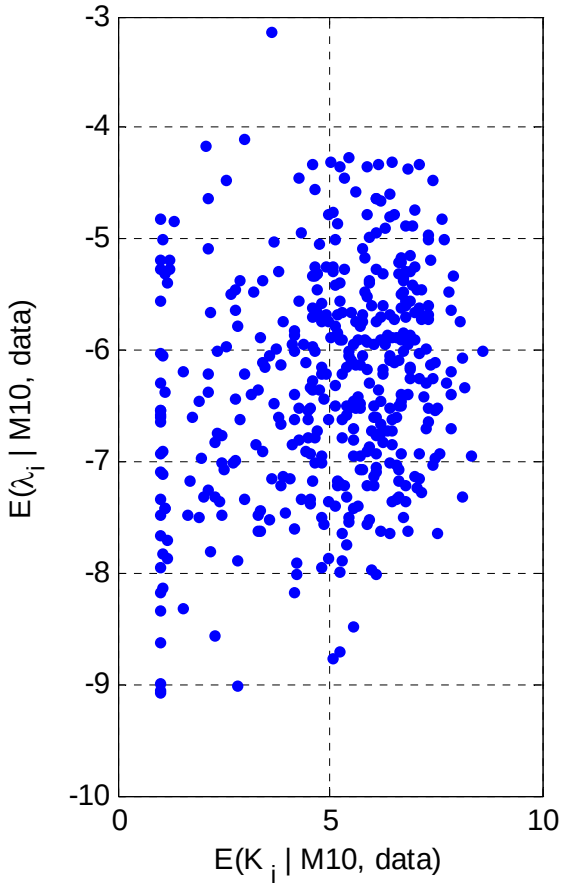
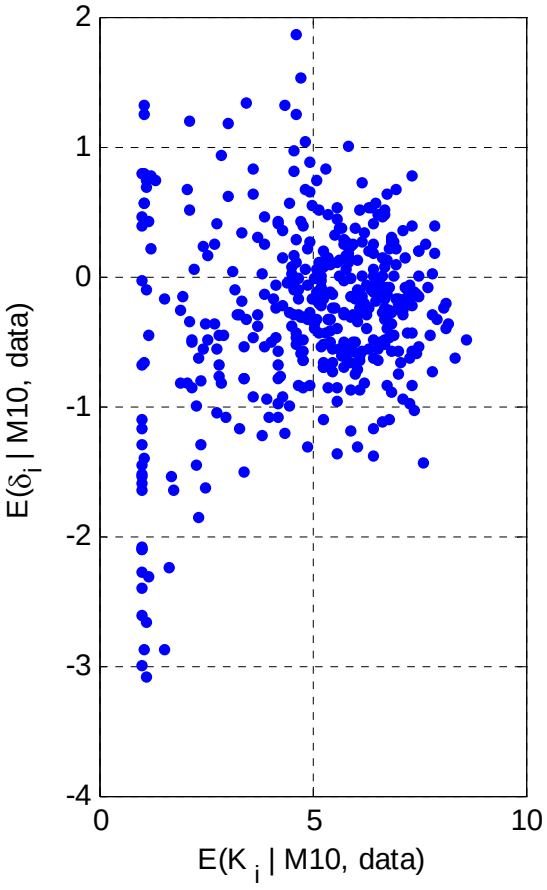


Figure 6a K_i versus rate of memory based processing



6b K_i versus maximum rate of constrained processing

Figures 6a and 6b plot posterior expectations of K_i against posterior expectations of $E_m(\lambda_{0,i} + \lambda_{1,i}\, m)$ and $E_m(\delta_{0,i} + \delta_{1,i}\, m)$, respectively, where m indexes choice sets. The plot supports the prior assumption of independence between the rates of processing, i.e. the parameters in $U(t)$, and $\{K_i\}$. All respondents are characterized by $E_m(\lambda_{0,i} + \lambda_{1,i}\, m)$ considerably smaller than zero and relatively smaller than the corresponding $E_m(\delta_{0,i} + \delta_{1,i}\, m)$. This is in line

