ABSTRACT

Title of Dissertation: EVALUATING COGNITIVE SEQUENTIAL RISK-TAKING MODELS: MANIPULATIONS OF THE STOCHASTIC PROCESS

Timothy J. Pleskac, Ph.D., 2004

Dissertation Directed By: Professor Thomas S. Wallsten, Department of Psychology

This dissertation evaluates, refines, and extends to a new paradigm, a set of stochastic models that describe the cognitive processes of individuals while they complete multiple trials of the Balloon Analogue Risk Task (BART; Lejuez et al., 2002). Wallsten, Pleskac, and Lejuez (2004) designed the models using prospect theory and a Bayesian learning process to better understand why the BART correlates so well with self-reported risky behaviors. The models differed in terms of the individuals' beliefs of the task's probabilistic structure and when option evaluations occur. The models revealed that although respondents use a Bayesian learning process to understand the task, they misunderstand the BART's stochastic process as stationary. Results also indicated that individuals' attitudes toward outcomes are, in part, a source of the BART's success. From these conclusions a new task was developed that allows manipulations of both the actual stochastic structure and the individuals' level of knowledge regarding the structure. Participants (N = 71) completed four different conditions of the task. Fitting the various cognitive models to each individual's data revealed that only a subset of the models correctly distinguished between the stochastic processes underlying the different conditions. Incorporating prospect theory's weighting function and a trial-dependent bias component into the models accounted for performance differences between conditions. Of the assorted model parameters, only prospect theory's value function correlated with external self-reported risky drug use. The results also showed that the learning component of the original BART may cloud its association to risky behaviors. Implications in terms of gambling tasks and the cognitive models will be discussed.

EVALUATING COGNITIVE SEQUENTIAL-RISK-TAKING MODELS:
MANIPULATIONS OF THE STOCHASTIC PROCESS

By

Timothy J. Pleskac

Dissertation submitted to the Faculty of the Graduate School of the
University of Maryland, College Park, in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
2004

Advisory Committee:
Professor Thomas S. Wallsten, Chair
Assistant Professor Michael R. P. Dougherty
Assistant Professor Dave E. Huber
Assistant Professor Carl W. Lejuez
Professor Joe Oppenheimer

Dedication

This work is dedicated to my father Joe Pleskac and my grandfather Art Johnston. Together they taught me how to use my hands, my head and most importantly my heart in everything I do.

Acknowledgements

I would like to thank my advisor Thomas S. Wallsten. My thinking in psychology and in life is clearer because of his support and guidance. I would also like to thank my committee members, Carl Lejuez, Michael Dougherty, David Huber, and Joe Oppenheimer, for their feedback and direction during this project and in my training. Finally, I would like to thank my wife Kate for her loving support and patience throughout my education. The grant R21-DA14699 partially funded the study.

Table of Contents

List of Figures ... vi

List of Tables

List of Figures

Introduction

Different methodologies can be used to identify individuals predisposed to risky behavior. A common method is to ask them to complete scales, such as the Sensation Seeking Scale (Zuckerman, 1990) or the Domain Specific Risk Attitude Scales (Weber, Blais, & Betz, 2002). Laboratory-based gambling paradigms are another option. Rather than reading and answering questions, respondents play a game for multiple trials involving real money. These paradigms have been successful in both identifying individuals predisposed to risky behavior and investigating neuropsychological components involved in decision making (see Bechara, Damasio, Damasio, & Anderson, 1994; Hoffrage, Weber, Hertwig, & Chase, 2003; Lejuez et al., 2002; Slovic, 1966).

The Balloon Analogue Risk Task (BART; see Lejuez et al., 2002) is one such paradigm. Performance in the BART correlates with numerous self-reported risky behaviors, including drinking alcohol, smoking cigarettes, using illegal drugs, gambling, not wearing a seat belt, engaging in unprotected sex, and stealing (see Lejuez, Aklin, Jones et al., 2003; Lejuez, Aklin, Zvolensky, & Pedulla, 2003; Lejuez et al., 2002). Although laboratory-based gambling tasks provide an appealing alternative to scales and questionnaires, little is known as to why they are successful at identifying risky individuals. To this end, Wallsten, Pleskac, and Lejuez (2004) formally modeled the behavior of decision makers (DMs) in the BART to isolate the cognitive processes they were using.[1] The work showed that multiple processes are used during the task to learn, evaluate options and choose. Moreover, the modeling process isolated the BART's association with risk propensity as residing in both the evaluation and response processes;

[1] Busemeyer and Stout (2002) have done a similar procedure to analyze the cognitive processes involved in the Bechara gambling task.

thereby suggesting that excessive risk taking, in part, may be due to both DMs' attitudes towards outcomes, and their insensitivity to an evaluation.

The model development also unearthed several issues regarding the BART and the models themselves. A set of these issues revolve around the stochastic process that controls the task, and individuals' level of knowledge about the process. Interestingly, the models revealed that a Bayesian process is used to learn about the task, but that DMs incorrectly assume the task is controlled by a stationary process. In this paper, I will use systematic changes to a new task's probabilistic structure to (1) investigate how performance is affected at the empirical level, (2) evaluate our (Wallsten et al., 2004) four most successful cognitive models, (3) examine possible modifications to the models, and (4) provide external and empirical validation for the model(s) that best describes the data. Next, I will describe the BART and introduce the four most successful cognitive models.

The BART and cognitive models of performance

During the BART, participants successively face a series of h simulated balloons on a computer. For each balloon, they sequentially click a button on the screen to inflate it, placing x¢ in a temporary bank for each click or pump. But, with each pump, the balloon has a chance of exploding. In fact, unbeknownst to them, the allowable number of pumps for each balloon is set at $n = 128$, with each pump a priori equally likely to produce an explosion. Consequently, after each successful pump the probability of an explosion increases for the next, with the n^{th} pump resulting in a certain explosion. Two events end the trial: an explosion or when DMs choose to stop. If an explosion occurs, then the money in the temporary bank is lost. But, if they stop pumping, then money moves from a temporary to a permanent bank. Typically, the BART is played for 30 trials (balloons) and the measure of performance is the average number of pumps per balloon excluding ones that exploded (adjusted BART score).

Our (Wallsten et al., 2004) models operate on a more fine-grained level of performance than the adjusted BART score, focusing on each of the DMs' choices to pump. Each model predicts the probability of pumping balloon *h* at pump *i*. The four most successful models presume DMs evaluate the gains and/or losses for each pump and then probabilistically choose to pump or stop based on a response rule incorporating their evaluation. Finally, they learn from experience, updating their opinion about the likelihood of the balloon exploding in subsequent trials. The models differ in the DMs' representations of the probabilistic structure of the balloon and when option evaluations occur. Table 1 provides a summary of the four models. Next, I will describe the models in terms of the two differences among them, beginning with the possible beliefs of the stochastic process.

Table 1. A display of the four most successful models and their relation to each other.

	DM's representation of the stochastic process		
Time of evaluation process		***N***on-stationary process, ***i***ncreasing probability	***S***tationary process
	Prior to beginning each balloon	PENi	PES
	Sequentially with each pump	SENi	SES

The DM's representation of the stochastic processes

The BART's instructions are vague as to what determines the balloon's explosion. This leaves an individual DM left to draw his/her own conclusions regarding the stochastic process governing the balloon. Two plausible beliefs are of a stationary process resulting in a constant explosion probability across pumps or a non-stationary process where explosion probability increases with each pump. I will describe the latter first, which also introduces how the balloon is actually programmed to explode.

Non-stationary stochastic process with increasing probability. There are many ways the DM could characterize a non-stationary process. One possibility is to assume the correct representation, but be unsure of the parameters governing the process. In the task, the computerized balloon allows a maximum of n pumps and a priori is scheduled to explode on a random pump between 1 and n, with the a priori probability of any given pump being $1/n$. Thus, the probability of it exploding on the first pump is $1/n$, on the second pump given that it didn't explode on the first $1/(n-1)$, etc. In general, the probability of an explosion on pump i given i - 1 successful pumps is expressed as, $p_i = 1/(n - i + 1)$, where p_i is the probability that the balloon will burst on pump i. The BART is usually programmed so that $n = 128$. Without that information, the DM might understand the general structure, but be unsure of n's value. Hence, we modeled his/her prior opinion of n for balloon 1 with a discretized gamma distribution over n (see Figure 1), fully described by its mean, μ_G, and variance, σ_G^2.[2,3]

---

[2] The gamma distribution is a continuous distribution that is sometimes specified by the parameters ν and τ, Where $\mu_G = \nu\tau$ and $\sigma_G^2 = \nu\tau^2$. The continuous gamma distribution function is $f(x) = \frac{1}{\Gamma(\nu)\tau^\nu} x^{\nu-1} e^{-x/\tau}$.

[3] To obtain the discrete approximation to the gamma distribution, we integrate the distribution from $x = n - 0.5$ to $x = n + 0.5$ for each $n = 1, 2, \ldots, \infty$ and then normalize to account for the lost area from 0 to 0.5.

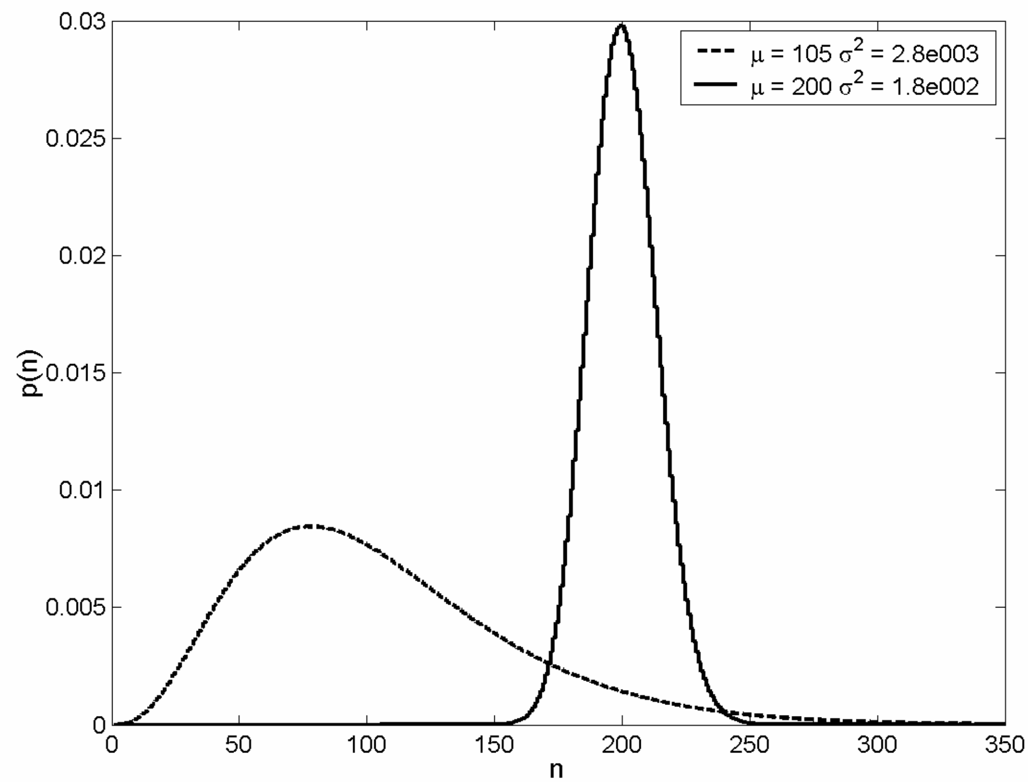

Figure 1. Two discretized gamma distributions with different means and variances.

The mean of the gamma distribution represents the DM's best estimate, prior to the first balloon, of the maximum number of pumps allowed, $\hat{n}_1$. The variance indexes his/her confidence in that guess. Both the mean and variance of the gamma distribution are free parameters estimated from the choice data. Using the estimated mean of the gamma distribution, the subjective probability of the first balloon exploding on pump *i* is $\hat{p}_{1i}$ =1/($\hat{n}_1$ - i +1). Notice that we have added an additional subscript to index the DM's changing opinion over balloons. In general, we will use *h* to index balloon or trial number. Thus, for example, we will write $\hat{p}_{hi}$ to index his/her estimate of balloon *h* exploding on pump *i*.

We assumed after each balloon the DM learns from his/her experience updating his/her prior distribution over *n, p(n)*. The updated distributions do not retain the properties of a gamma distribution; consequently, the revision process is fairly involved. However, the following equation captures the process,

$$p(n \mid c_1, d_1, \cdots, c_h, d_h) = \frac{\prod_{h'=1}^{h} \left[s_{h'} (n - c_{h'})^{d_{h'}} \; p(n) \big/ n^{h'} \right]}{\sum_{n'=1}^{\infty} \prod_{h'=1}^{h} \left[s_{h'} (n' - c_{h'})^{d_{h'}} \; p(n') \big/ n'^{h'} \right]} \tag{1}$$

where $c_{h'}$ is the number of pumps taken on balloon h', $d_{h'} = \begin{cases} 0 \text{ if balloon } h' \text{ popped} \\ 1 \text{ if balloon } h' \text{ did not pop} \end{cases}$

and $s_{h'} = \begin{cases} 0 & if \quad n < c^* \\ 1 & if \quad n \geq c^* \end{cases}$ for $c^* = Max(c_1, \cdots, c_h)$. A proof of this result can be found in Wallsten et al. (2004) and Appendix A. The expected value for the updated distribution for balloon $h+1$, following balloon h, is used to represent the new best estimate of the maximum number of pumps allowed, $\hat{n}_{h+1}$, and is subsequently used for the subjective probability of balloon h+1 bursting, $\hat{p}_{h+1,i}$. This process exemplifies the role of learning in the task and allows n**.** to vary for each balloon.

Stationary process. Alternatively, the DM could mistakenly characterize the balloon as governed by a stationary Bernoulli process with the probability of the balloon exploding, p_h, and not exploding, $q_h = 1 - p_h$, remaining constant over pumps. We modeled the initial uncertainty in q_1 with a beta distribution described by parameters a_0 and m_0, subject to the constraint that $m_0 > a_0 > 0$ (see Figure 2). [4]

---

[4] The beta distribution function is

$$f(x) = \frac{\Gamma(a_0 + b_0)}{\Gamma(a_0)\Gamma(b_0)} x^{(a_0 - 1)} (1 - x)^{(b_0 - 1)}, b_0 = m_0 - a_0 \qquad and \qquad 0 < x < 1.$$

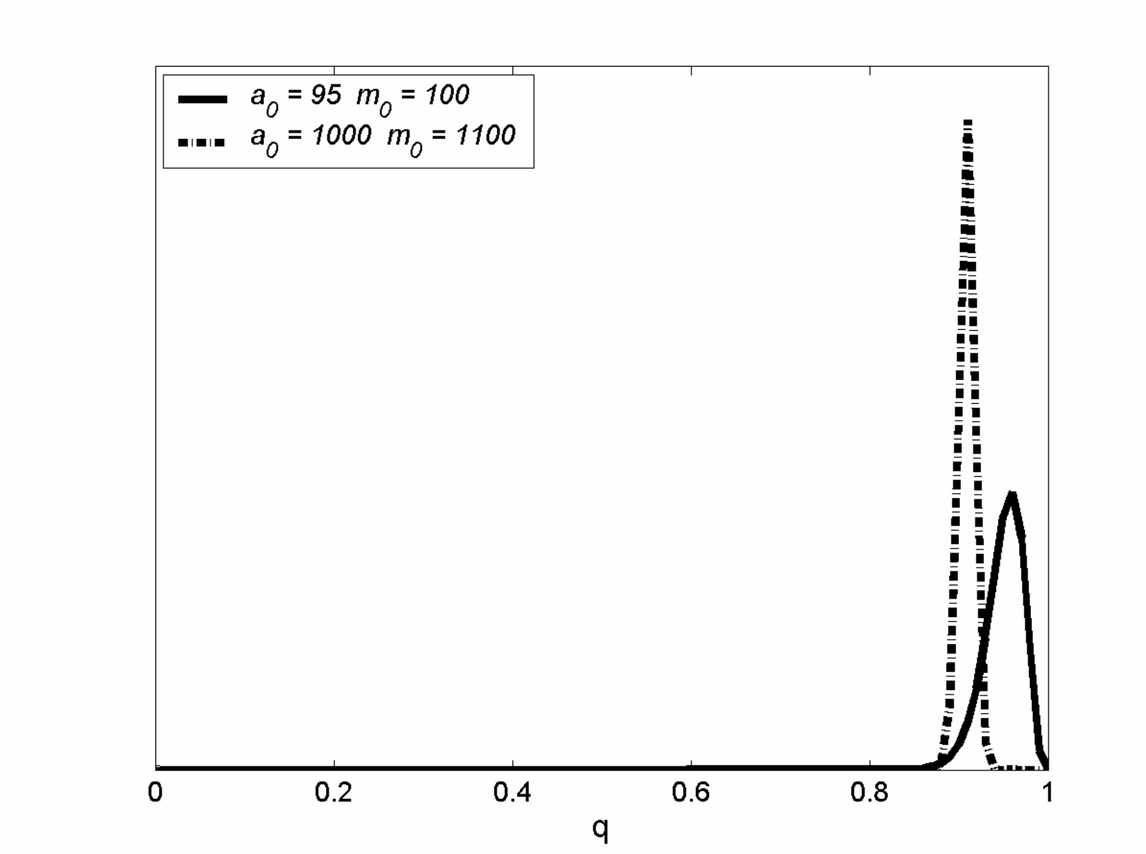

Figure 2. Two beta distributions differing in their estimate and confidence of q_1.

