Journal of Experimental Psychology:
Learning, Memory, and Cognition
2008, Vol. 34, No. 1, 167–185

0278-7393/08/$12.00 DOI: 10.1037/0278-7393.34.1.167

Decision Making and Learning While Taking Sequential Risks

Timothy J. Pleskac
Indiana University

A sequential risk-taking paradigm used to identify real-world risk takers invokes both learning and decision processes. This article expands the paradigm to a larger class of tasks with different stochastic environments and different learning requirements. Generalizing a Bayesian sequential risk-taking model to the larger set of tasks clarifies the roles of learning and decision making during sequential risky choice. Results show that respondents adapt their learning processes and associated mental representations of the task to the stochastic environment. Furthermore, their Bayesian learning processes are shown to interfere with the paradigm's identification of risky drug use, whereas the decision-making process facilitates its diagnosticity. Theoretical implications of the results in terms of both understanding risk-taking behavior and improving risk-taking assessment methods are discussed.

Keywords: risk taking, learning, Bayesian, individual differences, cognitive psychometrics

Learning and decision making are conceptually linked. Typically only after decision makers (DMs) make a decision do they observe or experience the precise outcome of that decision. For example, only after commuters select a traffic route do they determine its effectiveness, and only after athletes use a steroid do they learn about the precise properties it has on their body. These observations better inform DMs about the precise properties of their choice options and shape their next decision among the same or similar options. Despite this natural association between decision and learning processes, most decision-making theories fail to incorporate or explicate a learning component (e.g., Busemeyer & Townsend, 1993; González-Vallejo, 2002; Kahneman & Tversky, 1979). Yet, how DMs learn from experience has proven an important process in understanding risk-taking behavior. It can, for example, create an aversion toward risky alternatives in the gain domain and an attraction toward risky alternatives in the loss domain—a pattern typically attributed to how DMs evaluate outcomes (Denrell, 2007; March, 1996). The learning process can even produce the opposite pattern (Erev & Barron, 2005; Hertwig, Barron, Weber, & Erev, 2004; Weber, Shafir, & Blais, 2004).

Applying theories of decision making to the Balloon Analogue Risk Task (BART; Lejuez et al., 2002) or to the Iowa Gambling Task (Bechara, Damasio, Damasio, & Anderson, 1994) also exposes the necessity of learning. Clinicians use these laboratory-based gambling tasks to study and identify people with specific clinical or neurological deficits. Cognitive models of these tasks reveal that decision and learning processes are necessary to account for choices made by both clinical and normal populations (Busemeyer & Stout, 2002; Wallsten, Pleskac, & Lejuez, 2005). Besides describing behavior during the tasks, the models also show how the populations differ on the underlying cognitive dimensions captured within the models. For example, during the BART, Wallsten et al. (2005) found that people who take unhealthy and unsafe risks tend to differ from normal populations in how they evaluate payoffs and the consistency of their responses.

What is unclear, however, is the extent to which real-world risk takers systematically differ in the learning process used during the BART. There are two possible roles the learning process could play. On the one hand, learning may not play a significant role at all. This prediction comes from studies showing Slovic's (1966) devil task discriminates between risk takers without requiring learning (Hoffrage, Weber, Hertwig, & Chase, 2003). Slovic's devil task (henceforth devil task) is of interest because it has an identical structure to the BART but does not require learning. On the other hand, the learning process may aid in the BART's ability to identify risk takers. In the Iowa Gambling Task, for example, the learning process is necessary to understand how specific clinical deficits lead to individual differences in risky decision making (Stout, Rock, Campbell, Busemeyer, & Finn, 2005; Yechiam, Busemeyer, Stout, & Bechara, 2005). To directly address these conflicting predictions, this article develops a larger class of sequential risk-taking tasks: the Angling Risk Tasks (ART). This

---

Timothy J. Pleskac, Cognitive Science Program, Indiana University.

A portion of this article is based on a dissertation submitted in partial fulfillment of the doctoral requirements at the University of Maryland–College Park. Parts of this article were also presented as posters at the 2004 annual meeting of the Society for Judgment and Decision Research, Minneapolis, Minnesota, and the 2005 biannual meeting of the European Association of Decision Making, SPUDM-20, Stockholm. National Institute on Drug Abuse Grant R21-DA14699 (principal investigator: Carl W. Lejuez) partially funded the study, and National Institute of Mental Health Research Service Award MH019879, awarded to Indiana University, supported the writing of this article.

I extend my gratitude to Thomas S. Wallsten, Carl W. Lejuez, and Ralph Hertwig for their continuous support throughout this project. I thank Neil Bearden, Ana Franco-Watkins, and Eldad Yechiam for constructive comments on an earlier version of this article. I also thank Anthony Bishara, Sarah Queller, Julie Stout, Ido Erev, members of the Clinical and Cognitive Neuroscience Laboratory at Indiana University, and members of the ABC Research Group at the Max Planck Institute for Human Development for their input on the project. Finally, I am indebted to Laura Wiles for detailed editing of the article.

Correspondence concerning this article should be addressed to Timothy J. Pleskac, who is now at the Department of Psychology, 282A Psychology Building, Michigan State University, East Lansing, MI 48824. E-mail: tim.pleskac@gmail.com

larger class experimentally dissociates the contribution of learning from the decision processes in a model-free manner. Consequently, we can address whether the learning process gives rise to systematic individual differences between populations or not.

Equally important to understanding how individual DMs differ in their learning process is to identify how the learning process changes across different task environments. One possibility is that DMs use the same learning process regardless of the task environment. This prediction is derived from the counterintuitive finding that DMs mentally model the BART, a nonstationary probabilistic environment, as a stationary or sampling-with-replacement environment (Wallsten et al., 2005). The mental representation in turn affects the specific learning process used. One would then expect that during a task with a stationary stochastic environment DMs would adopt the same stationary mental representation. Alternatively, akin to findings in multiattribute decision making, DMs may use different processes in different environments (see Gigerenzer, Todd, & the ABC Research Group, 1999; Payne, Bettman, & Johnson, 1993). To address this issue, DMs completed the ART with different probabilistic structures (stationary and nonstationary probabilistic environments). Anticipating the results, this article will show that DMs do adjust their mental representations to different stochastic environments. Furthermore, during the ART, in stark contrast to during the BART, they adopt mental representations that match the task's actual probabilistic structure.

The article is structured as followed. First, I introduce the sequential risk-taking paradigm and compare two exemplars from it: the BART and Slovic's (1966) devil task. Next, to derive predictions about how learning may either aid or obstruct in identifying risk takers, I introduce Wallsten et al.'s (2005) Bayesian sequential risk-taking model (BSR) of the paradigm. I show how the model can be adapted to account for different mental representations of different stochastic environments. Then to test the predictions, I develop the ART and present the results of a study. During the experiment, participants played the four different conditions of the ART, reported their past drug-use activity, and completed the Domain-Specific Risk-Attitude Scale (DOSPERT; Weber, Blais, & Betz, 2002). The results are analyzed in terms of the cognitive processes at play during the four tasks and how risky drug users differ on the underlying cognitive dimensions of the model. Finally, in the Discussion section I address how the tandem use of cognitive models and the sequential risk-taking paradigm can lead to a more precise understanding of the underlying processes used in sequential risky choice.

The Sequential Risk-Taking Paradigm

Two exemplars of the sequential risk-taking paradigm are the BART and devil task. Both identify real-world risk takers. Performance on the BART, for instance, correlates with a variety of self-reported risky behaviors such as drinking alcohol, smoking cigarettes, using illegal drugs, gambling, not wearing a seatbelt, engaging in unprotected sex, and stealing (Lejuez, Aklin, Jones, et al., 2003; Lejuez, Aklin, Zvolensky, & Pedulla, 2003; Lejuez et al., 2002; Lejuez, Simmons, Aklin, Daughters, & Dvir, 2004). The devil task distinguishes among risk takers at different developmental stages (Montgomery, 1974; Slovic, 1966) and discriminates between children who cross the street dangerously (Hoffrage et al., 2003).

Both the BART and the devil task require participants to play the same structural game for up to a maximum of n choice trials. On each trial they make a choice between two options, a *stop* and a *play* option. If DMs choose to stop, then the game ends and they collect the reward in the bank. If they choose the risky play option, then with some probability one of two possible events occurs. A *successful event* can happen in which a constant reward is deposited into the bank, and DMs proceed to the next trial. Alternatively, a *failure event* can occur, ending the game, and participants lose the accumulated reward in the bank for that round.[1]

In both the BART and the devil task, the same probabilistic structure determines the probability of a failure event (and by implication the successful event). For each round, one of the n choice trials is randomly chosen to produce the failure event. Thus, all possible trials are a priori equally likely to result in a failure event. Yet, like drawing balls from an urn and not replacing them, the probability of a failure increases with each successful play option, making the probability of a successful event nonstationary. The games are typically played anywhere between 1 and 90 rounds, and at the beginning of each round the bank is empty.

Despite their underlying similarities, the BART and the devil task differ in the amount of information participants know about the task. The devil task is well defined for respondents. During the game participants—often children—are presented with a board with $n = 10$ switches making them visually aware of the number of choice trials available for each round.[2] True to the structure of the paradigm, nine of the switches produce a sticker or some fixed amount of candy when pulled, and the remaining switch yields a failure event (a devil sticker). Again, players must choose when to stop the round and collect their cumulated reward; otherwise, if they pull the devil the round ends and they lose their collected prizes for that round.

The BART, in comparison, is ill defined for players.[3] During the game, a balloon is shown on a computer monitor and participants inflate it by pressing a button at the bottom of the screen. With each successful inflation they receive a fixed amount of money, x¢. If the balloon explodes (a failure), participants lose the money for the round. The task is usually constructed to allow a total of $n = 128$ total possible inflations (choice trials) per round. The BART is ill defined because participants are unaware of the possible choice trials for each balloon, nor are they explicitly told that the chances of a failure (explosion) increase with each successful choice trial. Participants are told only that (a) they will earn x¢ with each successful pump option; (b) the balloon will explode somewhere between the first pump and when it fills the screen; (c) they must decide when to stop pumping and collect their reward

[1] This is a different structure from bandit problems (see Berry & Fristedt, 1985), typically used to study how DMs learn in uncertain situations. During bandit problems, DMs make repeated choices between two or more initially unknown options, and unlike in the sequential risk-taking paradigm, DMs have to learn the possible payoffs as well as their distribution associated with each option.

[2] The children are also shown a graph explaining that the likelihood of the failure event increases with each pulled switch (see Hoffrage et al., 2003; Jamieson, 1969; Slovic, 1966).

[3] The terms *ill-defined task* and *well-defined task* come from experimental economics, where they are frequently used in discussions about deception (see Hertwig & Ortmann, 2001; Hey, 1998).

before an explosion ends their round; and (d) they will play the BART for 30 rounds (Lejuez et al., 2002).

The success of both the BART and the devil task in identifying real-world risk takers suggests that the ill-defined characteristic of the BART may not be a necessary element to identify real-world risk takers. In fact, one aspect of the BSR model suggests that not fully informing respondents about the BART may even hurt its clinical diagnosticity. At the same time, though, the model indicates that learning might help its diagnosticity. These conflicting predictions are not intuitive, so in the next section I introduce the BSR model and the underlying predictions.