with our earlier conclusion that, in our application, processing under the capacity constraint is more important for signal generation than unconstrained, memory based processing. Interestingly the dispersion in $E_m(\delta_{0,i} + \delta_{1,i}\, m)$ across respondents appears to increase as K_i becomes smaller. Especially some of the 'logit- respondents' with a posterior $K_i = 1$ are characterized by very small maximum processing capacities; their $E_m(\delta_{0,i} + \delta_{1,i}\, m)$ are highly negative. These are respondents that take relatively long to generate and process signals and make choices (small $E_m(\delta_{0,i} + \delta_{1,i}\, m)$) that still do not appear to be well deliberated ($K_i = 1$), i.e. they are slow to think but quick to judge. This result graphically illustrates that response times are a function of both a respondent's diligence and his cognitive capacity. Differences in raw response times across respondents may be a result of differences in diligence and/or cognitive capacity, assuming equal preference structures.

Part-worths and their distribution

Table 4 compares estimates for the means and variances of the part-worths obtained from model 10 and the HB-logit. On average, part-worth estimates from the integrated model are closer to zero, and their variance in the population is estimated smaller than by the logit model. This pattern is in line with the notion that as the threshold, K, increases, choice probabilities increase in magnitude for the best alternative and decrease to zero for the other choice alternatives. Heterogeneity in the thresholds K_i results in heterogeneous choice probabilities even if the underlying rates are very similar in the population. Because the thresholds K_i, however, quantify respondents' differential diligence in making their choices independent of the attractiveness of the alternatives under study, the HB-logit model confounds diligence and attractiveness and thus, distorts the probability surface.

		Model 10		HB-logit, *(K ≡ 1)*	
	Attribute Levels	$\overline{\beta}$ (posterior std.dev.)	$diag(Q)$ (posterior std.dev.)	$\overline{\beta}$ (posterior std.dev.)	$diag(Q)$ (posterior std.dev.)
Attribute 1	JVC	-0.36 (0.04)	0.40 (0.05)	-0.86 (0.09)	2.45 (0.30)
	RCA	-0.22 (0.04)	0.51 (0.06)	-0.52 (0.10)	3.19 (0.36)
	Sony	Base = 0			
Attribute 2	25" screen	-0.40 (0.03)	0.21 (0.03)	-1.00 (0.08)	1.40 (0.19)
	26" screen	-0.20 (0.02)	0.09 (0.02)	-0.44 (0.06)	0.59 (0.10)
	27" screen	Base = 0			
Attribute 3	Mono sound	-0.78 (0.05)	0.72 (0.09)	-2.02 (0.13)	4.83 (0.55)
	Stereo sound	-0.18 (0.03)	0.19 (0.03)	-0.47 (0.07)	1.27 (0.16)
	Surround sound	Base = 0		Base = 0	
Attribute 4	No channel blockout	-0.34 (0.03)	0.28 (0.03)	-0.84 (0.08)	1.71 (0.20)
	Channel blockout	Base = 0			
Attribute 5	No picture in picture	-0.49 (0.03)	0.30 (0.03)	-1.23 (0.08)	1.95 (0.22)
	Picture in picture	Base = 0			
Attribute 6	$300	0.72 (0.05)	0.58 (0.07)	1.88 (0.12)	3.94 (0.48)
	$350	0.63 (0.04)	0.31 (0.04)	1.58 (0.10)	2.22 (0.32)
	$400	0.28 (0.03)	0.08 (0.02)	0.72 (0.08)	0.59 (0.15)
	$450	Base = 0			

Table 4 Part-worths and their distribution

In a posterior analysis we find only small correlations between the posterior expectations of the individual thresholds K_i and the part-worths (Table 5). Diligence and tastes can therefore be empirically distinguished. Also the assumption of prior independence between $\{K_i\}$ and $\{\boldsymbol{\beta}_i\}$ appears to be reasonable.