Its mean is used to represent the DM's estimated probability of no explosion for the first balloon exploding, $\hat{q}_1$. Formally, this is expressed as $\hat{q}_1 = a_0 / m_0$. Both a_0 and m_0 are free parameters that are estimated from the choice data and have a psychological interpretation. The more certain the DM is about the value of q prior to observing any data, the greater is m_0 and the greater the DM thinks q is prior to observing data, the greater is a_0 relative to m_0.

The beta distribution is a conjugate distribution of the binomial. If the DM assumes this representation, the balloon's explosion is a binomial event. Modeling the updating process of this representation is more straightforward. After observing the data the posterior distribution over q_{h+1} retains the properties of the prior distribution over q_h, but its parameters change. For example, after the first balloon, regardless if the DM stopped or the balloon exploded, m_0 is incremented by the number of pumps, c_1, made on the first balloon, $m_1 = m_0 + c_1$. If the balloon did not explode, a_0 is also incremented by c_1, $a_1 = a_0 + c_1$. However, if the balloon exploded then it is incremented only by the pumps that resulted in no explosion, $a_1 = a_0 + c_1$-1. The DM's estimate of $\hat{q}_2$ is a_1/m_1. In general, the expression for the DM's estimate of $\hat{q}_{h+1}$ following experience with h balloons can be written as

$$\hat{q}_{h+1} = \frac{a_0 + \sum_{h'=1}^{h-1}\left(c_{h'} - d_{h'}\right)}{m_0 + \sum_{h'=1}^{h-1} c_{h'}} \tag{2}$$

where $d_{h'} = \begin{Bmatrix} 1 \text{ if balloon } h' \text{ popped} \\ 0 \text{ if it did not} \end{Bmatrix}$. We turn next to the two possible evaluation processes, and then combine them with the two possible representation of the balloon's stochastic process just developed.

The DM's evaluation process

To model the DM's evaluations of pump options, we (Wallsten et al, 2004) incorporated prospect theory's value function (Kahneman & Tversky, 1979; Tversky & Kahneman, 1992). This theory presumes that the DM perceives values as changes from a reference point rather than in terms of absolute wealth. It also includes the idea that "losses loom larger than gains." More specifically, the DM considers the absolute value of losing \$10 to be greater than the absolute value of gaining \$10. The value function is usually expressed as a two-part power function:

$$v(x) = \begin{cases} x^{\gamma^+} & x \geq 0 \\ -\theta|-x|^{\gamma^-} & x < 0 \end{cases} \quad (3)$$

Where x is the amount gained on each pump and γ^+, γ^-, $\theta > 0$. In past work, γ^+ and γ^- were found to be less than one, suggesting diminishing sensitivity of gains and losses. Additionally, θ is usually greater than one indicating loss aversion. Figure 3 displays a value function with these characteristics.

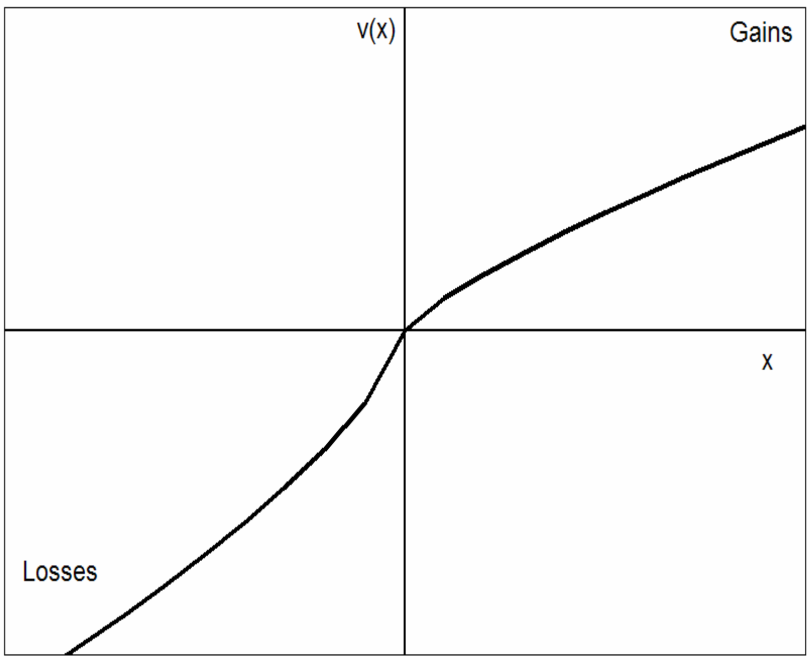

Figure 3. Prospect theory's value function for a range of gains and losses.

When applying these ideas to the BART, there are two different points in time that the DM might evaluate his/her per pump option of either pumping or stopping: (1) sequentially, prior to each pump of the balloon, or (2) prior to beginning to pump each balloon. With prospect theory the two approaches can lead to different choices.

Sequential evaluation. If the DM evaluates options sequentially, then the reference point is assumed to be the current pump, i, for balloon, h. Thus, with each pump, either *(i-1)x¢* can be lost or *x¢* gained. In addition, the chance of either a gain or loss is conditional on the fact of *i-1* successful pumps. Accordingly, a sequential evaluation of the option to pump leads to the following expression,

$$b_{hi} = E_{hi}(\text{pump}) = \hat{q}_{hi} x^{\gamma^+} - \hat{p}_{hi}\theta((i-1)x)^{\gamma^-} \tag{4}$$

where γ^+, γ^- and θ, are free parameters that must be estimated from the choice data.

Although the DM's evaluation is deterministic, his/her choice of whether or not to pump seems more plausibly described as probabilistic. If so, the probability of choosing to pump, r_{hi}, should strictly decrease with b_{hi} and assuming no response bias equal .5 when $b_{hi} = 0$. The response rule expressed as, $r_{hi} = \exp(\beta b_{hi})/[1+\exp(\beta b_{hi})]$, captures these properties. The free parameter β indexes the sensitivity to b_{hi}.

Table 2. A full specification of all four models. The number of free parameters vary for each model depending on whether or not respectfully the Bayesian subcomponent is needed.

Model	numb er pars	Evaluation of each pump	Maximizing pump	Response rule
SENi	6/4	$b_{hi} = E_{hi}(\text{pump}) = \left(\frac{\hat{n}_h - 1}{\hat{n}_h - i + 1}\right)x^{\gamma^+} - \left(\frac{1}{\hat{n}_h - i + 1}\right)\theta((i-1)x)^{\gamma^-}$		$r_{hi} = \frac{e^{\beta b_{hi}}}{1+e^{\beta b_{hi}}}$
PENi	4/2	$E(pump)_{hi} = \frac{\hat{n}_h - i}{\hat{n}_h}[ix]^{\gamma^+}$	$G_h = \frac{\gamma^+ \hat{n}_h}{(\gamma^+ + 1)}$	$r_{hi} = \frac{1}{1+e^{\beta d_{hi}}}$
SES	6/4	$b_{hi} = E_{hi}(\text{pump}) = \hat{q}_h x^{\gamma^+} - \hat{p}_h\theta((i-1)x)^{\gamma^-}$		$r_{hi} = \frac{e^{\beta b_{hi}}}{1+e^{\beta b_{hi}}}$
PES	4/2	$E(pump)_{hi} = \hat{q}_h^i (ix)^{\gamma^+}$	$G_h = \frac{-\gamma^+}{\ln(\hat{q}_h)}$	$r_{hi} = \frac{1}{1+e^{\beta d_{hi}}}$

Combined, the sequential evaluation subcomponent in equation 4 and the DM's two representations of the balloon's structure yield two sequential evaluation models. Each is shown in the third column of rows 1 and 3 of Table 2. Integrating the output of the evaluation subcomponent with the response rule produces two complete models:

SENi (*S*equential *E*valuation *N*on-stationary process *I*ncreasing probability) and SES (*S*equential *E*valuation *S*tationary process *I*ncreasing probability) found in Table 1.

Prior evaluation. In contrast to sequentially evaluating his/her options, the DM may plan how many pumps to carry out prior to each balloon, selecting the number of pumps that maximizes his/her expected gain. In this case, the reference point is located prior to the first pump for each of the h balloons. Due to the BART's payoff structure, evaluations are now only in terms of gains. In addition, each outcome is weighted by the joint probability of successfully pumping the h^{th} balloon i times. The expected gain for each pump is expressed with the following equation

$$E(pump)_{hi} = t_{hi}\left(ix\right)^{\gamma^+} \tag{5}$$

where t_{hi} is the probability of pumping balloon h, i times in succession without exploding. Taking the derivative of Equation 5, setting it equal to 0 and solving for i, produces the pump number that maximizes one's gains for balloon h, G_h. Each solution is specific to the presumed stochastic structure. The maximizing pump column in Table 2 lists the solution for the two models. The parameter γ^+ for both models must be estimated from the choice data.

Having selected G_h, the DM is assumed to probabilistically pump balloon h on pump i. In addition, we also assumed that the probability of taking the i'th pump on the h'th balloon, r_{hi}, strictly decreases with each pump and, is equal to .5 when $i = G_h$ without a bias. Formally, the response rule, $r_{hi} = 1/[1+\exp(\beta d_{hi})]$, captures these properties where $d_{hi} = i - G_h$, and β is a response parameter representing sensitivity to the evaluation. The parameter β also must be estimated from the data. Incorporating the prior evaluation sub-components with the above response rule produces two more fully specified models, PENi (*P*rior *E*valuation *N*on-stationary process *I*ncreasing probability) and PES (*P*rior *E*valuation *S*tationary process).

Past work and predictions

Wallsten et al. (2004) used maximum likelihood estimation procedures to fit a dataset of 58 participants and compared the four models presented here. The participants completed both the BART and a battery of self-reported risky-behavior questionnaires (see Lejuez, Aklin, Jones et al., 2003). In addition, they fit two other models to the data, a baseline model that was estimated directly from the data (see Appendix C), and a simple target model with non-Bayesian learning. The latter model presumed that (1) DMs selected a target pump and probabilistically pumped to their target, and (2) after each balloon individuals learned from their experience by adjusting the target up or down based on the previous outcome.

All four evaluation-based models fit the data substantially better than either of the alternative models; thereby, suggesting that (a) the DMs learn with experience in the task, (b) that this learning process is well approximated by a Bayesian process, and (c) that they evaluate possible outcomes rather than merely setting a target number of pumps. In addition, SES and PES, the models presuming a constant balloon explosion probability, had a better fit than PENi and SENi. To better distinguish between PES and SES, their MLL estimated parameters were correlated with participants' self-reported risky behaviors. Only PES's valuation parameter (γ^{+}) and response sensitivity (β) were significantly associated with the self reports. None of SES's parameters were significantly correlated with the self reports. As a result, Wallsten et al. (2004) selected

PES as the model that best represented the cognitive processes of participants during the BART.

These results and conclusion are relatively surprising. In particular, the suggestion that DMs use an optimal Bayesian learning process, albeit the wrong process for the task's actual structure, generates the need for further investigation of both the task and the models. A natural inquiry is whether the cognitive models could distinguish between DMs' different perceptions of the two stochastic processes under conditions in which they are made fully aware of the different structures. In terms of model comparison and selection, the models assuming a stationary stochastic process (PES and SES) may be in fact mathematically more flexible than the non-stationary models (PENi and SENi). Consequently, they may provide a better fit to the data regardless of whether or not the DM clearly understands the stochastic process. To investigate this issue, it is necessary to manipulate the task structure so that the stochastic process governing the outcomes is either stationary or non-stationary. Changing the structure also allows us to examine the generalizability of the task and its relation to self-reported risky behaviors to alternative stochastic environments.

A second inquiry resides with the lack of a correlation between the learning process parameters (e.g., a_0 and m_0) and risky-behavior. It implies that the BART's procedure of obscuring the correct structure may be adding unnecessary noise. In fact, work with similar tasks (see Hoffrage et al., 2003; Slovic, 1966) suggests the learning component is unnecessary when seeking to predict risky predisposition of individuals. In this case, the cognitive models would be simpler, no longer needing the Bayesian component. It is an open question as to how well the cognitive models fit the data and their MLL parameter estimates correlate with self-reports under these conditions.

To investigate both these issues simultaneously a different task is needed that both holds true to the BART's general scheme, but increases the transparency of the stochastic process. The Angling Risk Task (ART; see Figure 4) does that. Briefly, the ART, as the name implies, is a fishing game analogous to the BART and judgment and decision making's task of "balls in the urn". During the task, participants take a trip to a pond that has 1 blue fish and n -1 red fish. With each cast of a fishing rod, participants hook a fish (each fish is a priori equally likely to be caught). If it is red, then they earn x¢ and can cast again. But, if it is blue, then the trip to the pond ends and the money earned on that trip is lost. The pond's release law can be changed, thereby changing its stochastic structure. Participants can be forced to practice catch 'n' release, creating a stationary process, or catch 'n' keep, a non-stationary process. In addition, the parameters governing the stochastic processes can be masked by having the participant fish on a cloudy day so that they can not see how many fish are swimming in the pond, or can be exposed by having them fish on a sunny day so that the number of fish swimming in the pond are visible.

Having participants complete all 4 pond conditions allows us to examine the questions/hypotheses laid out in the prior sections and listed here. Do the cognitive models distinguish between participants' different perceptions of the two stochastic processes under conditions in which they are made fully aware of the different structures? To what extent does the correlation between task performance and self-reports generalize to alternative stochastic processes? Does the BART's procedure of obscuring the balloon's structure cloud the correlation between performance and self-reports? How does the fit of the cognitive models change when the Bayesian learning components are not needed? After describing the experiment and its results, I also expand the prior evaluation models to include both a trial-dependent bias component and prospect theory's weighting function to better handle the data.