The BSR Model

The BSR model was the best performing model of the many Wallsten et al. (2005) tested, accounting for both choice behavior during the BART and correlated with self-reported unhealthy and unsafe risky behaviors.[4] The model posits that DMs use three cognitive processes to complete a sequential risk-taking task: evaluation, response, and learning. Each component has at least one free parameter that is used to identify how each individual DM quantitatively differs on the underlying process. Each process is briefly described next, beginning with the evaluation process.

Evaluation

To make a choice between playing or stopping, DMs within the model adopt a choice policy whereby prior to each round they identify how many trials they should play to maximize expected gains. That is, at the beginning of each round, DMs can earn x¢, $2x$¢, . . ., ix¢, . . ., nx¢ on choice trials 1, 2, . . ., i,. . ., n with subjective probability $\pi_h(1)$, $\pi_h(2)$, . . ., $\pi_h(i)$, $\pi_h(n)$. Thus, the expected gain on round h for each choice trial i is

$$v_{h,i} = \pi_h(i)^*(ix)^{\gamma^+}. \tag{1}$$

The exponent γ^+ is akin to prospect theory's diminishing sensitivity parameter where γ^+ must be greater than 0 and typical participants have values less than 1 (Kahneman & Tverksy, 1979; Tversky & Kahneman, 1992). Lower values of γ^+ indicate less sensitivity to changes in payoffs, and higher values indicate greater sensitivity.[5] The precise subjective probability of a success $\pi_h(i)$ depends on DMs' mental representations of the stochastic process (stationary or nonstationary) and their cumulative experience from the past $h - 1$ rounds with the task. The best fitting BSR in the BART assumes participants have a stationary representation, where the probability of a success on any given trial is $\hat{q}_h$ and the subjective probability of i successes is $\pi_h(i) = \hat{q}_h^i$. The Bayesian learning process (specified later) describes how $\hat{q}_h$ changes with cumulative experience in the task.

The model assumes DMs identify the choice trial that maximizes expected payoffs and use it as a targeted trial, G_h. Optimizing Equation 1—assuming a stationary stochastic process—produces the following closed form solution:

$$G_h = \frac{-\gamma^+}{\ln(\hat{q}_h)}. \tag{2}$$

Equation 2 illustrates how DMs with different values of γ^+ will behave differently during sequential risk-taking tasks. Respondents with larger values of γ^+ will have a target that is greater than those for people with lower values of γ^+ and will typically choose the play option for more trials in a given round.

Response

DMs use the targeted trial G_h to probabilistically choose between playing or stopping on each trial. The response rule assumes that the probability of choosing the play option, $r_{h,i}$, during round h on trial i strictly decreases with each choice. Formally, the response rule

$$r_{h,i} = \frac{1}{1 + \exp\,(\beta d_{h,i} - \zeta_h)} \tag{3}$$

captures these properties, where $d_{h,i} = i - G_h$ and β is a free parameter representing how consistently DMs follow their targeted evaluation. DMs with lower values of β will be more variable in their overall gambling behavior during a sequential risk-taking task.

Departing from the model in Wallsten et al. (2005), the response function in Equation 3 also contains a response bias module, ζ_h, allowing DMs to have a bias in their response that changes over rounds. Some DMs, as Wallsten et al. (2005) report, have an exploratory bias whereby during the first few rounds they continue choosing the play option past the maximizing trial. Other participants nearing the end of the task tend to choose the play option much more than would be expected, perhaps exhibiting a house–money effect (e.g., Thaler & Johnson, 1990). To account for these two biases, ζ_h is set equal to the following expression:

$$\zeta_h = \exp[z(h - H/2)] - 1,$$

where H is the total number of rounds that DMs played during the task. The free parameter z identifies the type of bias a DM has. Negative and positive values of z characterize an early exploratory or a late bias, respectively. If $z = 0$, then the participant exhibits no round-dependent bias.

Learning

Reacting to participants' partial ignorance in ill-defined tasks like the BART, the BSR allows DMs to have different mental representations of the stochastic process and assumes they engage in a Bayesian learning process to discover the parameters of the task, given their mental representation. During the BART, participants typically represent the task as a stationary one (e.g., drawing successive balls from an urn and replacing them), despite its true nonstationary stochastic structure. This mental representation and Bayesian learning process is described next.

Stationary representation. Before the first round of each game, DMs have a prior opinion about the probability of a success, q. The prior beliefs are modeled with beta distributions, $f_1(q)$,

[4] The cognitive model is technically number 3 in Wallsten et al. (2005).

[5] Because of the nature of the BART, a probability weighting function could not be estimated, and so now weighting function was assumed. However, the larger set of tasks developed here does allow both the value and weighting function to be estimated for a given respondent and is therefore addressed later in the article.

whereby the mean of the beta distribution, $\hat{q}_1$, represents the estimated subjective probability of a success on the first round for any given trial and is used in Equation 2. It is a free parameter in the models.[6] DMs with higher levels of $\hat{q}_1$ are more optimistic and will tend to choose the play option more than would DMs with lower levels of $\hat{q}_1$.

After each round, DMs observe the total number of successes, c_h, and whether the round ended in a failure ($d_h = 1$) or not ($d_h = 0$). DMs then update their beliefs about the probability of a success using Bayes' rule

$$f_{h+1}(q|c_1,d_1,\ldots,c_n,d_n) = \frac{p(c_1,d_1,\ldots,c_h|q)f_1(q)}{\int p(c_1,d_1,\ldots,c_h,d_h|q)f_1(q)dq}. \quad (4)$$

The impact that the observed data have during the updating process depends on the DMs' uncertainty in their prior beliefs, δ_1. The free parameter δ_1 is the variance of the prior beta distribution. DMs with low uncertainty will have a low δ_1 that leads them to discount the observed data more than would DMs with a higher δ_1. Once the distributions describing respondents' beliefs in q are updated, their new means are used as the estimates of $\hat{q}_{h+1}$ for Round $h + 1$.

The mean ($\hat{q}_1$) and the variance (δ_1) of the prior distribution show how the learning process can either improve or impede the sequential risk-taking paradigm's clinical diagnosticity. On the one hand, real-world risk seekers and avoiders may have systematic differences in their prior beliefs (summarized by $\hat{q}_1$). Specifically, risk seekers could be more optimistic and believe the initial probability of a success to be higher (high $\hat{q}_1$) than their risk-avoiding counterparts. As a result they would choose to play for more trials in a given round, producing a correlation between the adjusted ART and risky behaviors.

On the other hand, the uncertainty across DMs in their beliefs (δ_1) may only hurt the diagnosticity of the task. In the model, δ_1 moderates the degree to which respondents' updated beliefs are a compromise between their observed data from past rounds and their initial beliefs. Higher levels of uncertainty (δ_1) mean that DMs will update their beliefs faster to reflect the observed data. And because the observed data are probabilistic, in the short run the beliefs of DMs will be more variable and thus reduce the reliability of correlations between performance and real-world behavior. Of course, this result is further compounded by differences in how far DMs continue to choose the play option for a given round.

Nonstationary mental representation. The stationary mental representation is surprising because both the BART and devil task have a nonstationary probabilistic structure where the a priori probability of a success, $s(i)$, decreases with each passing Choice Trial i,

$$s(i) = \frac{n - i}{n},$$

where n is the total possible number of choice trials. Examining performance in a larger class of task can address whether DMs use the same or different mental representation in different tasks.

This question of same or different mental representations can be examined with an alternative BSR model that allows for a correct nonstationary mental representation of the task (see also Wallsten et al., 2005 model number 7). The alternative model assumes that DMs adopt a correct mental representation but remain uncertain about the precise properties of the task. That is, during ill-defined conditions, DMs are uncertain of the maximum number of possible trials, n. A discretized gamma distribution over n, $p_1(n)$, describes their prior opinion about n's value for Round 1. The mean, $\hat{n}_1$, and variance, ν_1, of the gamma distribution are free parameters and carry the same psychological interpretation as their stationary counterparts.[7] The mean represents the best prior estimate of the maximum number of trials allowed on Round 1, $\hat{n}_1$, and scales how optimistic DMs are about the task. The variance, ν_1, indexes uncertainty and controls the impact that observed successes and failures have on the DMs' Bayesian updating process. In this mental representation, DMs learn and update their opinion of the likelihood of different values of n using Bayes' rule

$$p_{h+1}(n|c_1,d_1,\ldots,c_h,d_h) = \frac{(c_1,d_1,\ldots,c_h|n)p_1(n)}{\Sigma_n p(c_1,d_1,\ldots,c_h,\mathrm{d_h}|n')p_1(n')}. \quad (5)$$

.

A derivation of the revision process can be found in Appendix C of Wallsten et al. (2005). The expected value for the updated distribution after each round represents the new estimate of the maximum number of trials, $\hat{n}_{\mathrm{h}+1}$.

The remaining evaluation and response components are unchanged. DMs still identify the maximizing trial and probabilistically choose to play or stop based on the distance from that trial. However, the subjective probability in Equation 1 is set to $\pi_h(i) = (\hat{n}_h - i)/\hat{n}_h$, and the maximizing trial is

$$G_h = \frac{\hat{n}_h \gamma^+}{\gamma^+ + 1}, \quad (6)$$

where again $\hat{n}_h$ is either obtained from the DMs' estimated beliefs via the learning process in ill-defined tasks or directly observed in well-defined games. Comparing Equation 2 with Equation 6 also shows that γ^+ retains the same properties. In both cases, larger values of γ^+ lead to larger target values and consequently riskier play options taken in a given round.

Summary of Model and Predictions

In summary, the BSR model uses at most five parameters to predict the decision between playing and stopping during round h on choice trial i. It assumes DMs evaluate options prior to begin-

[6] The beta distribution was chosen due to its nice mathematical properties as a conjugate distribution to the binomial (see Gelman, Carlin, Stern, & Rubin, 2003). Typically the beta distribution is modeled with two free parameters, a_0 and b_0, where $a_0 \geq 0$ and $b_0 > 0$. The mean and variance are functions of these two parameters. In fact, the models were estimated with a_0 and b_0. For ease of interpretation, a_0 and b_0 are reparameterized into the mean and variance in the text.

[7] The gamma distribution is a continuous distribution that is sometimes specified by the parameters ν and τ, where $\mu_G = \nu\tau$ and $\sigma_G^2 = \nu\tau^2$ (see Gelman et al., 2003) . It was chosen because of its properties of being distributed over the positive reals and being flexible enough to characterize a wide range of different opinions. To obtain the discrete approximation to the gamma distribution, I integrated the distribution from $x = n - 0.5$ to $x = n + 0.5$ for each $n = 1, 2, \ldots, \infty$ and then normalized.

ning each round and identify a targeted choice trial that would maximize their expected gains. They then probabilistically choose to play or stop according to the distance from their target. Finally, during ill-defined games they use their observed data from past rounds to update their beliefs about the properties of the task. The model can account for both stationary and nonstationary mental representations.

There are two sets of predictions from the BSR that this article investigates. One set deals with the competing hypotheses regarding learning during ill-defined conditions. On the one hand, individual differences in prior beliefs captured with the parameter $\hat{p}_h$ or $\hat{n}_h$ may aid the paradigm's ability to identify real-world risk takers. On the other hand, DMs' general uncertainty seen with parameter δ_1 or ν_1 may do the opposite and hurt this ability. A second prediction is that DMs mentally model sequential risk-taking tasks as a stationary probabilistic structure even when the stochastic structure is truly nonstationary. Next, I develop a class of four sequential risk-taking tasks that I call the Angling Risk Tasks, which experimentally tests these predictions (see Figure 1).