	Attribute Levels	$C(\beta_i, K_i)$ (Model 10)
Attribute 1	JVC	-0.05
	RCA	-0.02
	Sony	Base = 0
Attribute 2	25" screen	-0.08
	26" screen	0.01
	27" screen	Base = 0
Attribute 3	Mono sound	-0.28
	Stereo sound	-0.18
	Surround sound	Base = 0
Attribute 4	No channel blockout	-0.13
	Channel blockout	Base = 0
Attribute 5	No picture in picture	-0.16
	Picture in picture	Base = 0
Attribute 6	$300	0.21
	$350	0.24
	$400	0.19
	$450	Base = 0

Table 5 Pearson correlations between part-worths and thresholds across respondents

6. Discussion

Our integrated model of choice and response times generalizes the standard logit model through the threshold coefficient, *K*, which is related to the diligence of respondents in their decision making. As *K* moves away from one, the structure of the implied probability surface changes and the logit model can only approximate that structure. Response time is incorporated into the model by exploiting the relationship between the cumulative distribution of a counting process and the cumulative distribution of the sum of the inter-arrival times.

Response time is completely endogenous to the system described. Long response times, *conditional* on choice set composition, cognitive capacity allocated to the task, and tastes are

associated with greater diligence and thresholds (K) leading to greater precision in posterior inference relative to respondents with short response times and small thresholds.

The way that time t is related to the rate at which signals are generated via $U(t)$ is an influential component of the model. Our empirical analysis supports the idea that response times mainly speak to the relative attractiveness of an alternative within a choice set which is consistent with the notion of a cognitive bottleneck. We also found tentative evidence for an unlimited capacity process that we speculate is driven by memory based recognition of good alternatives.

Our best integrated model had better in-sample fit and better out of sample performance than the HB-logit model and we documented that response times are informative about respondents' preferences, even though we had to rely on crude measures (rounded to the nearest second). However, we think that more importantly this style of modeling offers insights into the way respondents make these choices that are not accessible otherwise. For example, our findings imply that it takes respondents longer to choose from a choice set with equally attractive alternatives because they end up allocating limited cognitive capacity evenly across alternatives in this case. We take this research as a first step in a direction that holds promise of deeper insights into consumer decision making.

7. Concluding Remarks

Consumer behavior researchers have associated response times to constructs such as strength of preference (Aaker et al. 1980, MacLachlan and Myers 1983), accessibility of attitudes (Park and Hastak 1994), confidence (MacLachlan et al. 1979), and conflict (Tyebjee 1979). While our model assumes that the threshold parameter, K, is fixed, future research using the

integrated model can address the relationship of these constructs to changes in the threshold parameter. Conflict can be resolved, for example, by increasing the threshold value so that the difference between the signal counters of the best and second best alternative itself reaches a threshold. Similarly, the confidence with which respondents provide information can be associated with the difference in the values of the signal counters.

The primitive assumptions of counting processes give rise to an integrated model of choice and response times that describes how respondents make decisions. The advantage of this approach is that it provides the ability to derive inferences about theoretical constructs such as diligence and processing capacity. The disadvantage is that the explicit assumptions made about the process may be invalid in some contexts, and more descriptive models may provide more robust estimates. Future research will help to identify the boundaries of applications of the proposed model.

Appendix A - MCMC Algorithms

With observed response times as data

We describe the MCMC algorithm with the time map $U_{i,m}(t) = t\left(\exp\left(\delta_{0,i} + \delta_{1,i} m\right)\left(\sum_{j=1}^{J} \exp\left(\mathbf{x}'_{jm} \boldsymbol{\beta}_i \right) \right)^{-1} + \exp\left(\lambda_{0,i} + \lambda_{1,i} m\right) \right)$. Algorithmically, all the other maps from t into $U(t)$ discussed are special cases.