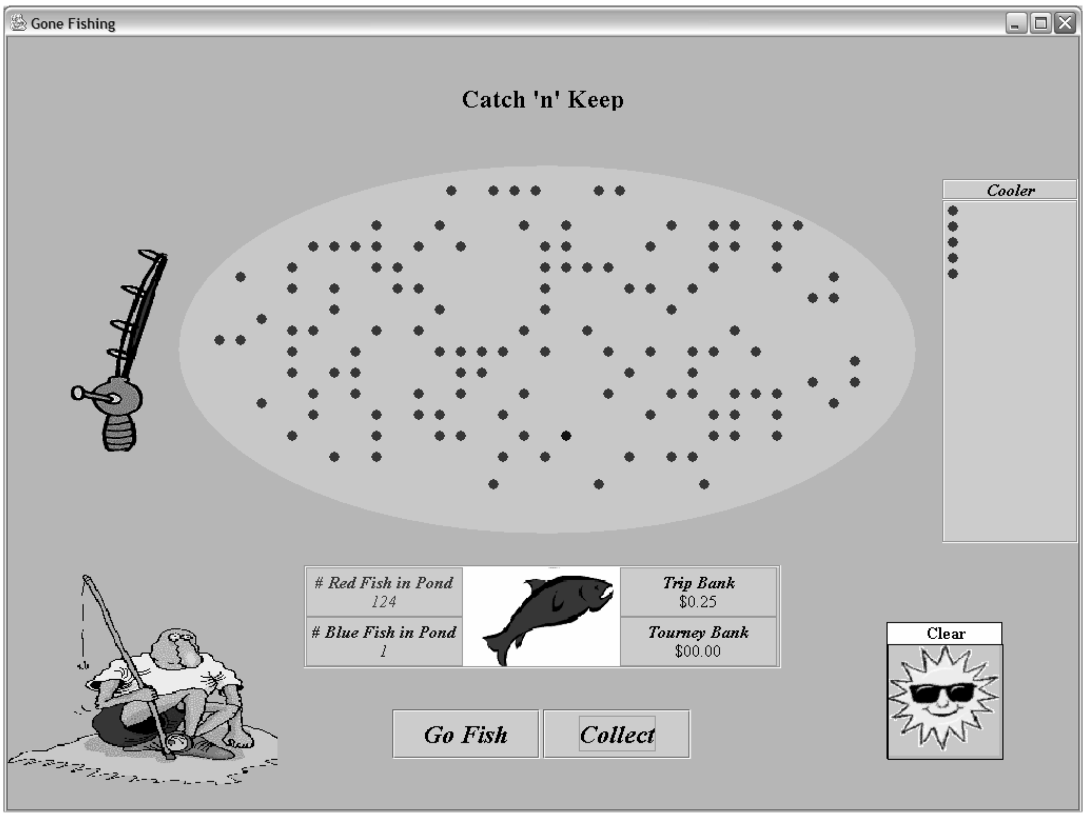

Figure 4. Screenshot of the ART. The weather conditions and conservation law change the task. During *cloudy* days, the fish in the pond are not shown to the participants. During *Catch 'n' Release*, the cooler is closed and the fish are returned to the pond rather than the cooler.

Method

Participants

A total of 72 participants were recruited from the University of Maryland community using advertisements placed throughout the campus. The sample consisted of 38 men and 34 women, ranging in age from 18 to 34 (M = 21.6 and SD = 3.9). Fifty-six percent were White, 18% were Black/African American, 17% were Asian/Southeast Asian, 4% were Hispanic/Latino and the remaining 6% marked other or chose not to respond to the question. They were paid $7 for their time. In addition, participants earned a bonus based on a random set of their pond trials.

Materials

The ART. During a trial (trip to a pond), participants were shown a pond on the screen (see Figure 4). At the beginning of each trip, the pond had 1 blue fish and n-1 red fish. Below the pond were two buttons and an information panel. One button was labeled, "Go Fish." Pressing it caused the rod on the left of the screen to cast a line into the pond and hook a fish. Each fish was equally likely to be caught on a given cast i.[5] If a red fish was caught, then x¢ was placed into the "Trip Bank" shown on the information panel. What happens next depends on the release law. If the law was catch 'n' keep then the red fish was placed in the cooler on the right of the screen, reducing the number of red fish in the pond by one. In contrast, if the law was catch 'n' release then the red fish was placed back into the pond. Either way, participants got another opportunity to cast the line into the pond for that trip. However, if a blue fish was caught, then the trip ended, participants lost their money in the "Trip Bank" and began a new trip. However, if participants decided to stop fishing during a particular trip before catching a blue fish, they pressed the "Collect" button to transfer the money to the "Tournament Bank" on the information panel and began a new trip.

[5] For the remaining of the paper, the terms pump, explosion, and balloon/trial used to describe the BART will be replaced with the ART terms of cast, blue fish, and trip/trial, respectively, for a particular pond. In addition, the term tournament will describe the particular conditions the participant fished under. For example, the participant competed in a *cloudy*, *Catch 'n' Release* tournament.

In addition to the two different release laws, there were two different types of weather. If the weather was sunny, as indicated by the weather forecast in the bottom right, the pond was clear and the participants could see how many fish were in it at all times. In addition, the information panel listed how many red and blue fish were in the pond before each cast. However, if the weather was cloudy, then the pond was murky concealing the number of fish in the pond and the information panel was blank. Combining the two release laws with the two weather forecasts produced four different fishing tournaments/conditions.

Drug and alcohol questionnaire. As a measure of risk propensity, participants completed a drug-use questionnaire, which referred to eleven categories of drugs including, cannabis, alcohol, cocaine, MDMA (ecstasy), stimulants (e.g., speed), sedatives/hypnotics, opiates, hallucinogens, PCP, inhalants, and nicotine. The questionnaire asked three questions each: (1) Have you ever used *drug* (Yes or no)?; (2) About how often did you use *drug* in the past year (Never, One time, Monthly or Less, 2 to 4 times a month, 2 to 3 times a week, or 4 or more times a week)?; (3) During the period in your life when you were using *drug* most frequently, about how often were you using (Never, One time, Monthly or Less, 2 to 4 times a month, 2 to 3 times a week, or 4 or more times a week)?

As a measure of propensity towards risky behavior, I used the following two indices based on participants' responses: (1) The total number of drug categories tried and (2) the weighted sum of drug categories tried, with the weights determined by responses to the third question. These measures or variants of them have been effectively utilized in past studies and the occurrences of these risk behaviors have been shown to correlate with paper and pencil measures of sensation seeking and impulsivity (see Lejuez et al., 2002; 2003a,b).

Domain specific risk-attitude scale (Weber, Blais, & Betz, 2002). This scale, developed and validated by Weber et al. (2002), contains 40-items that assess an individuals' likelihood to engage in risky behavior in 6 domains: ethics, investment, gambling, health/safety, recreational, and social. Two separate variants of the scale also assess an individual's perception of the magnitude of the risk for and expected benefit from each of the 40 risks.

Design and procedure

The study used a 2 (release law) x 2 (weather) within subject design. Participants fished in all four pond tournaments (conditions), and completed the four risk scales/questionnaires. Each tournament gave participants h = 30 trips to the pond to cast for as many red fish as they chose, earning 5¢ per cast. Each pond had n_{wl} fish, where w = s,c for sunny or cloudy, respectively, and l = k,r for keep or release, respectively. Both weather conditions of catch 'n' keep, began with $n_{\bullet k}$ = 128 fish in the pond, while the catch 'n' release conditions had $n_{\bullet r}$ = 65 fish. Thus, in terms of maximizing earnings in an expected value sense, the optimal number of casts in all four conditions was about 64. The order with which participants experienced each tournament and completed the risk questionnaires/scales was counterbalanced. All eight tasks were programmed using Sun Microsystem's Java language and are available upon request. The experiment was administered on PC computers in separate sound attenuated laboratory cubicles.

After reading and signing the informed consent form, participants read an introduction set of instructions on the computer. They were told that they would be playing four different fishing tournaments, each having different rules and conditions. The instructions then described the two different release laws and the two different weather conditions they would experience. In addition, the participants were informed that between each fishing tournament they would fill out a questionnaire assessing their own risky behavior.

Next, the participants completed four practice rounds, one for each tournament condition. This experience served to both reinforce their understanding of the different fishing tournaments and demonstrate that the ponds could have any number of fish. Before each practice round, participants were reminded of the conditions they would experience in the pond. They were then shown a window in which they were allowed to select how many fish they wanted in the pond (1 to 360). Finally, for each practice round they made two trips to the pond during which they cast for red fish as many times as they chose to.

After completing the practice rounds, they began the experimental sessions, starting with a risk questionnaire and then alternating between questionnaire and tournament for the remainder of the experiment. Before each tournament, participants were briefly reminded of the rules governing the pond they were about to visit. At the end of the experimental session, they completed a set of questions regarding the strategy they used to fish in the tournaments. First, they were asked to describe their strategy by typing it into a window with the following instructions:

> *"The strategy you describe should specify how you played the games in such full detail – describing your action in every contingency— so that if you were to write this all down, hand it to someone else, and go on vacation, this other person acting as your representative could play the tournaments just as you would have played it…"*

After explaining their strategy, participants were asked to classify it into one of the six categories shown in Table 3 based only on the description.

Table 3. Five strategies among which participants selected as reflective of what they used during the fishing tournament, and the number of participants who chose each one. When choosing among strategies, they only saw the description.

Strategy name	Description	number of votes
Prior evaluation	Prior to each fishing trip I selected the number of 'Go Fish' presses or casts that I thought would maximize my earnings for that trip. I then pressed the 'Go Fish' button with that number in mind, but sometimes, on a whim, I would stop short. Other times, I might go past that number. But, by and large, I would stop after I reached that number.	39
Sequential Evaluation	Before I pressed the 'Go Fish' button I would assess my situation. I would weigh the benefit of catching one more red fish, against the cost of catching a blue fish. By and large, I would stop once I reached the point at which the costs outweighed the benefits. But, sometimes, on a whim, I would stop short, other times I would go past that point.	7
Satisficing	Before I pressed the 'Go Fish' button, I considered my present state of affairs and decided whether I had reached a satisfactory state. Although I possibly could have made more money, the place where I stopped was good enough for me.	10
Minimize regret	As I pressed the 'Go Fish' button, I chose to stop when I had reached the point at which I felt I would have the least amount of regret if I lost my winnings for that trial.	6
Target strategy	I did not really consider the money when I played the fishing game. Rather, before each visit, I selected how many times I expected I could press 'Go Fish' before catching a blue fish. Then I cast the line that many times.	4
None	None of the strategies listed describe what I used.	5

At the conclusion of the session, the computer produced four tables showing how much money participants earned on each trip (trial) during the four tournaments. A trip from each tournament was then chosen randomly (four trips total) and participants were paid based on the selected trials. Participants were guaranteed $7 plus the money earned in the above selected trials. The whole experimental session took a little over an hour to complete.

Results

The results are organized in the following manner. (1) The data are analyzed with conventional methods aggregating and averaging the adjusted ART score across participants. (2) Results from fitting and comparing the four original models at the individual level are presented. (3) Recognizing some needed extensions to the models, two additional subcomponents, a trial-dependent response bias and weighting function components, are developed. With these extensions the best-fitting models are refit to the empirical dataset. (4) The most useful model's MLL parameter estimates are correlated with the self-reported drug and alcohol use and are used to gain insight to an individual's performance during the tournaments.[6]

Model-free analyses

The model free analyses utilize the adjusted ART score as the dependent variable, which is the average number of casts participants made on fishing trips during a tournament for which they did not catch a blue fish. Using the adjusted ART score, Figure 5 shows that the participants' behavior changed depending on the fishing tournament they were in. Recall that in all four fishing tournaments, the optimal number of casts per trip to maximize expected value was 64. While less than this, the mean adjusted ART scores suggests that participants cast more frequently in the catch 'n' keep than the catch 'n' release condition ($F(1,70) = 17.90$, $p < .001$, $MSE = 105.17$), and more frequently in the sunny weather conditions than cloudy, ($F(1,70) = 16.25$, $p < .001$, MSE

[6] One participant grew agitated during the experiment and did not complete the session. he/she will not be included in subsequent analyses.

= 144.45). Finally, the change from sunny to cloudy had a larger effect in the catch ’n’ keep condition as indicated by a significant interaction between release law and weather ($F(1,70) = 4.80$, $p = .03$, $MSE = 38.79$). The catch ‘n’ keep mean adjusted ART significantly decreased from 38.96 under sunny weather to 31.59, using Tukey’s HSD $q(3,210) = 7.05$, $p < .01$ with the *MSE* from the interaction. While Catch ‘n’ Release decreased from 32.19 to 28.06, using Tukey’s HSD $q(3,210) = 3.95$, $p < .05$.

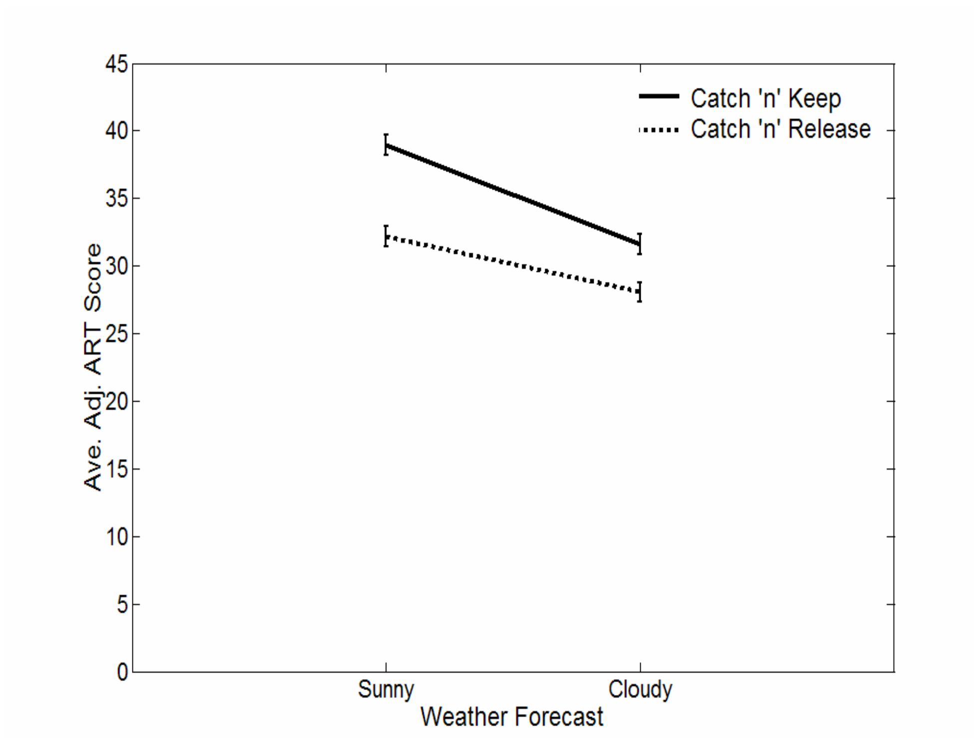

Figure 5. The average adjusted ART score across participants for the four different fishing tournaments. Points represent the average adjusted ART; vertical lines depict standard errors of the mean, estimated from the MSE of the interaction.

Table 4 shows correlations among the adjusted ART scores, demographic variables, Weber et al.’s (2003) six domains of risky behavior, and self-reported drug and alcohol use. Notice that the correlations between the adjusted ART scores and DrugSum, DrugWeighted, and nicotine, are larger in the sunny conditions than the cloudy conditions, providing preliminary evidence that the BART's concealment of the actual stochastic process reduces its correlation to real-world risky behaviors. Further discussion of these results are relegated to the discussion.

The correlation between the Weber et al.'s (2003) social domain and DrugSum on the one hand, and the lack of a correlation between the health domain and DrugSum on the other on the other, is rather surprising because the latter actually contained items on drug and alcohol use. Interestingly, none of the domains correlated with performance in the ART. This result in conjunction with the significant correlation between the ART's sunny conditions and the self-reported drug use, suggests that the two types of predictors are accounting for unique variance in self-reported drug use. This indeed is the case. The average adj. ART score significantly accounts for a unique proportion of the variance in DrugWeighted ($sr^2 = .14$, $t(65) = 3.42$, $p = .001$), as does the average response on the social domain ($sr^2 = .09$, $t(65) = 2.70$, $p = .008$). These analyses are not the whole story. For that we will evaluate and use the cognitive models.