The Angling Risk Tasks (ART)

The ART uses fishing tournaments akin to decision theory's prototypical "balls in the urn" laboratory task to aid participants' understanding of the stochastic environment. The premise of the ART lends itself well to manipulations of learning and the type of stochastic process. During a particular game, participants fish in a tournament for H rounds or trips (typically 30) in a pond that has 1 blue fish and $n - 1$ red fish. With each cast of a computerized fishing rod, they hook a fish (each fish is equally likely to be caught). If it is red, then they earn *5¢* and can cast again. But if it is blue, then the round ends and the money earned on that round is lost.

Different levels of learning can be manipulated by changing the weather conditions of the fishing tournament. For example, the tournament can take place on a sunny day, allowing participants to see how many fish are swimming in the pond and eliminating the need to learn their distribution. In contrast, the tournament can take place on a cloudy day, concealing the fish in the pond, which in turn forces participants to learn about how many potential fish are in the pond. If real-world risk takers do systematically differ in their prior beliefs ($\hat{q}_1$ or $\hat{n}_1$), then the ill-defined cloudy conditions should correlate equally well or even better than the well-defined sunny conditions.

Also, DMs' use of the same or different mental representations in different stochastic conditions can be tested by changing the pond's release law. Participants can catch and keep their fish (the catch 'n' keep tournament), a sampling-without-replacement process that is identical to the BART and devil task. Or participants can release their fish back into the pond (catch 'n' release tournament), a new task with a stationary sample-with-replacement process that matches the assumptions DMs hold about the BART.

In the experiment detailed next, respondents completed all four tournaments and also reported their past illegal drug use activity. The activity of drug use was selected because it has been one of the standard reported behaviors that the BART has been validated with (see for example Lejuez et al., 2002). In addition, risky drug use is an activity performed at different levels in and around college campuses, where this study took place (Johnston, O'Malley, Bachman, & Schulenberg, 2005).

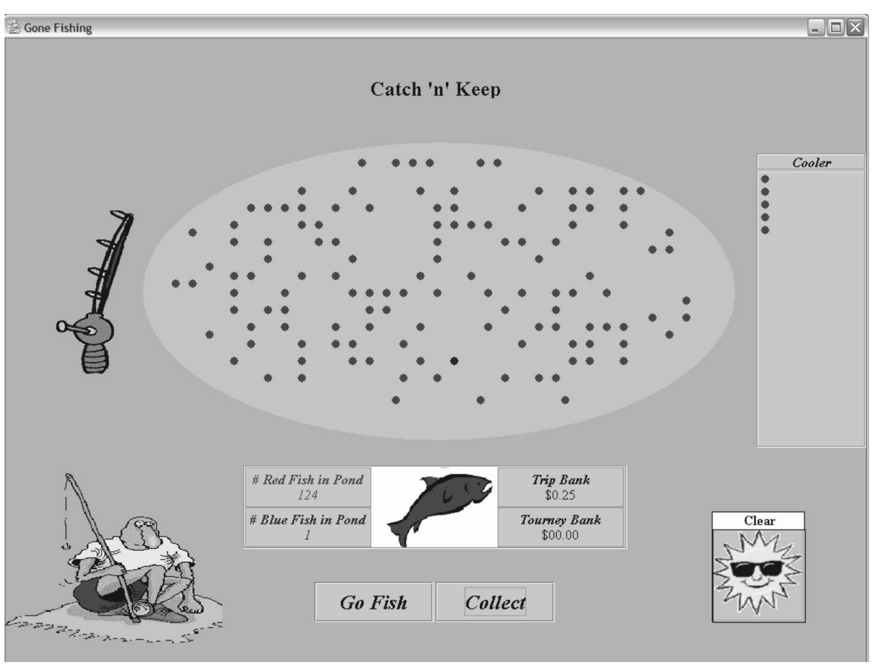

Figure 1. A screenshot of the Angling Risk Tasks' catch 'n' keep tournament on a sunny day. The screen changes for the three other tournaments. If the weather is cloudy, then the fish in the pond are not shown to the participants and no information about their number is given in the information panel. 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 with advertisements placed throughout the campus. The sample consisted of 38 men and 34 women, ranging in age from 18 to 34 years ($M = 21.6$, $SD = 3.9$). Fifty-six percent described themselves as White, 18% as Black or African American, 17% as Asian or Southeast Asian, and 4% as Hispanic or Latino; the remaining 6% marked “Other” or chose not to respond to the question. They were paid $7 for their time. In addition, participants could earn a bonus based on their winnings in the games. On average, people received an extra $5, but a few won nothing and one individual earned an extra $12.

Materials

The ART. In each of the four tournaments participants had $H = 30$ rounds during which they fished in a pond similar to the one shown in Figure 1. At the beginning of each round 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. If a red fish was caught, then *5¢* was placed into the “Trip Bank” shown on the information panel. What happened next depended 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 total number of red fish in the pond by one. If the law was catch ‘n’ release, then the red fish was placed back into the pond. Either way, the computerized fish swam around the pond and participants had another opportunity to cast the line into the pond for that round. However, if a blue fish was caught, then the round ended and participants lost their accumulated money in the “Trip Bank.” When participants decided to stop fishing, they pressed the “Collect” button to transfer the money to the “Tournament Bank” on the information panel and then began a new round.

In addition to the two release laws, there were two types of weather conditions. If the weather was sunny—as indicated by the weather forecast in the bottom right—the pond was clear and participants could see how many fish were in it at all times. Additionally, 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 left blank. Combining the two release laws with the two weather forecasts produced four different fishing tournaments (conditions). Participants completed all four tournaments.

Drug and alcohol questionnaire. As a measure of risky illegal/legal drug use, participants reported with yes/no responses the number of drug categories ever tried. The 11 different drug categories were cannabis, alcohol, cocaine, MDM (ecstasy), stimulants (e.g., speed), sedatives/hypnotics, opiates, hallucinogens, PCP, inhalants, and nicotine. The sum of the number of categories tried (polydrug) is a validated measure of risky drug use (see Babor et al., 1992; Grant, Contoreggi, & London, 2000). After each yes/no question, participants were also asked to report about how often they used the drug at the time in their life when they used it the most frequently (i.e., *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*). Weighting the categories that participants reported trying with the frequency rank and summing created the measure identified as *weighted polydrug*.

Domain-Specific Risk-Taking (DOSPERT) Scale. Participants also completed the DOSPERT Scale (Weber et al., 2002), which contains 40 items that assess the likelihood of engaging in risky behavior in six domains: ethics, investment, gambling, health/safety, recreational, and social. Two separate variants of the scale also assess the perception of the magnitude of the risk for and expected benefit from each of the 40 risks. The DOSPERT has also had success in identifying real-world risk takers (Hanoch, Johnson, & Wilke, 2006).

Design and Procedure

The study used a 2 (release law) × 2 (weather) within-subjects design. Participants fished in all four tournaments (conditions) and completed the drug and alcohol questionnaire along with the DOSPERT Scale. Participants had 30 rounds per tournament to cast for as many red fish as they chose, earning 5¢ per red fish caught. Intertask comparisons were facilitated by setting the ART parameter settings to be similar to those used in the BART. In the BART (a nonstationary environment), the total number of possible choice trials (pumps) is typically set at $n = 128$ (Lejuez et al., 2002). If the goal of participants is to maximize expected value and they have full knowledge of the structure of the task, then the strategy that maximizes earnings is for participants to choose the play option on approximately 64 of the choice trials. This setting has been found to be the best for distinguishing between real-world risk seekers and risk avoiders (Lejuez et al., 2002). Consequently, in the catch ‘n’ keep conditions there were $n = 128$ total number of fish or possible choice trials. Calibrating the catch ‘n’ release conditions in the ART so that choosing play 64 times per round was the maximizing strategy (assuming the same conditions as above) was done by setting the number of fish at $n = 65$ fish (see Equation 2 with $\gamma^{+} = 1$). Thus, if participants had a precise understanding of all four tasks and sought to maximize expected value, they should make an equal number of casts in all four tournaments.

The order in which participants completed each tournament and the risk questionnaires was counterbalanced. All eight tasks were programmed using Sun Microsystem’s Java language and are available upon request. The entire experiment was administered on PC computers in separate sound-attenuated laboratory cubicles.

After reading and signing the informed consent form, participants read an introductory set of instructions on the computer. They were told that they would be playing four different fishing tournaments with different rules and conditions. The instructions described the two different release laws and the two different weather conditions they would experience. Participants were also informed that between each fishing tournament they would fill out questionnaires assessing their own risky behavior. No further instructions were given about the stochastic structure of the tasks.

Next, the participants completed four practice games, one for each tournament. During the practice games participants were able to select how many fish they would like in the pond to practice with (1 to 360). This manipulation was done to demonstrate that the ponds in each of the experimental conditions could have any number of fish so as to minimize knowledge transfer from one tournament to the next. During the practice round participants played each condition for two rounds, during which they cast for red fish as many times as they chose to.

After completing the four practice games, they began the experimental session, alternating between the four questionnaires and the four tournaments. They started with a questionnaire. Before each tournament, participants were briefly reminded of the rules governing the pond they were about to visit. After completing all of the tournaments, they completed a set of questions regarding the strategy they used to fish in the tournaments. At the conclusion of the session, the computer produced four tables showing how much money participants earned on each round during the four tournaments. A round from each tournament was then chosen randomly using a bingo basket (four rounds total), and participants were paid based on their performance during these rounds.

Results

The Results section is organized into two subsections. The first addresses the cognitive processes used during the four different conditions of the ART (catch ‘n’ keep, sunny; catch ‘n’ keep, cloudy; catch ‘n’ release, sunny; and catch ‘n’ release, cloudy). The BSR model is used in this section to examine whether participants use the same mental representation in all four conditions of the ART, or whether they change their mental representation according to the stochastic structure of the task. Additionally, a set of BSR models with a probability weighting function is compared to the models without to examine whether DMs overweight rare events and underweight common events in sequential risk-taking tasks. The second section examines individual differences in the cognitive processes and their association to the risky use of drugs. One participant did not complete the experimental session and is not included in subsequent analyses.

Cognitive Processes During the ART

Both the average adjusted ART scores and proportion of rounds ending in a blue fish are listed in Table 1. The adjusted ART score is the average number of casts taken on rounds when participants chose to stop fishing. The adjusted score is the standard behavioral measure of performance in sequential risk-taking tasks (Lejuez et al., 2002; Pleskac, Wallsten, Wang, & Lejuez, 2007). Hence, it is used in all subsequent analyses at the behavioral level.

Table 1
Mean (SD) Adjusted ART Scores and Percentage of Rounds Ending With a Participant Catching a Blue Fish

	Adjusted ART scores		Percentage of rounds ending with a blue fish	
Release law	Sunny day	Cloudy day	Sunny day	Cloudy day
Catch ‘n’ keep	38.96 (18.77)	31.59 (15.35)	.34 (.16)	.25 (.15)
Catch ‘n’ release	32.19 (15.95)	28.06 (15.22)	.38 (.14)	.35 (.14)

Note. Calculations are based on a total of 30 rounds. ART = Angling Risk Tasks.