Conditional on starting values we first sample from

$$p\left(t^*_{i,m} \middle| \bullet\right) \propto p\left(t_{i,m} \middle| t^*_{i,m}\right) p\left(y_{i,m}, t^*_{i,m} \middle| \boldsymbol{\beta}_i, \boldsymbol{\delta}_i, \boldsymbol{\lambda}_i, K_i\right)$$

using a Metropolis-Hastings algorithm. We generate proposals from the first factor in the likelihood (equation 8a) given the observed response times and other parameters

$$Gamma\left(K_i, \exp\left(\mathbf{x}'_{i,m} \boldsymbol{\beta}_i\right) U'_{i,m}\left(t^*_{i,m}\right)\right) \times I\left(t_{i,m} - .5 \le t^*_{i,m} < t_{i,m} + .5\right)$$

using the inverse cdf method. Note that $U'_{i,m}\left(t^*_{i,m}\right)$ is independent of $t^*_{i,m}$ and $\mathbf{x}_{im}$ is the design vector of the alternative chosen by respondent i from choice set m, i.e. y_{im}. The acceptance ratio of the Metropolis step then involves only the second factor in the likelihood (equation 8a).

The second block of the sampler generates $\boldsymbol{\beta}_\mathbf{i}, \boldsymbol{\delta}_\mathbf{i}, \boldsymbol{\lambda}_\mathbf{i}$ from

$$p\left(\boldsymbol{\beta}_\mathbf{i}, \boldsymbol{\delta}_i, \boldsymbol{\lambda}_i \middle| \bullet\right) \propto p\left(\mathbf{y}_\mathbf{i}, \mathbf{t}^*_\mathbf{i} \middle| \boldsymbol{\beta}_\mathbf{i}, \boldsymbol{\delta}_i, \boldsymbol{\lambda}_i, K_i\right) p\left(\boldsymbol{\beta}_\mathbf{i} \middle| \bar{\boldsymbol{\beta}}, \mathbf{Q}\right) p\left(\boldsymbol{\delta}_i, \boldsymbol{\lambda}_i \middle| \boldsymbol{\alpha}, \boldsymbol{\Psi}\right)$$

using a Metropolis-Hastings algorithm with normal random walk proposal densities. The evaluation of the likelihood (see equations (8a) and (11)) that consists of independent Gamma

densities and independent Poisson cumulative distribution functions can be efficiently vectorized in packages like R or Matlab.

The third block of the sampler generates the thresholds K_i from

$$p\left(K_i \middle| \bullet\right) \propto p\left(\mathbf{y_i}, \mathbf{t^*_i} \middle| \boldsymbol{\beta}_\mathbf{i}, \boldsymbol{\delta}_i, \boldsymbol{\lambda}_i, K_i\right) p\left(K_i - 1 \middle| \theta\right)$$

where the proposal is uniform on $\{K_i - 1, K_i + 1\}$. If $K_i = 1$ and the proposal results in $K_i = 0$ we keep the old value.

The fourth block of the sampler generates θ from

$$p\left(\theta \middle| \bullet\right) \propto \prod_{i=1}^{N} \frac{\exp(-\theta)\theta^{K_i - 1}}{(K_i - 1)!} N\left(\log(\theta) \middle| \theta^0, \sigma^2_{\theta^0}\right)$$

again using the Metropolis-Hastings algorithm with a normal random walk proposal density for $\log(\theta)$.

We update hierarchical priors for $\boldsymbol{\beta}_i$ and $\boldsymbol{\delta}_i, \boldsymbol{\lambda}_i$ in the fifth and the sixth block using standard conjugate results.

<u>Without</u> observed response times as data

Without observed response times as data we sample from

$$p\left(\boldsymbol{\beta}_\mathbf{i}, K_i \middle| \mathbf{y_i}\right) \propto p\left(\mathbf{y_i} \middle| \boldsymbol{\beta}_\mathbf{i}, K_i\right) p\left(\boldsymbol{\beta}_\mathbf{i} \middle| \bar{\boldsymbol{\beta}}, \mathbf{Q}\right) p\left(K_i - 1 \middle| \theta\right)$$

in one block using a Metropolis Hastings algorithm with the same random walk proposal densities as above but a smaller step-size for $\boldsymbol{\beta}_i$. Equation (4) illustrates the form of $p\left(\mathbf{y_i}\middle|\boldsymbol{\beta}_\mathbf{i}, K_i\right)$. The steps involved in updating the hierarchical priors are the same as above.