Table 4 Means, Standard Deviations, and Intercorrelations Among Each of the Conventional Independent and Dependent Variables.

Variable	MEAN	STD	N	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Demographic																		
1. Age	21.56	3.58	71	-														
2. Gender	0.54	0.5	71	0.05	-													
Predictor																		
3. ART (k, s)	38.96	18.77	71	0.14	0.16	-												
4. ART (k, c)	31.59	15.35	71	**0.24***	0.23	**0.69#**	-											
5. ART (r, s)	32.19	15.95	71	**0.26***	0.1	**0.75#**	**0.57#**	-										
6. ART (r , c)	28.06	15.22	71	**0.27***	0.08	**0.51#**	**0.72#**	**0.64#**	-									
7. D-Gambling	2	1.18	71	-0.03	-0.02	0.02	0.18	0.13	0.15	-								
8. D-Financial	3.23	1.14	71	-0.1	0.01	-0.11	-0.03	0.01	-0.01	-0.12	-							
9. D-Health	2.73	1.44	71	0.13	0.22	0.05	0.1	0.08	0.08	0.18	0.03	-						
10.D-Ethics	2.13	1.23	71	-0.12	0.05	0.16	-0.03	0.02	-0.2	0.14	-0.15	0.11	-					
11. D-Recreation	2.79	1.31	71	**-0.29***	0.22	0.04	0.12	0	0.03	**0.25***	**0.25***	0.02	-0.02	-				
12. D-Social	3.75	1.02	71	-0.07	-0.15	-0.07	-0.03	0	-0.08	0.15	0.06	0.14	0.04	0.15	-			
Risk Behavior																		
13. DugSum	2.68	2.58	68	0.04	0	**0.38#**	0.15	**0.32#**	0.09	0.05	0.08	0.12	0.13	0.07	**0.27***	-		
14. DrugWeighted	7.44	7.99	68	0.09	-0.02	**0.36#**	0.14	**0.32#**	0.11	0.05	0.06	0.15	0.12	0.08	**0.28***	**0.97#**	-	
15. Alcohol	0.85	0.36	68	0.01	-0.15	0.05	0.07	0.04	0.02	**0.24***	-0.1	0.11	-0.04	0.02	**0.29***	**0.38#**	**0.38#**	-
16. Nicotine	0.53	0.5	68	0.01	-0.03	**0.30***	0.21	**0.32#**	0	**0.27***	**0.24***	0.13	0.16	0.13	0.12	**0.58#**	**0.54#**	**0.27***

Note. Gender was coded 0 for female, and 1 for male; The four ART conditions use the Adj Cast score, which is the average number of casts which did not result in a blue fish; Weber et al.'s (2003) six risk domains are listed, and use the average response within each domain set; DrugSum is the number of drug categories tried; DrugWeighted weights the drug categories tried by the frequency of use under their heaviest use; Alcohol and Nicotine are yes/no answers as to whether a participant has ever used either drug. They are included in both the DrugSum and DrugWeighted measure. Three participants failed to complete the drug/alcohol questionnaire.

* = $p < 0.05$, # = $p < 0.01$

Model analyses

The four cognitive models summarized in Table 1 and Table 2 each predict the probability, r_{hi}, of cast i on trip h. They differ in when the DM evaluates options and in his/her representation of the stochastic process. The models are estimated at the level of the individual using his/her entire dataset, not just the trials when he/she chose to stop. Consequently, the models can offer a different account of performance during the fishing tournaments.

The data. Figure 6 uses the baseline model (see Appendix B) to characterize the data in terms of r_{hi}. Briefly, the baseline model uses the proportion of casts made for each cast opportunity i across all 30 trips as an estimate for r_{hi}, assuming that if DMs chose to stop on cast i they would stop on all subsequent casts. Averaging these estimates across all 71 participants for each tournament produces Figure 6. It shows that both within and between tournaments there is a large amount of variability in the data. In fact, the adjusted ART score appears to capture only a small portion of the variability. The graphs demonstrate that participants did in fact cast beyond the adjusted ART score. In fact, the catch ‘n’ release condition by definition allowed them to cast more than the catch ‘n’ keep. This indeed happened as the graph and data show. In addition, they also caught a higher proportion of blue fish in these conditions.

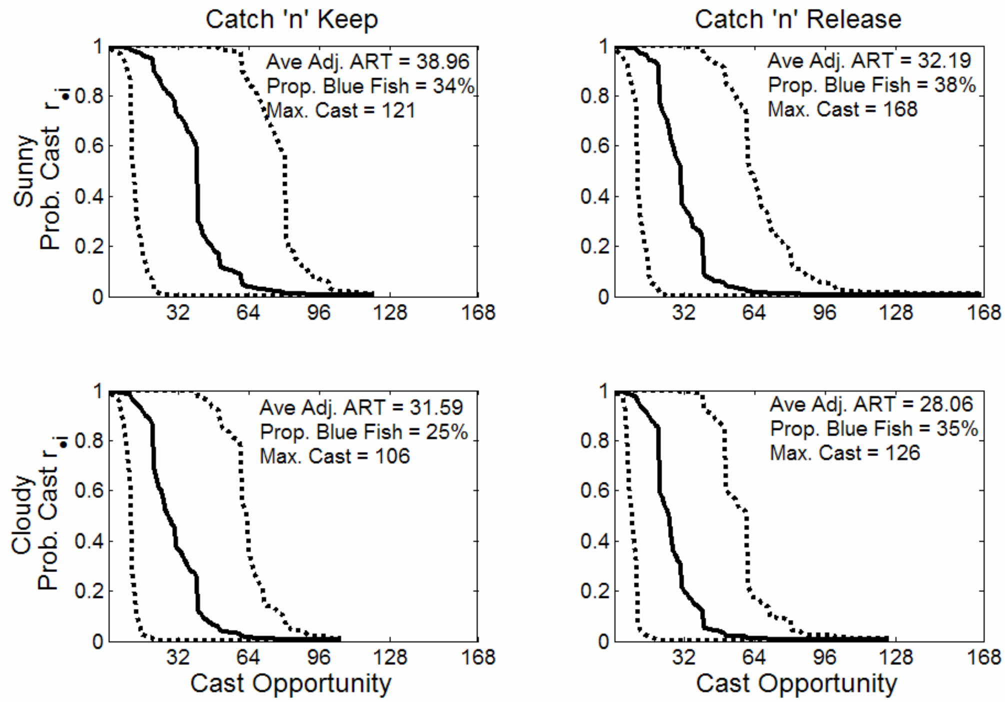

Figure 6. The estimated average probability of casting for each cast opportunity in the four fishing tournaments. The averages (solid lines) are based on the baseline model estimates from all 71 participants (see Appendix B). The solid lines only extend out to the maximum cast taken per tournament among all 71 participants. The dotted lines represent the 95% confidence interval for each cast's estimate. As one can see there is a large amount of variability in the data, which itself varies between fishing tournaments.

Model estimation. Although the baseline model was easily estimated, the remaining cognitive models required more effort. Each model was fit to each individual's data from each tournament using maximum likelihood methods. Let the vector $Y_{w,l} = \left(c_1, d_1, c_2, d_2, \cdots, c_{h'}, d_{h'}, \cdots, c_{30}, d_{30}\right)$ be the observed data from tournament *w,l*, where $w = s$ or c for sunny or cloudy, respectively, $l = k$ or r for keep or release, respectively, $c_{h'}$ is the number of casts for attempt *h*', and $d_{h'}$ is whether the DM stopped (1) or not (0). The log likelihood of the observed data, $Y_{w,l}$, for each of the models is defined as:

$$L_{w,l} = \left(c_1, d_1, c_2, d_2, \cdots, c_{h'}, d_{h'}, \cdots, c_{30}, d_{30}\right) = \sum_{h'=1}^{30} \sum_{i'}^{c_{h'}-1} \left[\ln\left(r_{h'i'}\right)\right] + \left(1-d_{h'}\right)\ln\left(r_{h'c_{h'}}\right) + d_{h'}\ln\left(1-r_{h'c_{h'}}\right) \qquad (7)$$

where each model predicts r_{hi}.

None of the cognitive models have a closed form solution to find the maximum log likelihood (MLL) estimates of the parameters. Consequently, the solutions were estimated with numerical optimization techniques, of which there are many. Both past experience and simulations have shown that a Nelder-Mead downhill simplex routine (available in Mathwork's Matlab) combined with a grid-search technique is the most successful at both reaching a solution and guarding against local maxima (see Appendix C). To estimate the models, I imposed constraints on some of the parameters to facilitate the optimization procedure. In particular, the valuation exponents and the mean of the discretized gamma distribution were constrained such that, $0 \le \gamma^{+}, \gamma^{-} \le 3$ and $0 \le \mu_0 \le 1000$.

Fitting the models to the data from the sunny weather tournaments also proved informative as to the limits of different stochastic representations. Recall that during these conditions the parameters controlling the pond's probabilistic structure were transparent to the participants. In other words, they knew that the catch 'n' release pond had 65 fish and the catch 'n' keep pond had 128 fish, making the models' Bayesian subcomponent unnecessary and reducing the number of model parameters by two. However, this alteration makes fitting PENi and SENi to the catch 'n' release, sunny tournament problematic, as the models don't sensibly allow for a DM to make more than 65 casts. Similarly, although PES and SES do allow for the behavior observed in catch'n'keep, fitting them to the data leads to extreme and unreasonable parameter estimates. Close inspection of the prior evaluation models (PE•) also revealed that, if allowed to, both PE• models would produce the same fit to the data. Consequently, only the prior and sequential evaluation models that assume the correct stochastic structure (non-stationary or stationary) were fit to the data to yield the subsequent results.

Model comparisons. The models have different numbers of parameters (2 or 4 under the sunny weather condition and 4 or 6 under cloudy) and are not nested; thus, standard maximum-likelihood ratio tests are not available to evaluate them. Akaike's information criterion (AIC; Akaike, 1973) is one common method used to compare the fit of non-nested models at a descriptive level. It is a function of both the maximum log likelihood of the data given the model (LL) and the number of parameters in the model. The latter is used as a heuristic measure of complexity (i.e., the more parameters the more complex the model). AIC is calculated as, $AIC = -2LL + 2k$. The model with the smallest criterion measure is selected as the best-fitting model, handicapping models with more parameters. However, simulations that I will present shortly suggest that AIC does not necessarily lead to the correct conclusion. As a result, I present comparisons at the level of the individual based on both AIC and the maximum LL.

Tables 5 and 6 show how many participants were best fit by each model using either LL or AIC. Comparing the LL of each model at the level of the participant, Table 5 shows that under cloudy conditions in catch 'n' keep a majority of participants were fit best by SES, which presumes an incorrect stationary process. However, when the model fits were handicapped by the number of parameters, AIC resulted in a plurality of participants best fit by PENi, which assumes the correct stochastic process. A similar pattern emerges in the sunny conditions of catch 'n' keep. SENi, the sequential evaluation model, fits a majority using LL as a measure of fit, but PENi fits a majority with AIC. Noticeably the baseline model does appear to fit a few participants best in each condition. Closer scrutiny of the data of these individuals showed that these participants were quite consistent in their behavior (i.e., always casting 15 or 16 times). Table 6 exhibits analogous results for the catch ‘n’ release tournaments. Under both weather conditions, SES is fit best by a plurality of participants using LL, while PES is fit best by a plurality using AIC.

Table 5 Model comparison analysis of the cloudy and sunny catch ‘n’ keep tournaments. The *df* for each model are in parentheses next to the respective model. The df for the baseline model ranged between 13 and 106 for the cloudy condition, and 13 and 121 for the sunny condition. In the sunny conditions, only the models that assumed the correct stochastic process could reasonably be fit to the respective conditions.

	Catch 'n' Keep							
	Cloudy				Sunny			
Model	Mean LL	Num. DM's best fit with LL	Mean AIC	Num. DM's best fit with AIC	Mean LL	Num. DM's best fit with LL	Mean AIC	Num. DM's best fit with AIC
Baseline	-209.90	1	525.99	0	-211.05	4	549.87	1
PES (4,2)	-73.84	3	155.68	17				
PENi (4,2)	-73.19	10	154.37	27	-78.06	23	160.12	52
SES (6,4)	-71.56	42	155.12	20				
SENi (6,4)	-72.92	15	157.83	7	-73.94	44	155.88	18

Table 6 Model comparison analysis of the cloudy and sunny catch ‘n’ release tournaments. The *df* for each model are in parentheses next to the respective model. The df for the baseline model ranged between 13 and 106 for the cloudy condition, and 7 and 168 for the sunny condition. In the sunny conditions, only the models that assumed the correct stochastic process could reasonably be fit to the respective conditions.

	Catch 'n' Release							
	Cloudy				Sunny			
Model	Mean LL	Num. DM's best fit with LL	Mean AIC	Num. DM's best fit with AIC	Mean LL	Num. DM's best fit with LL	Mean AIC	Num. DM's best fit with AIC
Baseline	-159.59	3	414.25	0	-184.45	3	483.30	1
PES (4,2)	-60.70	9	129.40	31	-69.26	4	142.51	60
PENi (4,2)	-61.46	9	130.92	18				
SES (6,4)	-58.87	33	129.74	12	-68.32	64	144.65	10
SENi (6,4)	-60.30	17	132.60	7				

These results are a little puzzling. On the one hand, the LL comparisons lead to a conclusion that participants are evaluating their options consistent with the sequential evaluation models (SE•). But, SES incorrectly fits the data best under both cloudy tournaments, indicating that it may be too flexible a model. On the other hand, the AIC comparisons lead to a conclusion that participants are performing consistent with the prior evaluation models (PE•). Under cloudy conditions the prior evaluation models appear to correctly distinguish between the stochastic processes, which is a nice result considering past work showed that the DM incorrectly believed the non-stationary process to be stationary (i.e., PES fit the BART best). Table 7 looks at this result more closely and shows that this indeed is the case when focusing only on the prior evaluation models under cloudy weather. The column labeled *prior* compares the number of DM's best fit with the two prior evaluation models in both release conditions, removing individuals best fit by the baseline model. The column labeled *sequential* does the same for the two sequential evaluation models and shows that the sequential models do not differentiate the processes.

Table 7 The number of DM's best fit within the PE• and SE• models under cloudy conditions, removing the participants for whom the Baseline was the best fit.

	Prior		Sequential	
	PES	PENi	SES	SENi
Catch 'n' Keep	24	46	52	18
Catch 'n' Release	44	24	45	23

Rather than conditionalizing on the time of evaluation, we can conditionalize on the presumed stochastic process. The top half of Table 8 does so, comparing the prior and sequential evaluation models under cloudy conditions, assuming the correct stochastic process. The empirical dataset row substantiates the result that the sequential evaluation models (SE•) are selected with the LL comparisons while the prior evaluation models (PE•) are selected with AIC, holding all else constant.

Table 8 The number of DM's best fit within •Ni and •S under cloudy conditions for both the empirical and simulated datasets. In the empirical dataset, participants best fit by the baseline model were removed from these comparisons.

			Catch 'n' Keep		Catch 'n' Release	
			PENi	SENi	PES	SES
Emprical dataset		Num. DM's best fit with LL	33	37	17	51
		Num. DM's best fit with AIC	59	11	50	18
Simulated dataset	*with* Prior evaluation	Num. DM's best fit with LL	49	51	58	42
		Num. DM's best fit with AIC	69	31	94	6
	with Sequential evaluation	Num. DM's best fit with LL	27	73	17	83
		Num. DM's best fit with AIC	51	49	65	35

Table 9 The number DM's best fit in the simulated dataset within •Ni and •S under sunny conditions. In the empirical dataset, participants best fit by the Baseline model were removed from these comparisons.