To analyze the effect of the weather conditions and release conditions, I conducted an ANOVA on the adjusted ART scores where each experimental condition was a within-groups variable. To test for order effects, two other between-groups variables were entered: (a) the order of completing sunny and cloudy tournaments; and (b) the order of completing the release conditions. The ANOVA utilized the full model, allowing interactions between the individual participant and the variables of interest. The appropriate error terms are therefore the interactions between the participant and the variable of interest, MSPar × Variable (see Howell, 1997).

The left panel in Figure 2 shows the average adjusted ART score in the four experimental conditions.[8] Participants had a tendency to cast more frequently in the catch ‘n’ keep condition than during catch ‘n’ release, $F(1, 67) = 17.70$, $p < .01$, MSPar × Release Rule = 103.84. They also made more casts in the sunny condition than in the cloudy condition, $F(1, 67) = 18.18$, $p < .01$, MSPar × Weather Condition = 121.55. Finally, there was a significant interaction between release law and weather, $F(1, 67) = 6.01$, $p < .05$, MSPar × Release Rule × Weather Condition = 35.76. As Figure 2 indicates, the change from sunny to cloudy had a larger effect in the catch ‘n’ keep condition than in catch ‘n’ release, $t(67) = 2.46$, $p < .05$.

There were no main effects related to the order in which participants completed the four fishing conditions. Participants who completed sunny tournaments before cloudy tournaments did not have significantly different adjusted ART scores. Nor did the order in which they completed the different release laws affect their casting behavior. Despite all efforts to minimize information transformation from the sunny to cloudy condition, there were two significant interactions with the order of completing the weather conditions. The order of completing the different release laws interacted with the weather variable, $F(1, 67) = 6.00$, $p < .05$, MSPar × Weather Condition = 121.55. Paired comparisons revealed participants did cast significantly more in sunny conditions ($M = 36.7$) than in cloudy ($M = 27.9$) when completing catch ‘n’ release conditions first, $t(67) = 4.8$, $p < .05$. But when the catch ‘n’ keep conditions were completed before catch ‘n’ release conditions, there was not a significant difference between casting behavior between sunny ($M = 34.2$) and cloudy ($M = 31.8$) conditions, $t(67) = 1.3$. The order of completing the weather conditions also interacted with the weather manipulation, $F(1, 67) = 7.55$, $p < .05$, MSPar × Weather Condition = 121.55. The adjusted ART scores for the cloudy conditions were significantly greater when participants completed sunny conditions first ($M = 33.4$) as compared with cloudy conditions first ($M = 26.4$), $t(67) = 3.8$, $p < .05$. But there was not a significant difference between adjusted ART scores on the sunny conditions based on the order in which they completed the weather conditions ($M = 35.4$ and 35.5, respectively). The only other significant order effect was the four-way interaction, but it was due to the two significant two-way interactions. Neither significant interaction pertaining to the order in which participants completed the four conditions influenced any

[8] The mean and median number of casts made on rounds that participants stopped on is approximately equal. All conclusions drawn from analyses based on the mean, including all the ANOVAs and all the correlations, are identical to those done with the median. Consequently, only the means will be reported in the text.

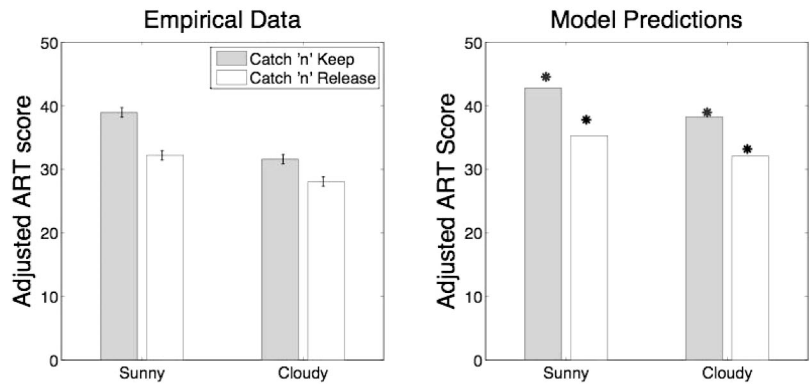

Figure 2. Average adjusted Angling Risk Tasks (ART) scores across participants for the four different fishing tournaments. In the left panel, the bars represent the average adjusted ART scores and the vertical lines denote standard errors of the mean, estimated from the MSPar × Release Rule × Weather Condition interaction. The right panel plots the predicted average adjusted ART score calculated from the 71 sets of individual parameter estimates using the Bayesian sequential risk-taking model with correct mental representations (see Appendix B for the analytic solutions for predicting the adjusted ART scores from the cognitive models). Stars indicate the average number of casts that the decision makers intended to take on all rounds (see Appendix B).

of the conclusions drawn from all subsequent analyses and will not be considered further.

A final observation concerning the adjusted ART score is that choice behavior in the ART conditions appears relatively consistent with that of the BART. The average adjusted ART scores for all four conditions were near or slightly greater than the adjusted score typically observed with the BART ($M = 30.3$; $N = 448$; $SE = 2.2$; 95% confidence interval = $26.0 < \text{score} < 34.6$; see Aklin, Lejuez, Zvolensky, Kahler, & Gwadz, 2005; Crowley, Raymond, Mikulich-Gilbertson, Thompson, & Lejuez, 2006; Jones & Lejuez, 2005; Lejuez, Aklin, Bornovalova, & Moolchan, 2005; Lejuez, Aklin, Jones, et al., 2003; Lejuez, Aklin, Zvolensky, & Pedulla, 2003; Lejuez et al., 2002). Next I use the two versions of the BSR model to test whether respondents use the same mental representation in all four conditions or whether they adapt their mental representation to the different conditions.

BSR Analyses

Two BSR models were fit to the data from all four ART conditions at the individual level. Both predict the probability of casting, $r_{h,i}$, during round h on trial i. The first model—the best fitting model for the BART—tests the hypothesis that participants believe a stationary stochastic process governs the task. For cloudy conditions it has one evaluation parameter, two response parameters, and two learning parameters. In sunny conditions there is no learning and $\hat{q}_h$ in Equation 2 is set to the observed parameters of the task (e.g., 64/65 in catch 'n' release). The second model tests whether participants mentally represent the structure as nonstationary. It has the same allocation of free parameters, and in sunny conditions $\hat{n}_h$ is set to the observed number of fish (e.g., 128 in catch 'n' keep). In addition, a baseline statistical model was estimated from the data (see Wallsten et al., 2005, for details). It has one parameter, the probability of casting on trial i. It assumes no cognitive processes and serves as an initial test for the BSR models.

Model estimation and comparison. The models were fit to each individual's choice data for all rounds from each fishing tournament using maximum likelihood estimation methods described in Appendix A. The data have a substantial number of observations per participant, but the precise number of independent observations depends on the intercorrelation between trials (captured within the model) and differs for each individual. As a rough estimate, the number of independent observations for each tournament ranged between 30 rounds and the total number of trials on which participants actually chose between playing and stopping, which was on average 840.4 trials per person per tournament.

Standard maximum likelihood ratio tests are not available to evaluate and compare model fits due to the nonnested nature of the models. Instead descriptive-level comparisons were made with the Bayesian information criterion (BIC; Schwarz, 1978; Wasserman, 2000). The BIC is one method to compare the fit of nonnested models with different numbers of parameters, where

$$\text{BIC} = [-2 \times ML] + [k \times \log(N)]. \quad (7)$$

In the expression, ML is the maximum log-likelihood of the data given the model, k is the number of parameters in the model, and N is the number of independent observations that contributes to the likelihood. The term on the right-hand side of Equation 7 accounts for model complexity (i.e., the more parameters the more complex the model). The model with the lowest BIC is chosen as the best fitting model. As a very liberal estimate of N, I used the total number of choices a participant made in a given tournament. This estimate puts the greatest handicap on models with more parameters.

Table 2 shows how many participants were best fit by each model in the two sunny tournaments using BIC. Recall no learning process parameters for the BSR models were fit in these conditions, and the number of free parameters reflects this. The model fits in the sunny tournaments also reveal that the models with congruent representations fit the data better than does the statistical baseline model.

However, as is apparent in Table 2, in the sunny tournaments the BSR model that assumes participants have an incorrect mental representation of the task could not fit the data. Here is why. Recall that in the well-defined sunny tournaments participants knew how many fish were in the pond. Consequently, in the catch 'n' release tournaments the model assuming a nonstationary representation would set the value of $\hat{n}_h$ to 65 in Equation 6 for all 30 rounds (the number of fish in the pond at the beginning of the round). With this setting, the model does not predict that DMs would make more than 65 casts on a given round. That is because in their sample-without-replacement model they would have caught all 65 fish in the pond, leaving none in the pond. Yet, a third of the participants in the sunny catch 'n' release tournaments cast more than 65 times for at least one round. For the remaining participants the evaluation parameter would have to reflect extreme sensitivity to payoffs ($\gamma^+ > 1.5$ in Equation 6) to capture their data—an estimate that is highly unlikely and inconsistent with decision theory literature (e.g., Tversky & Kahneman, 1992).

The same issue arises when fitting the model that posits a stationary representation to the data from the sunny catch 'n' keep condition. Here $\hat{q}_h$ would be set to 127/128 in Equation 2. Using this value, all participants would appear incredibly insensitive to outcomes ($\gamma^+ < 0.3$)—an estimate that is quite incongruent with the literature. Thus, these results paired with the superior performance of the BSR model over the baseline model offer preliminary support that in nonlearning conditions DMs can correctly adapt to and represent the probabilistic structure of the sequential task they face. This result is in contrast to the hypothesis derived from the BART that DMs use the same representation in both conditions.

The results from the cloudy tournaments support the same conclusion of correct mental representations. Table 3 shows that both BSR models fit the data better than do the baseline statistical models in the catch 'n' keep and catch 'n' release conditions. Moreover, the models that assume participants correctly represent and learn about the task structure best fit a majority of participants in both conditions. That is, the stationary model fits best in catch 'n' release, and the nonstationary fits best in catch 'n' keep. These results are in contrast with the BART results, where the model assuming a stationary representation fit the nonstationary task best (Bishara et al., 2007; Wallsten et al., 2005).[9]

Table 2
Model Comparisons of the Sunny Tournaments

Model	*df*	Catch 'n' keep: Mean BIC	Catch 'n' keep: Number of DMs best fit with BIC	Catch 'n' release: Mean BIC	Catch 'n' release: Number of DMs best fit with BIC
Baseline	1	192.16	4	176.15	1
Stationary mental representation	3			145.53	70
Nonstationary mental representation	3	156.86	67		

Note. During sunny conditions, only the Bayesian sequential risk-taking models that assumed the correct stochastic process could be fit to the respective conditions. BIC = Bayesian information criterion; DM = decision maker.

Table 3
Model Comparisons of the Cloudy Tournaments

Model	*df*	Catch 'n' keep: Mean BIC	Catch 'n' keep: Number of DMs best fit with BIC	Catch 'n' release: Mean BIC	Catch 'n' release: Number of DMs best fit with BIC
Baseline	1	206.00	4	176.34	3
Stationary mental representation	5	179.35	3	144.83	49
Nonstationary mental representation	5	171.51	65	146.34	19

Note. BIC = Bayesian information criterion; DM = decision maker.

Finally, the best-fitting BSR models also reproduce the adjusted ART scores well. The bar graph in the right panel of Figure 2 plots the predicted adjusted ART scores for the four conditions averaging across all participants and all rounds. The predicted adjusted ART scores were calculated for each individual and for each round using their respective parameter estimates from each condition. The analytical solutions for the predictions are developed in Appendix B.