An alternative sampler augments the unobserved finishing times of all alternatives in a choice set where the full conditional distributions are truncated Gamma. Because updating $\{\boldsymbol{\beta}_i, K_i\}$ conditional on the unobserved finishing times still involves densities that cannot be sampled from directly we opted for the former sampler.

Appendix B – Likelihood Calculations

Let $\vartheta = \left(\{\boldsymbol{\beta}_\mathbf{i}, K_i, \boldsymbol{\delta}_i, \boldsymbol{\lambda}_i\}, \overline{\boldsymbol{\beta}}, \mathbf{Q}, \theta, \boldsymbol{\alpha}, \boldsymbol{\Psi}\right)$ and $\vartheta_i = \left(\boldsymbol{\beta}_\mathbf{i}, K_i, \boldsymbol{\delta}_i, \boldsymbol{\lambda}_i\right)$. We compute

$$p\left(\{\mathbf{y_i}, \mathbf{t_i}\}\middle|\vartheta\right) = \int_{\mathbb{R}^{N\times M,+}} p\left(\{\mathbf{y_i}, \mathbf{t_i}, \mathbf{t^*_i}\}\middle|\vartheta\right) d\left\{\mathbf{t^*_i}\right\} = \prod_{i=1}^{N}\prod_{m=1}^{M} \int_{t^*_{i,m}=t_{i,m}-.5}^{t^*_{i,m}=t_{i,m}+.5} p\left(y_{i,m}, t_{i,m}, t^*_{i,m}\middle|\vartheta_i\right) d\left\{t^*_{i,m}\right\}$$ by

numerical integration. We obtain estimates of $p(\{\mathbf{y}_i, \mathbf{t}_i\} \mid \text{Model})$ defined as $\int p\left(\{\mathbf{y_i}, \mathbf{t_i}\}\middle|\vartheta\right)\pi\left(\vartheta\middle|M\right)\nu\left(d\vartheta\right)$ using the Harmonic Mean estimator.

References

Aaker, David A., Richard P. Bagozzi, James M. Carman, and James M. MacLachlan (1980), On Using Response Latency to Measure Preference, *Journal of Marketing Research*, 17, 2, 237-244.

Anderson, John R., Daniel Bothell, Michael D. Byrne, Scott Douglass, Christian Lebiere, and Yulin Qin (2004), An Integrated Theory of Mind, *Psychological Review*, 111, 4, 1036-1060.

Fazio, Russel H. and Michael A. Olson (2003), Implicit Measures in Social Cognition Research: Their Meaning and Use, *Annual Review of Psychology*, 54, 297-327.

Haaijer, Rinus, Wagner Kamakura, and Michel Wedel (2000), Response Latencies in the Analysis of Conjoint Experiments, *Journal of Marketing Research*, 37 (August), 376-382.

MacLachlan, James and John G. Myers (1983), Using Response Latency to Identify Commercials That Motivate, *Journal of Advertising Research*, 23, 5, 51-57.

MacLachlan, James, John Czepiel, and Priscilla LaBarbera (1979), Implementation of Response Latency Measures, *Journal of Marketing Research*, 16, 4, 573-577.

Newton, Michael A. and Adrian E. Raftery (1994), Approximate Bayesian Inference with the Weighted Likelihood Bootstrap, *Journal of the Royal Statistical Society*, Series B, 56, 3 – 48.

Park, Jong-Won and Manoj Hastak (1994), Memory-Based Product Judgements: Effects of Involvement at Encoding and Retrieval, *Journal of Consumer Research*, 21, 3, 534-547.

Smith, Philip L. and Trisha van Zandt (2000), Time-dependent Poisson counter models of response latency in simple judgment, *British Journal of Mathematical and Statistical Psychology*, 53, 293-315.

Townsend, James T. and F. Gregory Ashby (1983), *Stochastic Modeling of Elemenatary Psychological Processes,* Cambridge University Press, Cambridge.

Tyebjee, Tyzoon (1979), Response Time, Conflict, and Involvement in Brand Choice, *Journal of Consumer Research*, 6, 3, 295-304.