		Catch 'n' Keep		Catch 'n' Release	
		PENi	SENi	PES	SES
Prior Evaluation	Num. DM's best fit with LL	36	64	4	96
	Num. DM's best fit with AIC	97	3	96	4
Sequential Evaluation	Num. DM's best fit with LL	12	88	1	99
	Num. DM's best fit with AIC	80	20	94	6

A simulated dataset was produced to further investigate this comparison. To do so, the MLL parameter estimates for 10 random participants were used to generate a dataset in which the simulated participants played all 4 tournaments with the prior evaluation models and the sequential evaluation models, assuming the correct stochastic process. Each simulation was repeated 10 times per simulated participant, resulting in a dataset with 100 tournament plays per evaluation time (PE• or SE•) across participants. The bottom rows in Table 8 labeled simulated dataset include the results from fitting both PE• and SE• models to the respective conditions generated under cloudy conditions.

Unfortunately, the same patterns of results occur. The sequential evaluation models (SE•) are selected with the LL comparisons while the prior evaluation models (PE•) are selected with AIC, even under conditions in which sequential evaluation models actually generated the data. Table 9 confirms the identical pattern in the sunny conditions. These troubling results suggest that neither LL nor AIC necessarily identify the correct model. The former does not sufficiently account for the complexity of the model, while the latter overcompensates. Other measures are available to select among models that attempt to account for both a model's goodness of fit and complexity. An explanation of these measures is left for the discussion.

The prior evaluation models do appear to best describe the cognitive processes used during the fishing tournaments. First, past work and the present study jointly demonstrate that the prior evaluation models can discriminate between circumstances in which the DM incorrectly and correctly represents the stochastic process. Second, the present experiment suggests that the sequential evaluation model SES is too flexible a model, fitting all four tournaments the best using LL as a goodness of fit measure. Finally, a majority of participants identified the prior evaluation strategy as consistent with their own strategy (see Table 3). Tables 10 and 11 summarize the MLE parameter estimates for the prior evaluation models with the correct stochastic process in each tournament. For the remainder of the paper, I will focus on the prior evaluation models, PES and PENi, and their fit to their respective release law conditions. Next I will examine possible extensions to them, incorporating prospect theory's weighting function and a bias component. Although both components could easily be included in the sequential models, evidence already suggests that at least SES is already too complex.

Table 10 MLE parameter estimates summary for PES in Catch 'n' Release tournaments. The last two columns summarize the mean and variance of the initial beta distribution and are calculated from a_0 and m_0.

	Catch 'n' Release: *PES*							
	Sunny		Cloudy					
	β	γ^+	B	γ^+	A_0	m_0	$E(q_1)$	$\mathrm{var}(\hat{q}_1)$
Mean	0.16	1.17	0.93	0.78	5.94E+06	6.93E+06	0.98	2.49E-04
1st Quartile	0.06	0.63	0.14	0.38	102.47	104.85	0.97	5.00E-06
Median	0.11	0.98	0.19	0.65	328.71	334.73	0.98	4.44E-05
3rd Quartile	0.15	1.45	0.34	0.99	1465.59	1494.97	0.99	1.80E-04
IQR	0.09	0.82	0.20	0.62	1363.12	1390.12	0.02	1.75E-04

Table 11 MLE parameter estimates summary for PENi in Catch 'n' Keep tournaments

	Catch 'n' Keep: *PENi*					
	Sunny		Cloudy			
	β	γ^+	B	γ^+	μ_0	$\sigma_0 2$
Mean	0.69	1.74	0.29	1.09	137.06	3.67E+17
1st Quartile	0.08	0.59	0.10	0.53	43.68	1.82E+02
Median	0.11	1.78	0.16	0.90	82.48	2.12E+03
3rd Quartile	0.16	2.97	0.31	1.25	160.27	5.56E+03
IQR	0.08	2.38	0.21	0.73	116.59	5.38E+03

Extending the models

During this section I develop and test two additional subcomponents of the prior evaluation models (PE•): a trip-dependent bias component and prospect theory's weighting function (see Kahneman & Tversky, 1979; Tversky & Kahneman, 1992). The former captures unique strategies of participants and the latter accounts for their perception of probabilities. Each is developed in turn.

The bias component. Our initial work with past data found that an added bias parameter in the response function that remained constant across trials did not significantly improve the fit of the models.[7] However, close investigation of the exit

[7]In the current dataset, fitting the models with a constant bias parameter in the sunny conditions did significantly improve the fit of the models, but not in the cloudy conditions. The trip dependent bias component improved the fit in all four conditions and better captured the strategies explained in this section.

interviews revealed a number of participants reported doing either of two actions during the tournaments. One behavior is best identified as "Testing the waters." One participant described this stating, "I started each round by casting out as many times as I could until I caught a blue fish." In other words, an initial bias existed to go past the cast that maximizes gains. A second and related tactic labeled "Go for broke" by one participant occurred when he/she got bored during the tournament and decided to see how far he/she can get. Another participant described this stating, "After a while I got bored and started pushing to see how high I could get up to around \$4." To account for these trip sensitive actions a bias component can be added to the response function. This hypothesis is formalized by assuming the bias changes over trips and is most prevalent at either the beginning or the end of a tournament according to this expression:

$$\zeta_h = \exp\left[z\left(h-15\right)\right]-1 \tag{7}$$

Negative and positive values of z characterizes the "Test the waters" and "Go for broke" strategies, respectively. If $z = 0$ then the participant exhibits no trip dependent bias. The response function with the bias component is now expressed as

$$r_{hi} = \frac{1}{1+e^{\beta d_{hi}-\zeta_h}} \tag{8}$$

The addition of the bias component can be tested against its absence with the likelihood-ratio test, $G^2 = 2\left[L\left(M\right)-L\left(M'\right)\right]$. Where *L(M)* and *L(M')* are likelihoods of the general and restricted models, respectively. The statistic is asymptotically χ^2 distributed with *df* equal to the difference in the number of the parameters. In this case, the general models are those that contain the bias component. The bias-free models are the restricted models with $z = 0$. Fitting the prior evaluation models (PES and PENi) to the respective tournaments and summing the log-likelihoods across all 71 participants' results in one ratio test per tournament with 71 *df* (1 parameter difference for 71 participants). All four tests indicated the necessity of a trip dependent bias component. The tests for the bias component in PENi under sunny and cloudy conditions resulted in, $G^2 = 915.87$, $p < .01$ and $G^2 = 105.54$, $p < .01$, respectively. A similar result was found with PES, $G^2 = 436.85$, $p < .01$ and $G^2 = 168.34$, $p < .01$. Although particular individuals in specific tournaments did not statistically need the bias component, the extensive amount of individual differences among participants within and between each tournament resulted in its need across participants and tournaments. Based on this need, the bias component will be integrated into the models as the weighting function's use is examined.

The weighting function. Kahneman and Tversky's (1979; 1992) prospect theory hypothesizes that individuals distort probabilities in a nonlinear fashion, overweighting small probabilities and underweighting large probabilities. Like the value function, the weighting function also tends to exhibit diminishing sensitivity as one moves away from the reference points of 0 and 1. In other words, increasing the probability of winning a prize from 0 to .1 or decreasing from 1 to .9 has more impact than a change from .3 to .4 or .7 to .6. These properties give rise to a function that is concave near zero and convex near one, as displayed in the far left panel in Figure 7 (see Gonzalez & Wu, 1999; Kahneman & Tversky, 1979; Luce, 2000; Prelec, 1998; Tversky & Kahneman, 1992; Tversky & Wakker, 1995; Wu & Gonzalez, 1996, 1999). Incorporating the weighting function with the value function leads to the fourfold pattern of risk attitudes documented by Tversky and Kahneman (1992). For events of low probability, the DM is risk seeking for gains (e.g., lottery tickets) and risk averse for losses (e.g., insurance). At the same time, for events of high probability, the DM is risk averse for gains and risk seeking for losses.

Several functional forms of the weighting function are available (see Gonzalez & Wu, 1998; Luce, 2000). Prelec's (1998) function proved the most tractable for these models. It assumes a weighting function for gains and losses, jointly characterized by three positive parameters, $0 \le \alpha \le 1$ and $0 < \delta^{+}, \delta$:

$$
\begin{aligned}
w^{+}(p) &= \exp\left(-\delta^{+}\left(-\log(p)\right)^{\alpha}\right), \quad x \ge 0 \\
w^{+}(p) &= \exp\left(-\delta^{-}\left(-\log(p)\right)^{\alpha}\right), \quad x < 0
\end{aligned}
\tag{9}
$$

If $\alpha, \delta^{+}, \delta^{-} = 1$, the weighting functions are linear, illustrating Wallsten et al.'s (2004) initial assumptions. The parameter α controls the degree of over/underweighting of probabilities. As α increases DM exhibits more discriminability between option likelihoods (see the middle panel of Figure 7). The parameter $\delta^{\bullet}$ controls the inflection point of the weighting function or the elevation of the function (see the right panel of Figure 7).

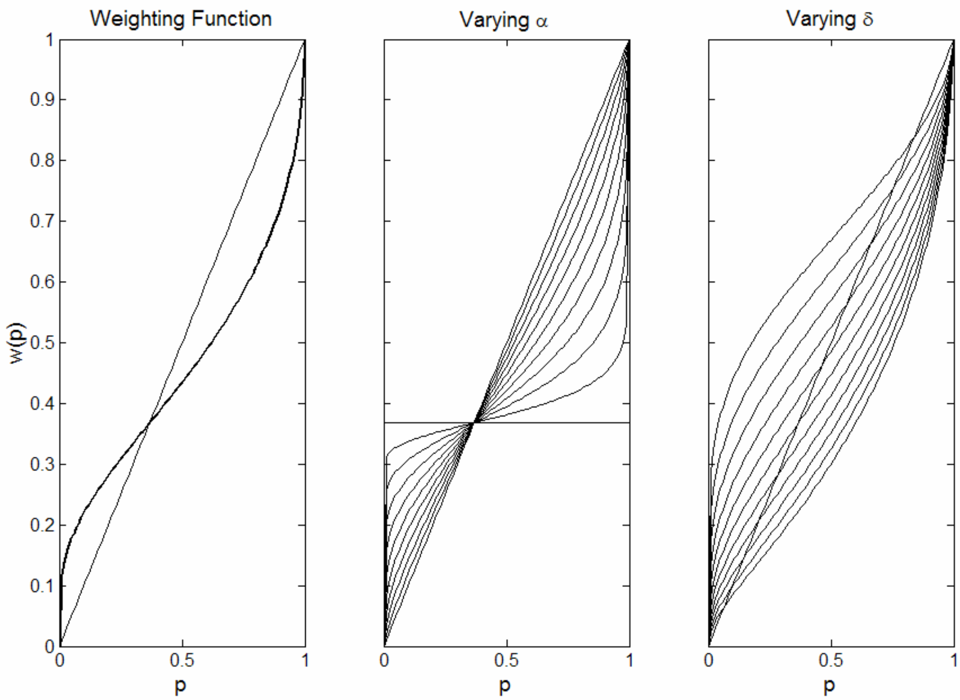

Figure 7 Plots of the weighting function and properties of the weighting function. The first panel shows a weighting function exhibiting overweighting of low probabilities and underweighting high probabilities. The middle panel fixes δ = 1 and varies α between 0 and 1. The third panel fixes α = 0.6 and varies δ between 0.5 and 1.5.

Incorporating the weighting function into the prior evaluation expression (equation 5) produces the following function:

$$E(cast)_{hi} = w^{+}(t_{hi})(ix)^{\gamma^{+}} . \tag{10}$$

Taking the derivative of equation 10, setting it equal to 0, and solving for *i*, produces the cast number that maximizes one's gains for trip *h*, G_h. Each solution is specific to the stochastic process. PESw's (the *w* signifies a weighting function) closed form solution is,

$$G_h = \frac{-\left(\gamma^{+} / \delta^{+}\alpha \right)^{1/\alpha}}{\ln(q_h)} . \tag{11}$$

PENiw does not have a closed form solution, but G_h for *PENiw* can easily be found with numerical methods.[8]

There are several observations from equation 11 that have implications for fitting the models to the data that also hold true for PENIw's numerical estimates. First, the presence of the weighting function may counteract, or at least serve as an alternative to γ^+ going above 1. For example, setting $\gamma^+ = 1$, as the weighting function becomes increasing nonlinear, G_h grows larger than the optimal pump number in an expected value sense (i.e. $G_h > 64$). Second, the valuation parameter, γ^+, and the weighting parameter, δ^+, can not be estimated independently. This is not necessarily detrimental, as δ^+ only controls the inflection point. In fact, past work has primarily focused on allowing α to vary while setting $\delta^+ = 1$ (see Prelec, 1998; Tversky & Kahneman, 1992), which I will do as well.[9] Third, all three parameters in the numerator cannot be estimated independently of each other when $\hat{q}_h$ remains constant. This situation occurs under sunny weather conditions or in cloudy conditions when the DM is extremely confident of himself. This issue also is solved by resolving a final less evident issue: α and γ^+ are not identifiable even when the denominator (i.e. $\hat{q}_h$ or $\hat{n}_h$) is allowed to vary. The problem and its solution are shown in Figure 8.

The top two plots in Figure 8 demonstrate this issue for both PESw and PENiw. They plot the maximizing cast, $G_\bullet$, for one stochastic process parameter (e.g., $\hat{q}_h = 64/65$ for PESw or $\hat{n}_h = 128$ for PENiw) against corresponding values of $G_\bullet$ for a

---

[8] This result is true also for Kahneman and Tversky's (1992) linear in log odds functional form also. In fact, their function also does not have a closed form solution for PESw.

[9] Gonzalez and Wu (1998) found that although the one-parameter functions captured group level data adequately, there were sufficient differences between individuals that the two parameter functions were required at the level of the individual.

different value of the parameter, across values of γ^+ and at a fixed level of α. Note the axes are different labels reflecting the different models and different stochastic process parameters. The lines fall on top of each other indicating γ^+ and α are not identifiable. However, they are identifiable when the parameters are set equal across stochastic processes, as the bottom plot shows by plotting PENiw's values of $G_\bullet$ against PESw's. In this case the two models share common values for γ^+ and α. I will call this constraint Model 2, treating all four tournament plays by one participant as one experimental session and the 2 prior evaluation models fit to the four tournaments with the aforementioned constraints as one model. Incorporating the bias component, ζ, Model 2 has 16 parameters: 4 bias parameters z_{wl}; 4 sensitivity parameters β_{wl}; 2 value parameters $\gamma^+_{w\bullet}$, 2 weighting parameters $\alpha_{w\bullet}$; and 4 parameters controlling the DM's representation of the stochastic process $a_{0\text{cr}}$, $m_{0\text{ cr}}$, $\mu_{0\text{ck}}$, σ^2_{0ck}, where $w = s,c$ for *sunny* or *cloudy* and $l = k,r$ for keep or release, respectively. Model 1, the set of models without a weighting function, also has 16 parameters. It has no weighting function parameters and 4 value parameters, γ^+_{wl}.