Comparing the predicted and empirical scores does show a slight miscalibration in the models. For all four conditions, the predicted adjusted ART scores are about five casts higher than in the empirical data. This is likely due to the fact that the empirically estimated adjusted ART scores are calculated from only the rounds that the participant ended with a stop choice. The models, in contrast, are estimated from all the rounds. To understand why this is important, recall first that in the ART the more that participants cast the more likely the round is to end in a blue fish. Thus, the rounds that end with a blue fish are also the rounds in which participants are more likely to exhibit the greatest level of risk seeking. The models—being fit to all the rounds—reveal this aspect of increased risk taking in their parameters, and as a result the predicted adjusted ART scores (even when conditionalized on the stop rounds) will reflect a slightly more risk-seeking nature than in the empirical scores. The models in fact even permit us to estimate the intended casting behavior of participants for all rounds (see Appendix B for a solution). These predictions are plotted as stars in the right panel of Figure 2. They show that if casting behavior were observable on all rounds, then the scores would be even slightly higher.[10] Further implications of these results are addressed in the discussion. Next, I examine whether the models should be expanded to include a probability weighting function.

Probability weighting function. One final untested hypothesis is whether during the evaluation process rare events are overweighted and common events underweighted. A probability

[9] Note Wallsten et al. (2005) used Akaike's information criterion (AIC) as a goodness of fit measure. BIC and AIC give rise to the same conclusions in this dataset.

[10] Appendix B also uses the analytic solutions for the adjusted ART scores to show an error theory is not sufficient alone to account for the observed data.

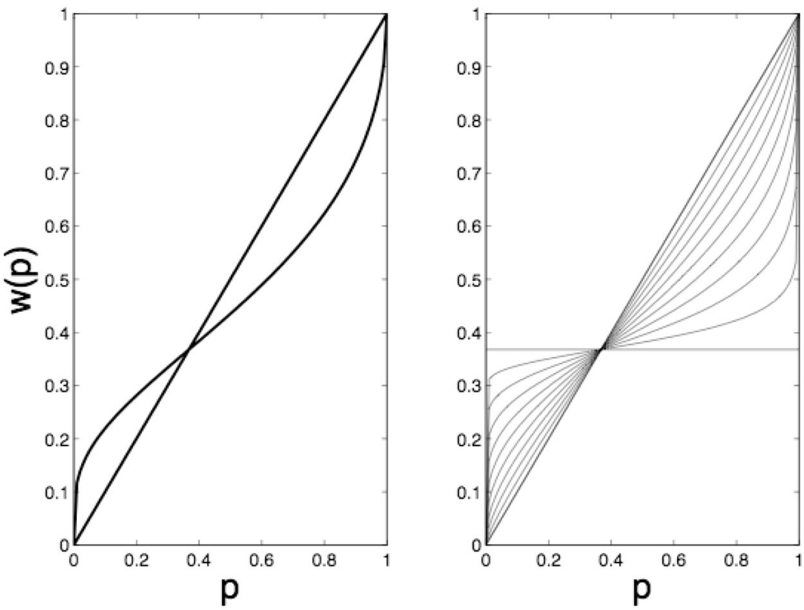

Figure 3. Plots and properties of the weighting function. The left panel shows a weighting function exhibiting overweighting of low probabilities and underweighting of high probabilities. The right panel varies α between 0 and 1; as α decreases, the weighting function approaches a step function.

weighting function with these properties is displayed in the left panel in Figure 3. These properties have been found to be a necessary component in explaining preferences over static gambles (see Fox & Tversky, 1998; Gonzalez & Wu, 1999; Kahneman & Tversky, 1979; Luce, 2000; Prelec, 1998; Tversky & Kahneman, 1992; Wakker & Tversky, 1996; Wu & Gonzalez, 1996, 1999). A probability-weighting function may be necessary in sequential risk-taking tasks as well. This study presents a unique opportunity to investigate this hypothesis because participants used two different mental representations in two different tasks. This occurrence makes it possible to simultaneously estimate a probability weighting function and a value function (like the one used in Equation 1) due to the fact that participants completed more than one sequential risk-taking task and used different mental representations (see Pleskac, 2004). If these conditions are not met—as in Wallsten et al. (2005)—then the two functions are not simultaneously identifiable.

To expand the BSR model to include a weighting function, I used Prelec's (1998) one-parameter function[11]: $w(t) = \exp(-[-\log(t)]^{\alpha})$, where $0 \leq \alpha \leq 1$ and t is the probability of an event. For $\alpha = 1$ the weighting function is an identity function, illustrating Wallsten et al.'s (2005) assumptions. The parameter α measures the degree of over/underweighting of probabilities. As α decreases there is more over/underweighting (see the right panel of Figure 3). In Equation 1, $\pi_h(i)$ is now $w[\pi_h(i)]$. Consequently, the maximizing trial for either mental representation is a function of α as well. If we revise Equation 2, the closed form solution for the maximizing trial assuming a stationary mental representation is

$$G_h = \frac{-\left(\dfrac{\gamma^+}{\alpha}\right)^{1/\alpha}}{\ln(q_h)}. \qquad (8)$$

There is not a closed form solution for the maximizing trial in the nonstationary mental representation (see Equation 6 for the expression without a weighting function), but it can be estimated numerically. The weighting function has similar implications for both mental representations. With γ^+ held constant, as the degree of over/underweighting of probabilities increases (α decreases), the maximizing trial increases and DMs are predicted to make more and more casts. In comparison, if the weighting function is held constant, as individuals exhibit less and less diminishing sensitivity to changes in payoffs (γ^+ decreases), then the maximizing trial decreases.

Equation 8 reveals that for a given condition of ART, α and γ^+ cannot be simultaneously estimated. The same holds for the nonstationary model. However, the parameters can be estimated if they are set equal across tasks in which participants use different mental representations (see Pleskac, 2004). In this experiment we have two conditions in which participants used a stationary representation (the catch 'n' release conditions) and two with a nonstationary representation (the catch 'n' keep conditions; see Tables 2 and 3). The most reasonable constraint for model estimation is to set γ^+ and α equal across catch 'n' keep and catch 'n' release for a given weather condition. Using this constraint the BSR models with a weighting function were fit at the individual level for all four conditions using maximum likelihood procedures. The reader is referred to Appendix C for a detailed explanation of the estimation and model comparison.

Briefly, the results of the model comparison showed that a weighting function made a substantial contribution in describing the data only when DMs were constrained to show only diminishing sensitivity ($0 < \gamma^+ \leq 1$). If, however, $\gamma^+$was allowed to take

[11] The two-parameter version is not identifiable in the present models.

Table 4

Summary Statistics of and Correlation With MLE Parameters for the Cloudy Tournaments, Assuming Mental Representations Were Congruent With the Structure of the Task

			Catch 'n' keep				Catch 'n' release			
	Evaluation		Learning		Response		Learning		Response	
Statistics	γ^+	α	$\hat{n}_1$	ν_1	z	β	$\hat{q}_1$	δ_1	z	β
Mean	0.59	0.75	145.3	9.74×10^6	0.01	0.3	0.92	1.73×10^{-3}	−0.01	1.26
SD	0.25	0.20	141.5	8.20×10^7	0.05	0.58	0.21	8.11×10^{-3}	0.08	8.18
1st quartile	0.43	0.61	50.78	83.08	−0.01	0.09	0.96	1.41×10^{-5}	−0.05	0.14
Median	0.59	0.79	96.79	645.22	0.01	0.15	0.98	6.50×10^{-5}	0.00	0.17
3rd quartile	0.84	0.98	183.65	3,867.18	0.05	0.28	0.99	3.31×10^{-4}	0.04	0.31
Adjusted ART score (KC)	.26**	−.07	.30**	.22	.08	−.34**	.20	−.09	.10	−.37**
Adjusted ART score (RC)	.24**	−.10	.13	.11	−.02	−.16	.14	−.11	.05	−.52**
Health Attitude score	.01	−.18	−.01	.01	.05	−.01	−.17	−.01	.16	−.19
Social Attitude score	−.07	−.22	.07	.04	.01	.05	−.01	−.18	.08	.03
Weighted polydrug	.24*	.01	−.09	.43x	−.10	−.15	−.37*	.01	−.04	−.10

Note. The evaluation parameters ($\gamma+$ and α) are set constant across the release conditions for a given weather condition. The Pearson *r* correlations between the MLE parameters and the adjusted ART scores, the DOSPERT scores from the Health Attitude and Social Attitude domains, and the self-reported weighted polydrug use are in the bottom five rows. MLE = maximum-likelihood estimates; ART = Angling Risk Tasks; KC = catch 'n' keep (cloudy day); RC = catch 'n' release (cloudy day); DOSPERT = Domain-Specific Risk-Taking Scale.
* $p < .05$. ** $p < .01$. x = not a meaningful correlation (see Footnotes 11 and 12).

any value between 0 and 3—meaning DMs could have diminishing, equal, or increasing sensitivity—then the original set of models taking probabilities at face value without a weighting function fit a majority of participants best.

Next, the models with the weighting function and the diminishing sensitivity constraint are used to assess individual differences in the cognitive processes. Though perhaps not the best fitting models—but still adequate (see Appendix C)—the models with a weighting function make it possible to test whether risk takers differ in how they evaluate payoffs, weight probabilities, or both. Behavioral decision theory would predict that risk takers might differ on both dimensions (Tversky & Fox, 1995; Tversky & Kahneman, 1992). Tables 4 and 5 summarize the maximum likelihood estimates (MLE) of the parameter values for models with the weighting function in cloudy and sunny tournaments, respectively.[12,13]

Individual Differences in Cognitive Processes

The relationship between performance during sequential risk-taking tasks and real-world risky behavior makes it possible to examine individual differences in the cognitive processes used during the tasks. In particular, the relationship between the four adjusted ART scores and risky behaviors can reveal whether learning helps or hurts the clinical diagnosticity of sequential risk-taking tasks. On average, participants reported using 2 of the 11 drug categories. The average score on weighted polydrug was 10.0 (*SD* = 10.5; *Mdn* = 7.0). The scale had good internal reliability with a coefficient alpha value of .89. Three participants did not complete the drug questionnaire, so they are not included in this set of analyses.

Table 6 lists the Pearson correlations between the adjusted ART scores and weighted polydrug.[14] The table shows that in the sunny conditions (when participants saw and knew how many fish were in the pond) there was a significant correlation between the adjusted ART score and weighted polydrug. However, in support of the prediction that the learning process hinders the diagnosticity of the ART, the adjusted ART scores in the cloudy tournaments were not significantly correlated with the self-reported risky use of drugs. An a priori planned comparison showed that the correlations in the sunny conditions were significantly greater than in the cloudy, $z = 3.10$, $p < .05$ (see Meng, Rosenthal, & Rubin, 1992). That is, only in well-defined versions of the ART that do not require a learning process do the adjusted ART scores identify the risky use of drugs. In fact, Table 7 shows that behavior was consistent enough across the sunny conditions that the adjusted ART scores calculated on the first, second, and last set of 10 rounds were also significantly correlated with weighted polydrug.

[12] The bias component significantly improved the fit of the models. The bias-free nonstationary representation models provided a significantly worse fit to the data in the catch 'n' keep conditions, $G^2(71) = 915.87$, $p < .01$ and $G^2(71) = 105.54$, $p < .01$, respectively. A similar result was found with the stationary representation in catch 'n' release, $G^2(71) = 436.85$, $p < .01$ and $G^2(71) = 168.34$, $p < .01$. All conclusions regarding the mental representation of respondents remain if the data are analyzed without the bias component.

[13] The mean MLE of α in Tables 4 and 5 indicate that on average individuals were less sensitive to changes in probability in the sunny tournaments than in the cloudy, $t(70) = 3.80$, $p < .01$. The remaining possible hypothesis tests regarding changes between conditions of cognitive processes with both least-squares methods and nested model comparisons did not identify any significant or consistent trends across the tournaments.

[14] All correlations were run with both Goodman and Kruskal's nonparametric measure of association (*K*) and Pearson's *r*. All conclusions drawn from the association with weighted polydrug remain, whether one uses a Pearson's *r* correlation or *K*. One minor inconsistency is that response sensitivity in the sunny catch 'n' keep condition was negatively associated only with the nonparametric measure, $K = -.29$, $p < .01$. Given that none of the other conditions showed this association, it will not be interpreted. Pearson's *r* correlations will be reported in the text to support the interpretation of the linear regression analyses.

Table 5
Summary Statistics of and Correlation With MLE Parameters for the Sunny Tournaments, Assuming Mental Representations Were Congruent With the Structure of the Task

Statistics	Evaluation		Catch 'n' keep response		Catch 'n' release response	
	γ^+	α	Z	β	z	β
Mean	0.68	0.59	−0.01	0.73	−0.01	0.15
SD	0.33	0.28	0.05	3.31	0.06	0.21
1st quartile	0.41	0.45	−0.06	0.1	−0.04	0.03
Median	0.87	0.55	−0.01	0.14	0.00	0.09
3rd quartile	0.99	0.89	0.06	0.19	0.02	0.18
Adjusted ART score (KS)	.74**	−.07	−.16	.00	−.15	−.46**
Adjusted ART score (RS)	.54**	−.16	−.08	−.07	−.07	−.40**
Health Attitude score	.24*	−.13	.17	.19	.19	−.30*
Social Attitude score	.00	−.10	.18	−.04	−.04	−.09
Weighted Polydrug score	.34**	−.08	.02	−.15	.16	−.22

Note. The evaluation parameters ($\gamma+$ and α) are set constant across the release conditions for a given weather condition. The Pearson *r* correlations with the DOSPERT risk attitude factor and self-reported drug use are in the bottom five rows. MLE = maximum-likelihood estimates; ART = Angling Risk Tasks; KS = catch 'n' keep (sunny day); RS = catch 'n' release (sunny day); DOSPERT = Domain-Specific Risk-Taking Scale.
* $p < .05$. ** $p < .01$.

In comparison, for the cloudy conditions, none of the adjusted ART scores shown in Table 7 were statistically significant.

To identify the cognitive processes responsible for the tasks' reliable identification of the risky use of drugs, the MLE parameters for each individual from the four conditions were correlated to weighted polydrug. The correlations with each individual parameter value are listed in the bottom rows of Tables 4 and 5. In both cloudy and sunny tournaments the payoff sensitivity parameter, γ^+, was significantly correlated to past drug use, not the probability weighting parameter, α. The significant correlations with the learning process parameters, $\hat{q}_1$ and υ_1, are due to a few extreme parameter values and do not remain with nonparametric associations.[15]

The significant correlation between weighted polydrug and γ^+ in the cloudy conditions (see Table 4) is particularly noteworthy. Recall that the adjusted ART scores were not correlated with weighted polydrug in these cloudy conditions (see Table 6). This highlights one merit of the models; more specifically, once the effect of different experiences on learning was controlled for within the BSR models, systematic differences in real-world risk takers were identified. Further discussion of the implications of this finding in terms of both the learning and evaluation processes is left for the discussion.

To establish whether the differences in how people evaluate payoffs (γ^+) during the ART were responsible for the correlation between the adjusted ART scores and risky drug use, I performed a series of linear regressions. This was done by first forming a Sunny Tournament factor for each participant from the two sunny adjusted ART scores by normalizing the two scores and averaging them together. The Sunny Tournament factor is correlated both with the MLE γ^+ from the sunny tournaments ($r = .70, p < .05$) and with weighted polydrug ($r = .37, p < .05$). Next the estimates for γ^+ in the sunny tournament and the Sunny Tournament factor were simultaneously entered in a linear regression predicting weighted polydrug. The linear regression model with both γ^+ and the Sunny Tournament factor was significant, adjusted $R^2 = .13$, $F(2, 65) = 5.77, p < .01$. But the standardized regression coefficient for γ^+ was not significant, $\beta = 0.17$, $sr^2 = .01$, $t(65) = 1.07$, nor was the Sunny Tournament factor, $\beta = 0.26$, $sr^2 = .03$, $t(65) = 1.64$. This pattern implies that the sensitivity that people have to payoffs is responsible for the individual differences of real-world risk takers observed in the ART.

In a final set of analyses, correlations between weighted polydrug and Weber et al.'s (2002) DOSPERT scale were computed. Recall the DOSPERT scale identifies a person's risk attitude toward six different domains of risk-taking behavior. The DOSPERT scores reveal that participants who reported greater drug use also tended to report that they were more likely to engage in health and social risks in the future (see Table 6 for the health and social domains). None of the other four subscales were significantly correlated with the risky use of drugs.[16] The DOSPERT score for each of the six risky domains were also not significantly correlated with the adjusted ART scores in the four fishing conditions (again, see Table 6 for the health and social domains).

This pattern of correlations suggests that both the DOSPERT and the sunny ART conditions account for unique variance in the risky use of drugs. As a test of this prediction, responses to the social and health subscales were normalized and averaged to form a Risk Attitudes factor. The Risk Attitudes factor is significantly correlated with weighted polydrug ($r = .34, p < .05$) but not with the Sunny Tournament factor ($r = .07$). Next a simultaneous linear regression was run using both the Risk Attitudes factor and the Sunny Tournament factor to predict weighted polydrug. Together they significantly accounted for 24% of the variance in weighted polydrug, $F(2, 65) = 10.03$, $p < .01$, $MSE = 86.54$. Individually both the Risk Attitudes factor, $\beta = .31$, $sr^2 = .10$, $t(65) = 2.88$, $p < .05$, and the Sunny

[15] This was revealed by a nonsignificant rank-ordinal *K* correlation (see footnote 11) for both of these parameters.

[16] Besides assessing the likelihood of engaging in risky behavior in each of the six domains, I asked respondents to rate the perceived benefits and perceived risks associated with each activity in the six domains. In all six domains the perceived benefits were significantly positively correlated with the likelihood ($\bar{r} = .60$) and perceived risks were significantly negatively correlated with the likelihood ($\bar{r} = -.52$).

Table 6
Intercorrelations Between ART Scores for Each of the Four Tournaments, Responses to the DOSPERT Health Attitude and Social Attitude Subscales, and Self-Reported Drug Use

Variable	Mean (*SD*)	Weighted polydrug	1	2	3	4	5
1. Adjusted ART score (KS)	38.96 (18.77)	.36**	—				
2. Adjusted ART score (KC)	31.59 (15.35)	.14	.69**	—			
3. Adjusted ART score (RS)	32.19 (15.95)	.32**	.75**	.57**	—		
4. Adjusted ART score (RC)	28.06 (15.22)	.11	.51**	.72**	.64**	—	
5. Health Attitude score	2.58 (0.64)	.29*	.04	.08	.21	.12	—
6. Social Attitude score	3.76 (0.60)	.27*	−.09	−.01	.10	.02	.28*

Note. A total of 71 participants completed the four ART tournaments and the Health Attitude and Social Attitude subscales, and 68 participants reported their past drug use. ART = Angling Risk Tasks; DOSPERT = Domain-Specific Risk-Taking Scale; KS = catch 'n' keep (sunny day); KC = catch 'n' keep (cloudy day); RS = catch 'n' release (sunny day); RC = catch 'n' release (cloudy day).
$^{*}p < .05$. $^{**}p < .01$.

Tournament factor, $\beta = .33$, $sr^2 = .11$, $t(65) = 3.12$, $p < .05$, accounted for unique variance in weighted polydrug.

To identify whether the DOSPERT accounted for significant incremental variance in weighted polydrug above participants' payoff sensitivity (γ^+) and the Sunny Tournament factor, I conducted a hierarchical linear regression. Table 8 shows that, consistent with the previous analysis, entering the Sunny Tournament factor after γ^+ does not produce a significant change in R^2. However, Step 3 of the regression reveals that Risk Attitudes accounts for unique variance in weighted polydrug over and above the Sunny Tournament factor and the cognitive parameter γ^+. This implies that the DOSPERT and sequential risk-taking tasks account for different psychological aspects in real-world risk taking. In the discussion, I will address the implication of this result as well as the more specific results dealing with the cognitive processes used during sequential risk taking.

Discussion

During sequential risk-taking tasks, DMs make consecutive decisions to either stop the current round and collect their earnings or continue to play and take a gamble. The gamble offers them the chance to win more money, but if they lose the gamble they lose their earned money for that round and the round ends. People differ in the risks they take during the laboratory paradigm, and these systematic differences can be attributed to differences in their cognitive processes. The purpose of this article was to examine whether the learning process contributed to the paradigm's clinical diagnosticity and whether the cognitive processes people use during the tasks depend on the stochastic environment. In the discussion that follows, I will summarize the results of the study. Then I will discuss their implications in terms of the specific cognitive dimensions of learning, mental representations, and decision making. Finally, I will address the broader issue of developing a multitheory framework for studying risk-taking behavior.

Summary

The BSR model describes how people make decisions during the gambling paradigm and reveals how they differ on the underlying cognitive dimensions of learning, evaluation, and response. The model hypothesizes that prior to each round DMs evaluate how many trials they should play that would maximize their subjective expected payoff for each round. Then they probabilistically choose to play or stop on each trial as a function of the distance from the maximizing trial. In ill-defined conditions, like the cloudy conditions of the ART, DMs use the data they observed on previous rounds to learn about the task and employ Bayes' rule to update their beliefs about the parameters of the task.

How people evaluate expected payoffs and how they learn about the task depends on their mental representation. The results from this article show that DMs react to changes in their environment by changing their mental representation of the task. In contrast to their

Table 7
Correlations Between Adjusted ART Scores Calculated From Three Sets of Rounds for the Two Sunny Tournaments

ART condition	Rounds 1–10	Rounds 11–20	Rounds 21–30
Catch 'n' keep			
Sunny day	.37**	.42**	.29*
Cloudy day	.13	.17	.11
Catch 'n' release			
Sunny day	.29*	.24*	.36**
Cloudy day	.14	.06	.13

Note. ART = Angling Risk Tasks.
$^{*}p < .05$. $^{**}p < .01$.