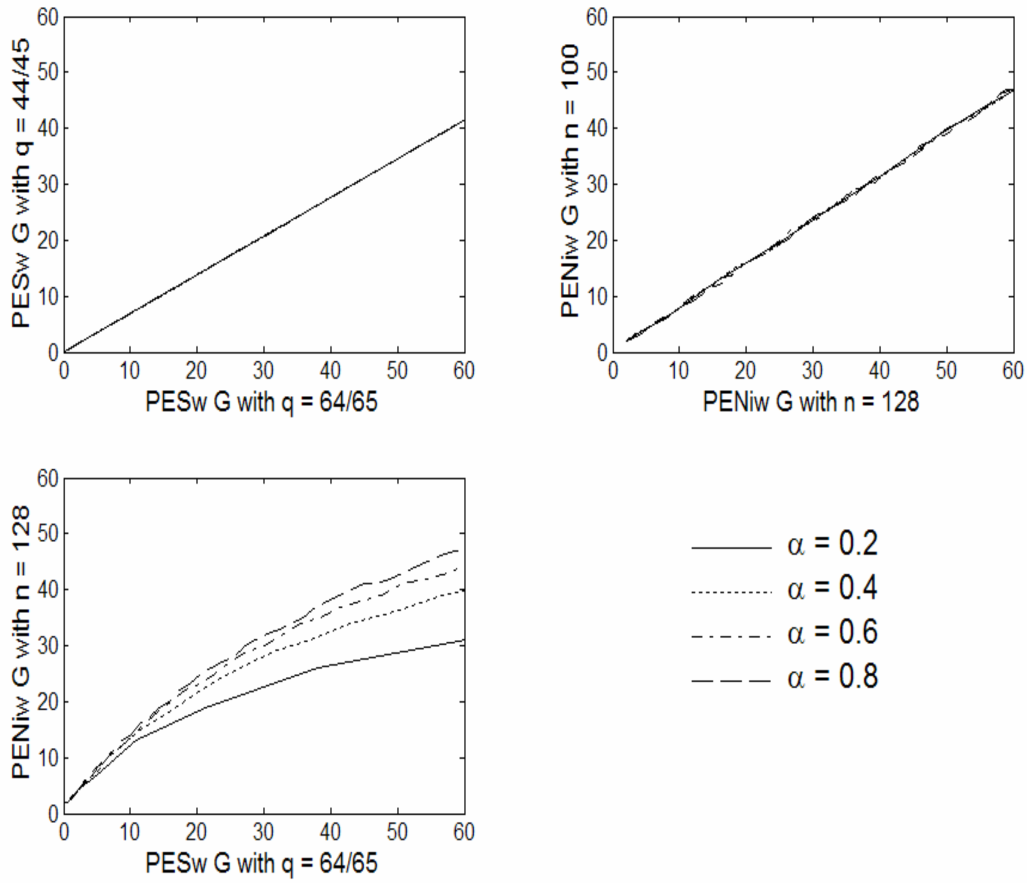

Figure 8 All 3 graphs plot the maximizing cast for the prior evaluation models, G•, for one stochastic process parameter (i.e. q = 64/65) against the corresponding G• values for another parameter (i.e., q = 44/45) across values of γ^+ fixing α. All three plot four lines, one for each specified level of α. The top row demonstrates that γ^+ and α cannot be independently estimated within a release law. However, systematic constraints of the parameters across the release laws allow for γ^+ and α to be identifiable. For example, the bottom panel demonstrates that, setting both γ^+ and α equal across release laws within weather conditions allows both to be estimated.

In addition to Model 2, there are two other possible constraints that result in the identifiability of γ^+ and α, each also with 16 parameters. Model 3 sets γ^+ equal across weather conditions and α equal across release laws ($\gamma^+_{\bullet l}$ and $\alpha_{w\bullet}$). Finally, Model 4 sets γ^+ equal across release laws and α equal across weather conditions ($\gamma^+_{w\bullet}$ and $\alpha_{\bullet l}$).

Before estimating the models briefly recall also in equation 11 that the sunny conditions α and γ^{+} when $G_{\bullet}$ did not change from trip to trip, which occurs in the sunny tournaments and the DM was highly confident of him/herself in the cloudy conditions. Constraining γ^{+} and α in Models 2 through 4 solve this difficulty by increasing the number of $G_{\bullet}$ to at least 2. To compare the models, I will use the log-likelihood of the entire session, *LLs*. Again the models are not nested so the goodness of fit measure can only be used at a descriptive level.

Table 12. Comparisons of models with different constraints to estimate a weighting function, with varying constraints on γ^{+}.

	$0 < \gamma^{+} < 3$		$0 < \gamma^{+} < 1$	
Model	Mean LLs	Num. DMs best fit	Mean LLs	Num. DMs best fit
Baseline	-764.99	2		2
Model 1 (γ^{+}_{wl} and $\alpha_{\bullet\bullet} = 1$)	-270.32	52	-343.98	31
Model 2 ($\gamma^{+}_{w\bullet}$ and $\alpha_{w\bullet}$)	-280.76	2	-282.27	30
Model 3 ($\gamma^{+}_{\bullet l}$ and $\alpha_{w\bullet}$)	-308.76	0	-349.89	8
Model 4 ($\gamma^{+}_{w\bullet}$ and $\alpha_{\bullet l}$)	-298.39	0	-352.17	0

Treating each participant's data for the experimental session separately, Table 12 shows how many participants were best fit by each model using LLs when the valuation parameter is subject to the constraints $0 < \gamma^{+} < 3$. Model 1, which does not incorporate a weighting function and allows γ^{+} to vary between all four tournaments fits a majority of the participants best. However, the first observation resulting from equation 11 points out that the weighting function may serve as an alternative to allowing γ^{+} to go above 1, keeping with a majority of the findings in behavioral decision making (see Gonzalez & Wu, 1998; Kahneman & Tversky, 1979; Luce, 2000; Wu & Gonzalez, 1995; Tversky & Kahneman, 1992). Table 12 shows that with this constraint Model 2 does just as well at fitting individuals as does Model 1. In addition, the mean LLs for Model 1 implies that for some individuals its fit is quite poor. To get a better sense of this result the absolute difference between LLs for Model 1 and Model 2 for each participant were calculated. The average deviation for the 31 individual for which Model 1 was greater than Model 2 was 6.01 (in log-likelihood space). But, for the 30 individuals for which Model 2 was greater than Model 1 the average deviation was 200.65.

This result is not conclusive by any means. Two outcomes, however, encourage me to advance Model 2 with constraints of $0 < \gamma^{+} < 1$ (Model 2*) as best describing the data. First, although forced, the model provides a reasonable fit and conforms to both behavioral decision theory and standard notions of diminishing sensitivity to gains. Second, under this model the MLL parameter estimates have the weighting function partialed from them. Consequently, the correlation between γ^{+} and self-reported risky behaviors can be examined free from any correlation with α. Table 13 summarizes the MLL parameter estimates for Model 2*. The next section explores correlations between the estimates and self-reported risky behaviors and DM's behavior during the fishing tournaments.

Table 13 Summary of MLL parameter estimates for Model 2* and Pearson r correlations with and self-reported drug use. * = $p < 0.05$ & # = $p < 0.01$

MLL Parameters	Average	1st Quartile	Median	3rd Quartile	*DrugSum*	*DrugWeighted*
$z_{s,k}$	-0.02	-0.06	-0.01	0.05	.01	.02
$z_{s,r}$	-0.01	-0.04	0.00	0.02	.12	.16
$z_{c,k}$	0.01	-0.01	0.01	0.05	-.07	-.10
$z_{c,r}$	-0.01	-0.04	0.00	0.04	-.03	-.04
$\beta_{s,k}$	0.42	0.10	0.14	0.19	-.16	-.15
$\beta_{s,r}$	0.12	0.03	0.08	0.17	-.23	-.22
$\beta_{c,k}$	0.27	0.09	0.15	0.26	-.13	-.15
$\beta_{c,r}$	0.29	0.14	0.17	0.29	-.09	-.10
$\gamma^{+}_{s,\bullet}$	0.69	0.41	0.87	0.99	**.36#**	**.34#**
$\gamma^{+}_{c,\bullet}$	0.59	0.43	0.59	0.82	.21	**.24***
$\alpha_{s,\bullet}$	0.59	0.45	0.55	0.89	-.11	-.08
$\alpha_{c,\bullet}$	0.74	0.61	0.76	0.95	.04	.01
μ_{0ck}	147.60	53.12	96.79	186.37	-.06	-.09
$\mathrm{var}\left(\hat{n}_{0ck}\right)$	1.00E+07	114.82	750.20	4019.41	**.40#**	**.43#**
$a_{0c,r}$	9.05E+13	41.78	211.42	707.60	-.08	-.08
$m_{0c,r}$	9.27E+13	42.4248	223.7765	748.165	-.08	-.08
$E\left(\hat{q}_{0c,r}\right)$	0.92	0.96	0.98	0.99	**-.31***	**-.37#**
$\mathrm{var}\left(\hat{q}_{0c,r}\right)$	1.78E-03	1.41E-05	6.50E-05	3.31E-04	-.02	.01

Brief model comparison summary. Before proceeding to the next section, brief reviews of both the modeling process and its conclusions to this point are necessary. The models differed with respect to two factors: (1) the DM's representation of the stochastic process (•Ni or •S) and (2) the point at which he/she evaluates options during a sequential risk-taking task (PE• or SE•).

To test the models, I developed the ART, which better informs participants about the structure of the environment. Manipulations of the stochastic process governing the task, and participants' level of knowledge of the process revealed several results. First, only the prior evaluation class of models could distinguish between the two different stochastic processes. SES, appeared to be too flexible, using LL as a measure of goodness of fit. Second, further investigations with simulations showed that standard measures of goodness of fit may not be able to distinguish between situations when the DMs are either evaluating options sequentially or prior to beginning a task. In addition, a majority of the participants identified the prior evaluation strategy as reflecting their own strategy. Thus, the prior evaluation class of models was taken as a better performing class. The prior evaluation class of models was then extended to include both a trip dependent bias component and a weighting function (Model 2*).

Validity of the model

The purpose of modeling the DM during either the ART or the BART is not solely to arrive at a model that fits the data well. The model also should provide insights into the DM's behavior during and external to the task. To that end, the parameter estimates of the models can be correlated to self-reported risky behaviors as an attempt to externally validate the models. In addition, the model predictions can be studied to gain further insight as to the DM's performance during the fishing tournaments.

Correlation to risky drug use. Recall that the participants completed a questionnaire obtaining their self-reported drug use. The last two columns of Table 13 list the correlations between the parameters and the indexes of self-reported drug use.[10] The primary area of interest focuses on the parameters involved in the DMs' evaluation processes. The table shows a significant correlation between the valuation parameter, $\gamma_{s,\bullet}^{+}$ for the sunny weather conditions and either version of the risky-drug use index. In addition, there is a correlation between the valuation parameter in cloudy weather, $\gamma_{c,\bullet}^{+}$, and the weighted index. The latter result is particularly interesting considering that

[10] None of the MLL parameter estimates correlated with Weber et al.'s (2002) scales designed to predict risky behavior in different domains of life and are not included in the results section.

neither of the ART scores from the cloudy conditions correlated with either index. Finally, neither of the probability weighting function parameters $\alpha_{w,\bullet}$ significantly correlated with risky drug use.

The significant correlation between the variance of the initial gamma distribution, $\text{var}\left(\hat{n}_{0ck}\right)$, and drug use is likely a partial function of a few extreme estimates. Taking the natural logarithm of the MLL estimate reduces the correlation to .22 (p = .07) and .18 (p = .14) for DrugSum and DrugWeighted, respectively. Theoretical interpretation of the correlations will be left for the discussion.

Model accounts of tournament behavior. The models also provide functional insights, statistically and graphically, regarding the behavior of DMs during the task, above and beyond that given by the typical adjusted ART score or other related empirical measures. The mean MLL α estimates in Table 13 indicate that on average individuals were less sensitive to changes in probability in the sunny conditions ($t(70) = 3.80$, $p = .0003$). The MLL γ estimates also hint at a marginal trend of a more linear value function in the sunny conditions, $t(70) = 1.83$, $p = .07$. The remaining possible hypothesis tests with both least squares methods and hierarchical model comparisons did not identify any significant or consistent trends across the tournaments.

Figures 9, 10, 11, and 12 plot predicted probabilities of casting as a function of count number from the prior evaluation models (PENiw and PESw estimated with Model 2 and the more restrictive γ^{+} constraints), holding true to the correct stochastic process. Figures 9 and 10 plot the predictions for three participants in the catch 'n' keep and catch 'n' release*,* sunny weather tournaments, respectively, and shows how the models account for individual differences. Each figure has the predictions for the 1st, 15th and 30th trip for three different participants. Progressing down the two figures, participants reported using fewer drug categories. The predicted curves reflect this by predicting fewer casts as a function of drug use. This trend generally follows across participants, but it is not perfect.

Two other observations are worthy of notice. First, Figures 9 and 10 demonstrate how the bias component, ζ_h, accounts for performance differences between and within participants. Participant 24's MLL estimates reflect a "Testing the waters" strategy in the catch 'n' keep condition (top panel Figure 9), while a "Go for broke" strategy in catch 'n' release (top panel Figure 10). In addition, Participant 20 exhibited little or no bias during the catch 'n' release, sunny tournament (bottom panel Figure 10). Second, the plots also illustrate the fact that the adjusted ART score may be misleading due to its omission of trials resulting in a blue fish. For example, participant 24's adjusted ART score for catch 'n' keep was 39, but his/her model predictions differ substantially from this. In fact, closer scrutiny of his/her data revealed that in this tournament he/she actually made over 60 casts on several trips, but caught a blue fish on a majority of them.

Figures 11 and 12 plot predictions in the catch 'n' keep and catch 'n' release, cloudy weather tournaments for participant 46. Together they illustrate how the models account for learning during the task by plotting trips 1-5, 13 – 17 and 26 – 30 in the top, middle, and bottom panels, respectively, for each model's figure. After experiencing a blue fish (starred lines), both models adjust their prediction downward, but gradually increase after successful trips to a pond (solid lines). The Bayesian learning component also adjusts the predictions as participant 46 progresses in the tournament. Initially, if an individual has low confidence in his/her representation, like participant 46, the models are much more sensitive to successes and failures. But, as he/she progresses in the task the individual settles into his/her beliefs and becomes more consistent in his/her behavior.