Table 8
Hierarchical Linear Regression Analysis Testing the Incremental Validity Accounted For in Risky Drug Use

Variable	R^2	Adjusted R^2	ΔR^2	*df* error	*F*	*p*
Step 1						
Payoff sensitivity (γ^+)[a]	.12	.10	.12	66	8.6	<.01
Step 2						
Adjusted ART scores[b]	.15	.13	.04	65	2.7	*ns*
Step 3						
Risk Attitudes factor[c]	.24	.21	.09	64	7.7	<.01

Note. ART = Angling Risk Tasks.
[a] Refers to participants' payoff sensitivity during sunny tournaments.
[b] Data refer to scores during sunny tournaments.
[c] Refers to the Health Attitude and Social Attitude domains from the Domain-Specific Risk-Taking Scale (DOSPERT). DOSPERT scores do account for significant unique variance in Weighted Polydrug scores.

approach on the BART (see Bishara et al., 2007; Wallsten et al., 2005), here DMs adopted mental representations that were congruent with their stochastic environment. They used stationary representations during catch 'n' release and nonstationary representations during catch 'n' keep. Finally, it was not the learning process but rather the decision-making process, specifically how sensitive people were to changes in payoffs, that accounted for differences between self-reported drug users and non-users. In fact, the Bayesian learning processes were shown to interfere with this association, reducing the paradigm's ability to identify risky drug use.

Learning

How people learn about choice options influences their preferences and consequently their risk-taking behavior (see Barron & Erev, 2003; Denrell, 2007; Erev & Barron, 2005; Hertwig et al., 2004; March, 1996; Weber et al., 2004). During the Iowa Gambling Task (Bechara et al., 1994), the learning process even dissociates between different clinical populations and the risks they take. For example, during the Iowa Gambling Task both individuals with bilateral damage to the ventral-medial prefrontal cortex and patients with Huntington's disease choose the riskier option more often than normals do, and this difference is due to how these populations learn about the task (Busemeyer & Stout, 2002; Yechiam et al., 2005). Stout et al. (2005) even found that when the payoff structure of the Iowa Gambling Task was made explicit to participants, thereby removing the learning requirement, its clinical diagnosticity disappeared.

In the sequential risk-taking paradigm, the results presented here suggest the opposite conclusion. Learning obstructs the paradigm's clinical diagnosticity, and removing the learning process improves its ability to identify real-world risk takers. Only in the well-defined sunny versions were the adjusted ART scores correlated with self-reported past drug use. The BSR model, in turn, reveals that the general uncertainty of DMs reduces the sequential risk-taking paradigm's reliability. Once uncertainty and different experiences are accounted for with the BSR models, there is a significant association between the evaluation process (γ^+) and real-world risky behaviors.

The next logical question is, why does learning not appear to affect BART's validity? The BSR model can be used to answer this question as well. During the BART, Lejuez and colleagues (Lejuez, Aklin, Jones, et al., 2003; Lejuez, Aklin, Zvolensky, & Pedulla, 2003; Lejuez et al., 2002) try to control the between-participants variance by fixing experience between participants so that all 30 balloons explode at the same trial number for all participants. Besides restricting conclusions drawn from the BART to one particular sequence of failures and successes, this can only partially account for experience-based learning differences. DMs still have different experiences based on their own stopping decisions in previous rounds. For example, Lejuez and colleagues often fix the first balloon to explode on pump 65. Some DMs stop before this point, whereas others reach the point and find that it explodes, leading to different beliefs for the next round. This article indicates that a better alternative—if one is only interested in identifying risk takers—is to eliminate the learning process entirely and use well-defined variants of the sequential gambling paradigm like the sunny tournaments. Only when theoretically necessary (e.g., if a clinical population is thought to have different levels of optimism in their prior beliefs) should the learning process be implicated within the paradigm.

Mental Representations

Although across individuals the learning process does not show systematic differences, there are differences in the mental representations across different stochastic environments. In this experiment, DMs were remarkably adaptive in their learning process and their more general mental representations. During the nonstationary tasks (catch 'n' keep), the best fitting models were those that assumed respondents adopted a nonstationary representation; and during stationary tasks (catch 'n' release), the best fitting models assumed a stationary representation.

Certainly this congruence between representation and environment does not occur with every task. In the BART (a nonstationary environment), the best fitting BSR model in two different datasets assumes a stationary representation (Bishara et al., 2007; Wallsten et al., 2005). The congruence does not even occur with every model positing different mental representations. When models assuming DMs sequentially evaluate their options after each play option (as opposed to the BSR model used in this article that assumes DMs evaluate options prior to each round) were fit to the four conditions of this study, a stationary representation fit all four conditions better than did the same model assuming a nonstationary representation. Importantly, the quantitative fits for the sequential models were dismal compared with those for the BSR models presented in this article, giving more support to the prior evaluation assumption and the change in mental representation results. For a more detailed assessment of the sequential evaluation models, see Wallsten et al. (2005) and Pleskac (2004).

The more informative fishing task, as well as the within-subjects design, may have contributed to the ability of participants to discriminate between the different stochastic environments. The study's design cannot distinguish between these two explanations. Regardless, the experiment shows that DMs can be remarkably adaptive to their environment and that the often-used implicit or explicit assumption that DMs use the same decision strategy in these or similar experience-based tasks (e.g., Barron & Erev, 2003; Busemeyer & Stout, 2002; Wallsten et al., 2005; Yechiam et al., 2005) may be an overly strong assumption (see Erev & Barron, 2005 for a similar conclusion). As in multiattribute decisions, the cognitive strategies likely depend on other variables related to the structure of the task, the cognitive ability of the DMs, and the social context (see Gigerenzer et al., 1999; Payne et al., 1993).

Decision Making

Decision making, more specifically the evaluation process, plays a crucial role in the sequential risk-taking paradigm. For one thing, it can explain the relationship seen between the adjusted ART scores in the catch 'n' release and catch 'n' keep conditions (see left panel in Figure 2). Take for instance a risk-averse person with a constant level of payoff sensitivity across environments ($\gamma^+ = 0.8$ in Equations 2 and 6) and no probability weighting function. This risk-averse person will cast less in the catch 'n' release than in the catch 'n' keep sunny tournaments. In fact, she will set a target of making 48 casts in the sunny catch 'n' keep tournament and 39 casts in the sunny catch 'n' release tournament.

The difference between the conditions converges as the person grows more sensitive to changes in payoffs and reverses when there is increasing sensitivity to larger payoffs ($\gamma^{+} > 1$). A similar pattern holds in cloudy conditions if one assumes participants adopt a similar level of relative optimism in both catch 'n' keep and catch 'n' release conditions.[17] This prediction and empirical result speaks to the strength of formal cognitive models to bring clarity to seemingly peculiar and counterintuitive results.

Furthermore, the evaluation process and how sensitive people were to changes in payoffs showed systematic differences across respondents. Participants who reported lower levels of risky drug use were also less sensitive to changes in payoff gains in all four conditions of the tasks (for a similar conclusion see Stout, Busemeyer, Lin, Grant, & Bonson, 2004; Wallsten et al., 2005). More importantly, the regression analyses show that payoff sensitivity is responsible for the ability of the adjusted ART score to identify real-world risk takers. No stable association was found between how respondents weighted probabilities and drug use. Taking this result in combination with the fact that the BSR model without the weighting function adequately accounted for the data suggests that the added weighting function component is unnecessary at this point.

The evaluation process has several implications for the sequential risk-taking paradigm itself. First, it suggests that the tedious button pressing needed in the BART (to pump the balloon) and the ART (to cast the rod) may be unnecessary. If participants already have a target number of play options in mind, then an alternative response mode would be for them to enter the number into the computer before each round and watch the computer play for them, an automatic type of game. Such adaptations have been developed, and in accordance with the model prediction, there is little difference in performance between the automatic version and the standard button-pressing version (see Pleskac et al., 2007). One advantage, besides decreasing the total amount of time required to complete the BART, is that the automatic version also makes it possible to empirically observe intended risky behavior in all rounds (see stars in the right panel of Figure 2 for the expected targeted casts for all rounds).

A second implication of the evaluation process is the role that losses play in the sequential gambling paradigm. The prior evaluation policy indicates that respondents frame the task as a gain domain, not a mixed outcome domain (see Equation 1 for an example). As a result, losses are not currently considered in the already clinically diagnostic ART. Behavioral decision theory, though, usually predicts that losses and aversion to losses play a critical role in risk-taking behavior (see Kahneman & Tversky, 1979). Consequently, one might predict that adjustments to the payoff structure to include losses would only improve the paradigm's already notable ability to identify real-world risk takers. This prediction is beyond the scope of this article, but certainly it opens an intriguing avenue of research—one that would not have been revealed without the cognitive model.

A Multitheory Framework for Understanding Risk Taking

Risk-taking behavior does not arise from one single psychological process. The challenge then, for social and cognitive scientists with a goal of understanding risky behavior, is to build a theoretical framework that draws on these different processes and reaches beyond the traditional boundaries in psychology. The results of such a framework have great potential. For example, in this article the combined approach of a sequential risk-taking paradigm, with its roots in the clinical sciences, and the DOSPERT, developed based on theories of risk perception, together explained 24% of the variance in the reported risky use of drugs. Separately they accounted for 12%–14% of the variance. Cognitive models, in turn, led to a better understanding of how learning and decision processes co-occur in the gambling paradigm and identified the processes critical for its assessment ability: In identifying risky drug users, evaluation processes are essential, whereas learning reduces the paradigm's reliability. Only by integrating these and other approaches will the social and cognitive sciences come to not only appreciate the complexity of risky behavior at the theoretical level, but also be able to make substantial contributions in understanding the potential public-health problems of unsafe and unhealthy risk taking.

[17] This equal optimism assumption is much like setting the catch 'n' release ponds to have 65 fish and the catch 'n' keep ponds to have 128 fish, so that if a DM was an expected value maximizer and had perfect knowledge of the ponds, she would cast approximately 64 times to maximize her earnings.

References

Aklin, W. M., Lejuez, C. W., Zvolensky, M. J., Kahler, C. W., & Gwadz, M. (2005). Evaluation of behavioral measures of risk taking propensity with inner city adolescents. *Behaviour Research and Therapy, 43,* 215–228.

Babor, T. F., Hofmann, M., DelBoca, F. K., Hasselbeck, V., Meyer, R., Dolinsky, Z. S., & Rounsaville, B. (1992). Types of alcoholics: I. Evidence for an empirically derived typology based on indicators of vulnerability and severity. *Archives of General Psychiatry, 49,* 599–608.

Barron, G., & Erev, I. (2003). Small feedback-based decisions and their limited correspondence to description-based decisions. *Journal of Behavioral Decision Making, 16,* 215–233.

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

Berry, D., & Fristedt, B. (1985). *Bandit problems.* London: Chapman and Hall.

Bishara, A. J., Pleskac, T. J., Fridberg, D. J., Yechiam, E., Lucas, J., Busemeyer, J. R., et al. (2007). Models of risky decision-making in stimulant and marijuana users. *Manuscript submitted for publication.*

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,* 253–262.

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

Crowley, T. J., Raymond, K. M., Mikulich-Gilbertson, S. K., Thompson, L. L., & Lejuez, C. W. (2006). A risk-taking "set" in a novel task among adolescents with serious conduct and substance problems. *Journal of the American Academy of Child and Adolescent Psychiatry, 45,* 175–183.

Denrell, J. (2007). Uncertainty avoidance in experiential learning. *Psychological Review, 114,* 177–187.

Erev, I., & Barron, G. (2005). On adaptation, maximization, and reinforcement learning among cognitive strategies. *Psychological Review, 112,* 912–931.

Fox, C. R., & Tversky, A. (1998). A belief-based account of decision under uncertainty. *Management Science, 44,* 879.