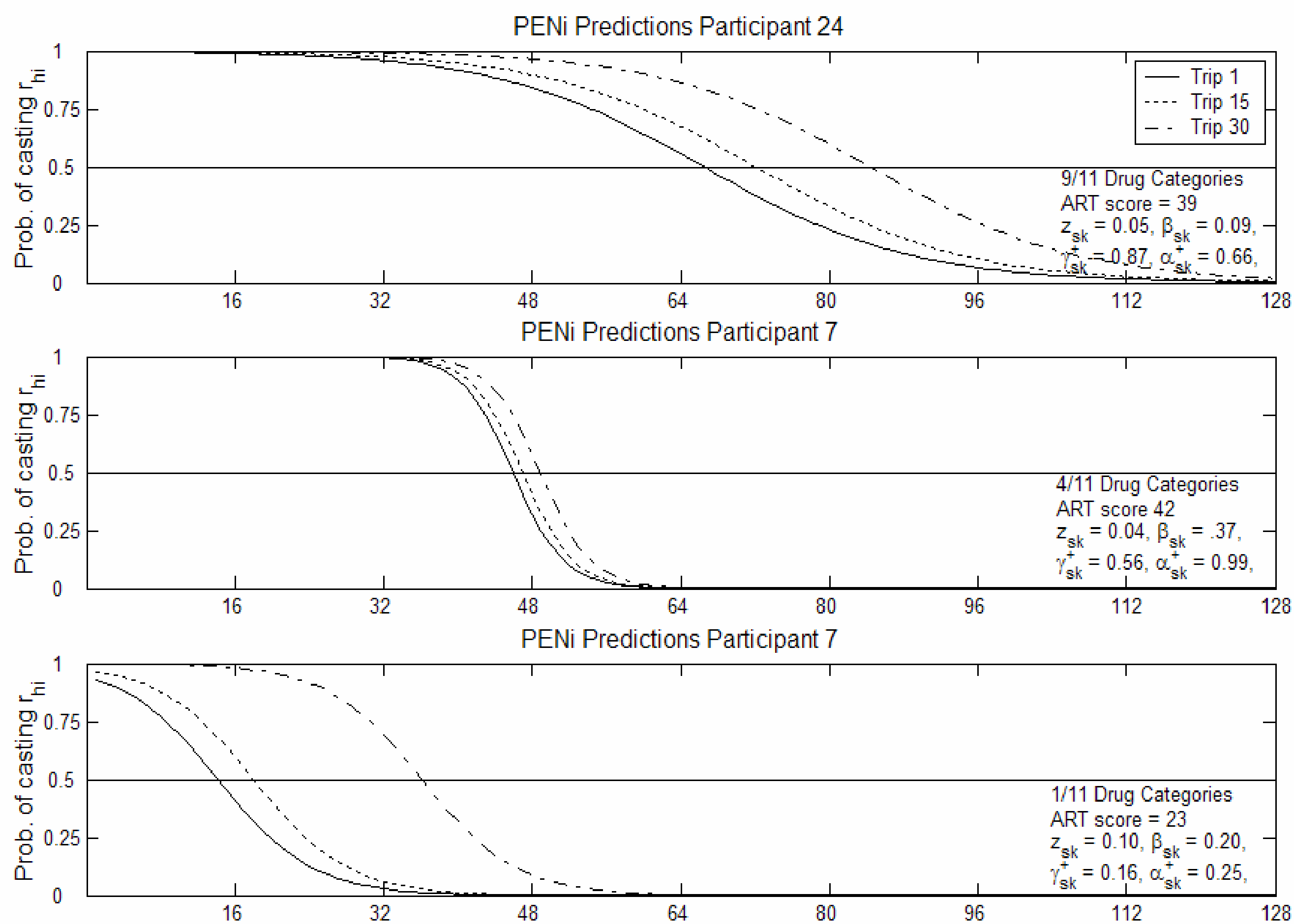

Figure 9 PENiw’s predictions for three participants in the catch ‘n’ keep, sunny fishing tournament. Each panel shows the prediction for Trip 1, Trip 15, and Trip 30.

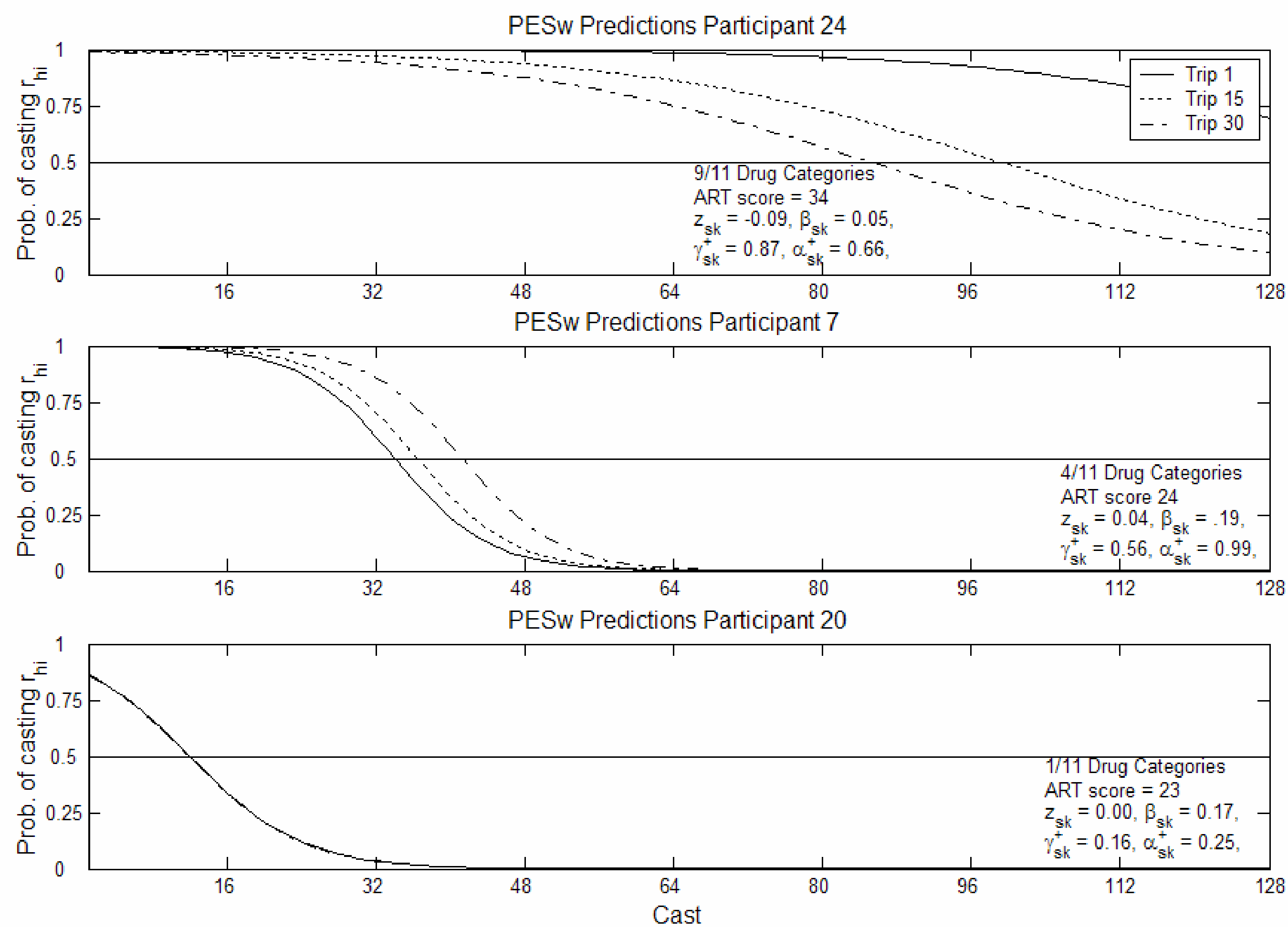

Figure 10 PESw's predictions for three participants in the catch 'n' release, sunny fishing tournament. Each panel shows the prediction for Trip 1, Trip 15, and Trip 30.

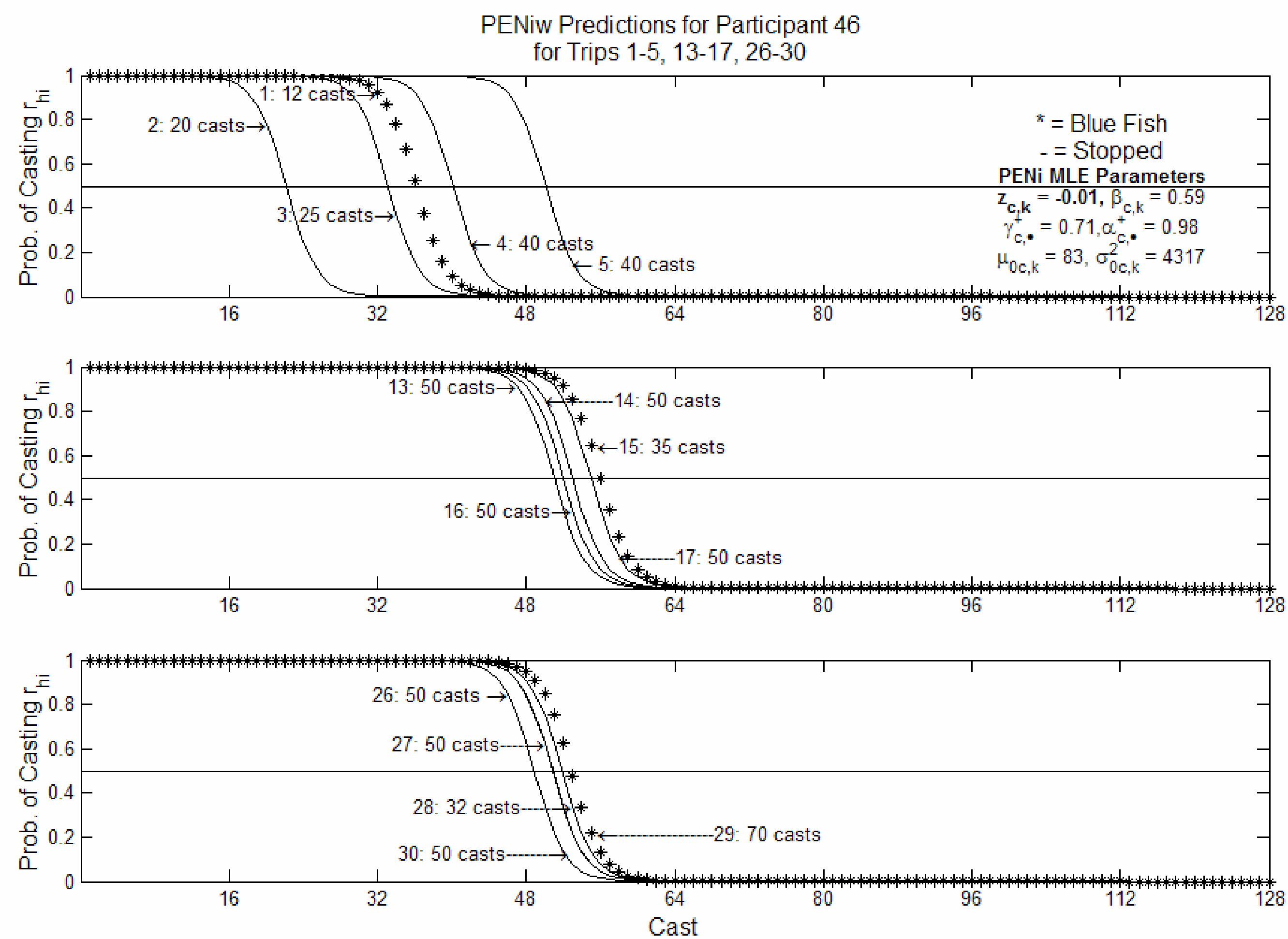

Figure 11 PENiw's predictions for participant 46 in the catch 'n' keep, cloudy fishing tournament. The panel has trips 1-5, middle has 13-17, and bottom has 26-30.

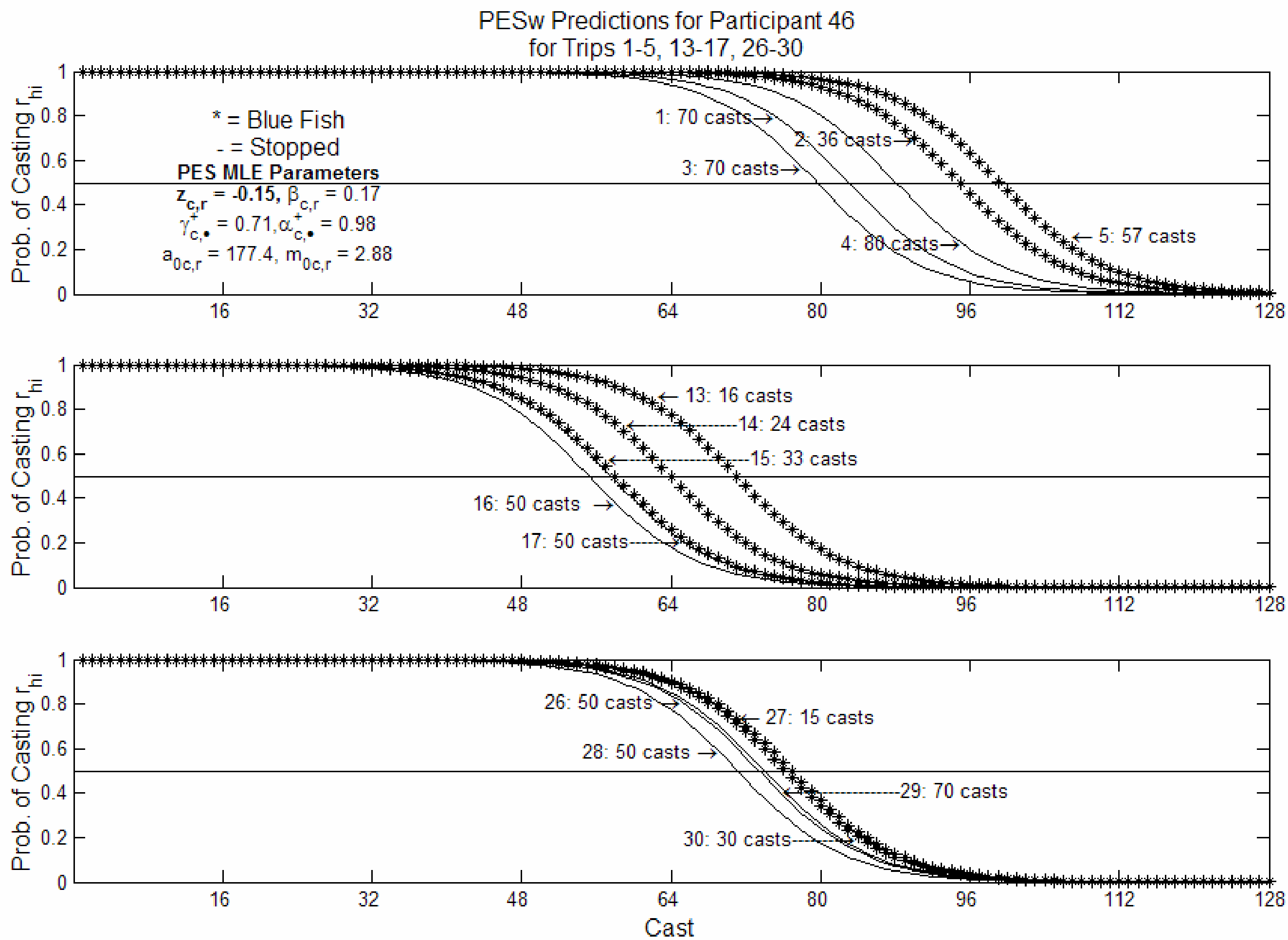

Figure 12 PESw's predictions for participant 46 in the catch 'n' release, cloudy fishing tournament. The panel has trips 1-5, middle has 13-17, and bottom has 26-30.

Discussion

The goals of this dissertation were to evaluate the BART itself and Wallsten et al.'s (2004) four most successful models of the DM completing it as well as to generalize both the paradigm and the models. To that end, the ART was developed. It held true to the general structure of the BART, while simultaneously allowing manipulations of both the environment's actual stochastic structure and the knowledge of the structure. Administering a completely within-subjects design using the ART revealed conclusions about the models and the task.

Modeling conclusions, issues, and insights

Recall that the models differed as to when the DMs evaluated their options and what their beliefs of the task's stochastic process were. Analyses focused on three topics about the models, themselves, (1) their ability to distinguish between the DMs' representation of the stochastic process, (2) their ability to identify the strategy the DMs used to evaluate his/her options, and (3) the incorporation of a weighting function in the models. I will address each in turn, and also how the best-fitting model aids in our understanding of the ART's correlation to risky drug use.

One necessary test, following our previous work, was to examine whether the models could discriminate between circumstances when DMs were and were not fully aware of the stochastic process. As the model estimation section pointed out, the critical subcomponent in this question is the Bayesian learning module. Without it, the estimation of the models either became unattainable, or the MLL parameter estimates were extreme. Additionally, the prior vs. sequential issue makes interpreting model fits difficult. With those contingencies in mind, the present study suggests that SES is too flexible. Using

LL, it fit both cloudy conditions best, regardless of the stochastic process and of which models it was compared against. However, the prior evaluation class of models (PE•) can distinguish between the different representations of the stochastic structures, both with AIC as a measure of fit and when conditionalizing on the point of evaluation. It is interesting to note that this result was identified with a Bayesian learning component, which is surprising considering the majority of present day behavioral decision theories hold the belief that individuals use non-optimal strategies (see Gigerenzer & Todd, 1999; Tversky & Kahneman, 1974).

Whether DMs plan their behavior prior to each trial, or they sequentially evaluate options, is less clear. Without accounting for model complexity, the LL comparisons suggest that DMs sequentially evaluate their options. But, using AIC to handicap models with more parameters, the comparisons suggest the prior evaluation method. Unfortunately, simulations showed that this same pattern held under conditions when the data were generated with either type of evaluation. This result could be due to the measures not distinguishing between the two strategies or the models making indistinguishable predictions. Both are possible and future work is needed, as the complexity of models is clearly an issue.

This is not the first time AIC has been shown to fail as a measure of a model's goodness of fit (see Myung, 2000). Other measures do exist, such as the Bayesian information criterion (BIC; Schwarz, 1978), the minimum description length (MDL; Pitt, Myung, & Zhang, 2002), or Bayesian model selection (BMS; Myung & Pitt, 1997). The latter two require the a priori specification of the prior distribution over parameters. All three require better specification of the degree of dependence among either the data points themselves or among the parameters. This is beyond the scope of this paper, but certainly must receive future attention.

Regardless, SES does not even distinguish among the different stochastic representations of DMs. Based on this result and the fact that a majority of participants identified prior evaluation as their own personal strategy, the prior evaluation class of models was put forth as better describing the data. Subsequently, they were extended to include a bias component and a weighting function. The bias component assisted the models in accounting for behavior participants reported using, (i.e., *testing the waters* or *going for broke*).

Turning to the weighting function, its necessity in this venture seems, in part, contingent on whether the value function is allowed to show increasing sensitivity. Certainly, the weighting function seems necessary to account for the typical *decisions from description* demonstrations used in prospect theory (e.g., choose between two lotteries A and B). But, choices in the ART would be classified as *decisions from experience* (see Hertwig, Barron, Weber, & Erev, *in press*). Incorporating the weighting function did indicate, however, differences in how individuals responded to the task. Surprisingly, it indicated that they were more sensitive to changes in probability in the cloudy tournaments; though, this may be because in cloudy conditions the weight was a function of the DMs' subjective probability, while in sunny conditions it was a function of the objective probability.

Incorporating the weighting function also allowed me to investigate the role that the value function, weighting function, and the remaining cognitive processes, have in the ART's correlation to risky drug use. The results indicate that the primary reason for its correlation resides in the DMs' attitudes towards outcomes, not their perceptions of the probabilities. Although this outcome may change for different domains of risk it is consistent with our previous findings. The remaining strong correlation with drug use was with the initial subjective probability of catching a red fish in the catch 'n' release, cloudy tournament. No theoretical explanation seems tenable, but one way to interpret this result is that less risky individuals simply thought there were more red fish in the pond. An alternative explanation is that these individuals are using an alternative model, not specified here, for these conditions.

A final point to be made in terms of the models is that the adjusted ART score and other similar measures ignore a substantial portion of the data. That is, it ignores all the trips on which a blue fish was caught (e.g., when perhaps the DM cast over 64 times). In the present study, blue fish were caught on about 1/3 of the trips. The cognitive models presented here use all the data and also allow individual differences among the DMs. In general, they suggest that DMs evaluate their options prior to beginning each trip, but, their evaluations are affected by their risk aversion/seeking. DMs then probabilistically choose to cast or not based on their evaluation, their sensitivity to the evaluation and their biases. Finally, when necessary, DMs learn about the task in Bayesian fashion, but again this depends on the person. Some are too confident in their representation of the task to change their opinion.

Task conclusions

To be sure, the BART is not precisely the ART. Conceptually, the BART lies somewhere in between the cloudy and sunny conditions with a catch 'n' keep law. The strength of the models is their ability see beyond the tasks' surface differences to the similarities beneath. With that, the results presented here are beneficial for future versions of the BART as for other gambling tasks.

From a modeling perspective, one of the interesting challenges with the BART has been its vagueness from the DM's perspective as to the process underlying the balloon's explosion. However, for the purpose of risk assessment, this elusiveness appears to be a downfall, adding unnecessary noise to the BART's correlation to risky behaviors. Both the analyses with the adjusted ART score and the cognitive models support this conclusion. In terms of the adjusted ART score, only conditions that revealed the stochastic structure (i.e., sunny) correlated with self-reported risky behaviors. Concealing the fish removed this correlation. Modeling the individuals' cognitive processes removed this problem by accounting for their initial opinion about the pond's structure. Consequently, the valuation parameter correlated with risky-drug use both when the DM knew and did not the number of fish in the pond. The BART is not the only task that incorporates a learning process. Busemeyer and Stout (2002) have found that a cognitive learning process is also involved in the Bechara card sort task. The results from the present experiment imply that the learning component of these tasks may hinder their clinical and neurological model-free assessment of risk taking.

The implications of different stochastic processes are less clear on this paradigm of laboratory-based gambling tasks. Participants both tended to report changing their behavior and appeared to change their behavior between catch 'n' keep and catch 'n' release. The clear inevitability of a blue fish in the catch 'n' keep tournaments appears to have increased the variation among individuals, as can be seen in Figure 6. The models themselves, in their present form, did not show any systematic interpretable differences, nor did the correlations with the adjusted ART score. Thus, no conclusive suggestions as to which process is better at correlating or predicting an individual's propensity towards risky behavior can be given. Perhaps different more disparate structures are needed.

Future directions

Like any good fishing story, this one may leave the reader wondering about the one that got away. In particular, the degree to which the ART/BART differentially correlates with certain domains of risk remains unaddressed. Although at the empirical level, the ART's sunny conditions correlated with Weber et al.'s (2002) social domains of risk, more work in this area is needed and seems promising. For instance, it remains to be seen whether the concepts developed here can be used to develop a larger class of gambling tasks each with different narratives and/or stochastic structures that differentially identify individuals predisposed to particular domains of risk

Appendix A

The discretized gamma distribution over *n*, *p(n)*, is updated with Bayes' Rule. The updated distribution does not necessarily retain the properties of the discretized gamma distribution. Consequently, the process is not straightforward. Wallsten et al. (2004) originally formalized the process, I will reconstruct it here.

To update the distribution consider the case of observing *c* pumps followed by the balloon exploding (or in ART's case a blue fish is caught) on the last opportunity. Recall that the expression, $p_{hi} = 1/(n-i+1)$, is the probability of an explosion after i -1 successful pumps. Therefore, the probability of no explosion is $q_{hi} = 1 - p_{hi} = (n-i)/(n-i+1)$**.** The probability of *c* pumps followed by an explosion (pop) given *n* is

$$p(c\, pumps\, with\; pop \mid n) = \begin{cases} \frac{n-1}{n}\frac{n-2}{n-1}\cdots\frac{n-c+1}{n-c+2}\frac{1}{n-c+1} = \frac{1}{n}, & when\; n \geq c \\ 0, & when\; n < c \end{cases} \tag{A1}$$

Thus, any sequence of pumps resulting in the balloon exploding has probability, $1/n$. Similarly, the probability of a sequence of *c* pumps without the balloon exploding is

$$p(c\, pumps\, with\; no\; pop \mid n) = \begin{cases} \frac{n-1}{n}\frac{n-2}{n-1}\cdots\frac{n-c+1}{n-c+2}\frac{n-c}{n-c+1} = \frac{n-c}{n}, & when\; n \geq c \\ 0, & when\; n < c \end{cases} \tag{A2}$$

. Thus, in general after *h* balloons the result can be expressed as

$$p(c_1, d_1, c_2, \cdots, c_h, d_h) = \prod_{h'=1}^{h}\left[\frac{s_{h'}(n-c_{h'})^{d_{h'}}}{n^{h'}}\right] \tag{A3}$$

where $d_{h'} = \begin{cases} 0\; if\;\; balloon\; h'\; popped \\ 1\; if\;\; balloon\; h'\; did\; not\; pop \end{cases}$ and $s_{h'} = \begin{cases} 0\; if\;\; n < c* \\ 1\; if\;\; n \geq c* \end{cases}$, where

$c* = Max(c_1, c_2, \cdots, c_h)$.

With Bayes' rule we can now obtain the expression for the updated distribution over n, $p\left(n \mid c_1, d_1, \cdots, c_h, d_h\right)$ The result is

$$p\left(n \mid c_1, d_1, \cdots, c_h, d_h\right) = \frac{\prod_{h'=1}^{h}\left[s_{h'} \left(n - c_{h'}\right)^{d_{h'}} p(n) \big/ n^{h'} \right]}{\sum_{n'=1}^{\infty} \prod_{h'=1}^{h}\left[s_{h'} \left(n' - c_{h'}\right)^{d_{h'}} p(n') \big/ n'^{h'} \right]} \tag{A4}$$

(see also equation 1).

Appendix B

The baseline model is a statistical model estimated directly from the data. In addition to describing the data, cognitive models are compared against it. The model is based on the binary event of the DM pumping the balloon or not at each opportunity. Collapsing across *h* balloons, the probability, r_i, that the DM pumps on opportunity is the proportion of opportunities over all balloons that the DM chose to pump, or

$$\hat{r}_i = \frac{1}{H_i}\sum_{h=1}^{H} w_{hi} \qquad \text{A1}$$

where H_i is the total Number of balloons that did not explode prior to opportunity *i* and $w_{h,i}$ = 1 or 0 if the balloon is pumped or not, respectively.

A critical assumption is that if DM chose to stop on opportunity *i*, then we assume that he/she would choose to stop on all subsequent opportunities. No such assumption is made for the balloons which exploded. However, this assumption renders the individual $\hat{r}_i$ non-independent of each other. We will use the maximum Number of opportunities over all 30 balloons/trips for which the DM actually made a choice to pump or stop as an estimate of the *df*. A detailed discussion of this assumption is in Wallsten et al. (2004).

Appendix C

The Nelder Mead downhill simplex routine (see Nelder & Mead, 1965) in conjunction with a grid-search technique uses a 2 step approach to arrive at a solution. During the first step, I divided the parameter space into three plausible sectors. For example, the plausible space for γ^+ was set between 0 and 3, but the divisions were weighted toward the lower spectrum of the space, (0, 0.5), (0.5, 1.5), and (1.5, 3). A starting value for each parameter was then randomly selected from one of its divisions. The set of starting values was then tested to insure the starting values would lead to a solution below a pre-specified criterion (e.g., $\ln(L) > -2000$). If not, then the set was iteratively perturbed with random noise, and tested, until the criterion was met or a cutoff was reached. If the set did meet the criterion, then they were input into the Nelder-Mead method, beginning the second step. The full two step process was then repeated for 50 to 100 iterations. The maximum ln(L) from the full set was taken as the MLL estimate.

I tried many other procedures. Examples of such procedures include: nonlinear programming, genetic algorithm, or Van Zandt's iterative annealing Nelder Mead method (see Van Zandt, 2000; Van Zandt, Colonius, & Proctor, 2000). However, simulations showed that the aforementioned procedure performed the best.

References

Akaike, H. (1973). Information theory and an extension of the maximum likelihood principle. In B. N. P. F. Caski (Ed.), *Second international symposium on information theory* (pp. 267-281). Budapest, Hungary: Akademiai Kiado.

Bechara, A., Damasio, A. R., Damasio, H., & Anderson, S. W. (1994). Insensitivity to future consequences following damage to human prefrontal cortex. *Cognition, 50*(1-3), 7-15.

Busemeyer, J. R., & Stout, J. C. (2002). A contribution of cognitive decision models to clinical assessment: Decomposing performance on the Bechara gambling task. *Psychological Assessment, 14*(3), 253-262.

Gigerenzer, G., & Todd, P. M. (1999). *Simple heuristics that make us smart.*: Oxford University Press.

Gonzalez, R., & Wu, G. (1999). On the Shape of the Probability Weighting Function., *Cognitive Psychology* (Vol. 38, pp. 129).

Hertwig, R., Barron, G., Weber, E.U., & Erev, I. (in press). Decisions from experience and the effects of rare events in risky choices. *Psychological Science.*

Hoffrage, U., Weber, A., Hertwig, R., & Chase, V. M. (2003). How to Keep Children Safe in Traffic: Find the Daredevils Early. *Journal of Experimental Psychology: Applied, 9*(4), 249-260.

Kahneman, D., & Tversky, A. (1979). Prospect theory: An analysis of decision under risk. *Econometrica, 47*(2), 263-291.

Lejuez, C. W., Aklin, W. M., Jones, H. A., Richards, J. B., Strong, D. R., Kahler, C. W., et al. (2003a). The balloon analogue risk task (BART) differentiates smokers and nonsmokers. *Experimental & Clinical Psychopharmacology, 11*(1), 26-33.

Lejuez, C. W., Aklin, W. M., Zvolensky, M. J., & Pedulla, C. M. (2003b). Evaluation of the Balloon Analogue Risk Task (BART) as a predictor of adolescent real-world risk-taking behaviours. *Journal of Adolescence, 26*(4), 475-479.

Lejuez, C. W., Read, J. P., Kahler, C. W., Richards, J. B., Ramsey, S. E., Stuart, G. L., et al. (2002). Evaluation of a behavioral measure of risk taking: The Balloon Analogue Risk Task (BART). *Journal of Experimental Psychology: Applied, 8*(2), 75-84.

Luce, R. D. (2000). *Utility of gains and lossess measurement-theoretical and experimental approaches*. London: Lawrence Erlbaum Associates.

Myung, I. J. (2000). The importance of complexity in model selection. *Journal of Mathematical Psychology, 44*(1), 190-204.

Myung, I. J., & Pitt, M. A. (1997). Applying Occam's razor in modeling cognition: A Bayesian approach. *Psychonomic Bulletin & Review, 4*(1), 79-95.

Nelder, J. A., & Mead, R. (1965). A Simplex-Method for Function Minimization. *Computer Journal, 7*(4), 308-313.

Pitt, M. A., Myung, I. J., & Zhang, S. (2002). Toward a method of selecting among computational models of cognition. *Psychological Review, 109*(3), 472-491.

Prelec, D. (1998). The probability weighting function. *Econometrica, 66*(3), 497-527.

Schwarz, G. (1978). Estimating the dimension of a model. *Annals of Statistics, 6*(2), 461.

Slovic, P. (1966). Risk-taking in children: Age and sex differences. *Child Development, 37*(1), 169-176.

Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. *Science, 185*(4157), 1124-1131.

Tversky, A., & Kahneman, D. (1992). Advances in Prospect-Theory - Cumulative Representation of Uncertainty. *Journal of Risk and Uncertainty, 5*(4), 297-323.

Tversky, A., & Wakker, P. (1995). Risk Attitudes and Decision Weights. *Econometrica, 63*(6), 1255-1280.

Van Zandt, T. (2000). How to fit a response time distribution. *Psychonomic Bulletin & Review, 7*(3), 424-465.

Van Zandt, T., Colonius, H., & Proctor, R. W. (2000). A comparison of two response time models applied to perceptual matching. *Psychonomic Bulletin & Review, 7*(2), 208-256.

Wallsten, T. S., Pleskac, T. J., Lejuez, C. W. (2004). Modeling a sequential risk-taking task. *Submitted for publication*.

Weber, E. U., Blais, A.-R., & Betz, N. E. (2002). A domain-specific risk-attitude scale: Measuring risk perceptions and risk behaviors. *Journal of Behavioral Decision Making, 15*(4), 263-290.

Wu, G., & Gonzalez, R. (1996). Curvature of the probability weighting function. *Management Science, 42*(12), 1676-1690.

Wu, G., & Gonzalez, R. (1999). Nonlinear decision weights in choice under uncertainty. *Management Science, 45*(1), 74-85.

Zuckerman, M., Ball, S., & Black, J. (1990). Influences of sensation seeking, gender, risk appraisal, and situational motivation on smoking. *Addictive Behaviors, 15*, 209.