Gelman, A., Carlin, J. B., Stern, H. S., & Rubin, D. B. (2003). *Bayesian data analysis* (2nd ed.). Boca Raton, FL: Chapman & Hall/CRC.

Gigerenzer, G., Todd, P. M., & the ABC Research Group (1999). *Simple heuristics that make us smart.* New York: Oxford University Press.

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

González-Vallejo, C. (2002). Making trade-offs: A probabilistic and context-sensitive model of choice behavior. *Psychological Review, 109,* 137–154.

Grant, S., Contoreggi, C., & London, E. D. (2000). Drug abusers show impaired performance in a laboratory test of decision making. *Neuropsychologia, 38,* 1180–1187.

Hanoch, Y., Johnson, J. G., & Wilke, A. (2006). Domain specificity in experimental measures and participant recruitment. *Psychological Science, 17,* 300–304.

Hertwig, R., Barron, G., Weber, E. U., & Erev, I. (2004). Decisions from experience and the effect of rare events in risky choice. *Psychological Science, 15,* 534–539.

Hertwig, R., & Ortmann, A. (2001). Experimental practices in economics: A methodological challenge for psychologists? *Behavioral and Brain Sciences, 24,* 383–403.

Hey, J. D. (1998). Experimental economics and deception: A comment. *Journal of Economic Psychology, 19,* 397–401.

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,* 249–260.

Howell, D. C. (1997). *Statistical methods for psychology.* Belmont, CA: Wadsworth Publishing.

Jamieson, B. D. (1969). The influences of birth order, family size, and sex differences on risk-taking behavior. *British Journal of Social and Clinical Psychology, 8,* 1–8.

Johnston, L. D., O'Malley, P. M., Bachman, J. G., & Schulenberg, J. E. (2005). *Monitoring the future: National survey results on drug use, 1975–2004: Vol. I. Secondary school students* (NIH Publication No. 05–5727). Bethesda, MD: National Institute on Drug Abuse.

Jones, H. A., & Lejuez, C. W. (2005). Personality correlates of caffeine dependence: The role of sensation seeking, impulsivity, and risk taking. *Experimental and Clinical Psychopharmacology, 13,* 259–266.

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

Lejuez, C. W., Aklin, W. M., Bornovalova, M. A., & Moolchan, E. T. (2005). Differences in risk-taking propensity across inner-city adolescent ever- and never-smokers. *Nicotine & Tobacco Research, 7,* 71–79.

Lejuez, C. W., Aklin, W. M., Jones, H. A., Richards, J. B., Strong, D. R., Kahler, C. W., et al. (2003). The Balloon Analogue Risk Task (BART) differentiates smokers and nonsmokers. *Experimental and Clinical Psychopharmacology, 11,* 26–33.

Lejuez, C. W., Aklin, W. M., Zvolensky, M. J., & Pedulla, C. M. (2003). Evaluation of the Balloon Analogue Risk Task (BART) as a predictor of adolescent real-world risk-taking behaviours. *Journal of Adolescence, 26,* 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,* 75–84.

Lejuez, C. W., Simmons, B. L., Aklin, W. M., Daughters, S. B., & Dvir, S. (2004). Risk-taking propensity and risky sexual behavior of individuals in residential substance use treatment. *Addictive Behaviors, 29,* 1643–1647.

Luce, R. D. (2000). *Utility of gains and losses measurement: Theoretical and experimental approaches.* Mahwah, NJ: Erlbaum.

March, J. G. (1996). Learning to be risk averse. *Psychological Review, 103,* 309–319.

Meng, X. I., Rosenthal, R., & Rubin, D. B. (1992). Comparing correlated correlation coefficients. *Psychological Bulletin, 111,* 172–175.

Montgomery, R. (1974). Transmission of risk-taking through modeling at two age levels. *Psychological Reports, 34,* 1187–1196.

Nelder, J. A., & Mead, R. (1965). A simplex method for function minimization. *Computer Journal, 7,* 308–313.

Payne, J. W., Bettman, J. R., & Johnson, E. J. (1993). *The adaptive decision maker.* New York: Cambridge University Press.

Pleskac, T. J. (2004). *Evaluating cognitive sequential risk-taking models: Manipulations of the stochastic process.* Unpublished dissertation, University of Maryland, College Park.

Pleskac, T. J., Wallsten, T. S., Wang, P., & Lejuez, C. W. (2007). *Valid score adjustments require valid assumptions: The case of the Balloon Analogue Risk Task.* Manuscript submitted for publication.

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

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

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

Stout, J. C., Busemeyer, J. R., Lin, A. L., Grant, S. J., & Bonson, K. R. (2004). Cognitive modeling analysis of decision-making processes in cocaine abusers. *Psychonomic Bulletin and Review, 11,* 742–747.

Stout, J. C., Rock, S. L., Campbell, M. C., Busemeyer, J. R., & Finn, P. R. (2005). Psychological processes underlying risky decisions in drug abusers. *Psychology of Addictive Behaviors, 19,* 148–157.

Thaler, R. H., & Johnson, E. J. (1990). Gambling with the house money and trying to break even: The effects of prior outcomes on risky choice. *Management Science, 36,* 643–660.

Tversky, A., & Fox, C. R. (1995). Weighing risk and uncertainty. *Psychological Review, 102,* 269–283.

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

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

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

Wakker, P., & Tversky, A. (1996). Properties of the weighting function in cumulative prospect theory. *Journal of Mathematical Psychology, 40,* 363–363.

Wallsten, T. S., Pleskac, T. J., & Lejuez, C. W. (2005). Modeling behavior in a clinically-diagnostic sequential risk-taking task. *Psychological Review, 112,* 862–880.

Wasserman, L. (2000). Bayesian model selection and model averaging. *Journal of Mathematical Psychology, 44,* 92–107.

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,* 263–290.

Weber, E. U., Shafir, S., & Blais, A. R. (2004). Predicting risk sensitivity in humans and lower animals: Risk as variance or coefficient of variation. *Psychological Review, 111,* 430–445.

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

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

Yechiam, E., Busemeyer, J. R., Stout, J. C., & Bechara, A. (2005). Using cognitive models to map relations between neuropsychological disorders and human decision-making deficits. *Psychological Science, 16,* 973–978.

Appendix A

The Likelihood Function for the Bayesian Sequential Risk-Taking (BSR) Model and Details on Fitting the Model

Each model was fit to each individual's data from each tournament using maximum likelihood methods. Let the vector $\mathbf{Y}_{w,l} = (c_{w,l,1}, d_{w,l,1}, \cdots, c_{w,l,h}, d_{w,l,h}, \cdots, c_{w,l,30}, d_{w,l,30})$ represent the observed data from tournament w,l, where $w = s$ or c for sunny or cloudy conditions, respectively; $l = k$ or r for keep or release, respectively; $c_{w,l,h}$ is the number of casts on Round h; and $d_{w,l,h}$ is whether the DM stopped (1) or not (0). The likelihood of the observed data, $\mathbf{Y}_{w,l}$, for each of the models is defined as

$$L(c_{w,l,1}, d_{w,l,1}, \ldots, c_{w,l,h}, d_{w,l,h}, \ldots, c_{w,l,30}, d_{w,l,30})$$

$$= \prod_{h=1}^{30} \prod_{i=1}^{c_{w,l,h}} \hat{r}_{h,i}(1-\hat{r}_{h,c_{w,l,h}+1})^{d_{w,l,h}}, \quad \text{(A1)}$$

where each model predicts $r_{h,i}$. Taking the natural logarithm gives the log-likelihood of the data.

None of the BSR models have a closed form solution to find the maximum likelihood estimates (MLE) of the parameters. Consequently, the solutions were estimated with numerical optimization techniques, of which there are many. For computational purposes, I constrained the valuation parameter to $0 \leq \gamma^{-}, \gamma^{+} \leq 3$ and the mean of the γ distribution to $0 < \mu_o \leq 1{,}000$.

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. The Nelder–Mead downhill simplex routine (see Nelder & Mead, 1965) in conjunction with a grid-search technique uses a two-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 ensure the starting values would lead to a solution below a prespecified criterion (e.g., log-likelihood > –2,000). If not, then the set was iteratively perturbed with random noise, and tested, until the criterion was met or a cutoff was reached. Parameter sets that did meet the criterion 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 log-likelihood from the full set was taken as the estimate.

Examples of other procedures used include nonlinear programming, genetic algorithm, or an iterative annealing Nelder–Mead method (see Van Zandt, 2000; Van Zandt, Colonius, & Proctor, 2000). However, simulations and past experience showed that the aforementioned procedure performed the best.

Appendix B

Derivation of the Predicted Adjusted Scores

To derive the predicted adjusted scores from the Bayesian sequential risk-taking (BSR) model, we must find the probability of observing score y given the round ended in a success, $P(y \mid s)$. First, note that the BSR model predicts the target number of trials for which DMs should select the play option for each round (see for example Equations 2 and 6). The intended casting behavior for each round, called score y, is a function of the response rule in Equation 3. This score is empirically unobservable, and the adjusted score is an estimate of it. Using the model, the probability of score y for all rounds is the probability of choosing the play option up to and stopping on Trial y,

$$P_h(y) = \prod_{i'=1}^{y-1} r_{i',h}[1 - r_{y,h}], \quad \text{(B1)}$$

where $r_{i,h}$ is the probability of choosing the play option and is found with Equation 3. In catch 'n' release conditions the score y can take the values $[0, \infty]$. In catch 'n' keep conditions y can only take values $[0, n]$. The expected score, or intended casting behavior, for each round is then

$$E(y) = \Sigma y P_{h(y)}. \quad \text{(B2)}$$

The adjusted score is an estimate of the expected score given the round ended in a success, $E_h(y \mid S)$.[B1] Bayes' rule identifies the probability of observing score y given a success, S, on Round h,

$$P_h(y|S) = \frac{P(S|y)P_h(y)}{\Sigma P(S|y)P_h(y)}. \quad \text{(B3)}$$

The values of $P(S \mid y)$ depend on the stochastic environment. In the catch 'n' keep conditions, the a priori probability of the failure event, F, on any trial is $1/n$; the probability of no failure (i.e., of a success S, given y sequential successful trials) is

$$P(S|y) = \frac{n - y}{n}. \quad \text{(B4)}$$

In catch 'n' release, the a priori and the conditional probability of any one trial leading to a failure is probability p. The probability, therefore, that y observations will all be successful is

[B1] My thanks to Thomas Wallsten for showing me the analytic solution for distribution of scores given a success.

(*Appendixes continue*)

$$P(S|y) = (1 - p)^y. \quad \text{(B5)}$$

Calculating $E_h(y \mid S)$ with Equation B5 is straightforward. In cloudy tournaments, both n and p are replaced with estimated values from the Bayesian learning models (see Equations 5 and 6).

The bar graph in the right panel of Figure 2 plots the expected adjusted ART scores averaged across all rounds and across all participants using the parameter estimates from the models estimated without a weighting function. The models with the weighting function produce a similar graph. The stars in the panel are the predicted scores from Equation B2. Pleskac et al. (2007) proved that in sequential risk-taking tasks, the adjusted scores will always be less than the predicted unconditional latent scores, $E_h(y \mid S) < E_h(y)$. This is because the more trials the DM chooses the risky play option on a given round, the more likely it is to end in a failure (i.e., catching a blue fish). Therefore, the adjusted score tends to filter out the longer response sequences.

This tendency of the longer response sequences to result in a failure and be excluded from the adjusted score suggests that an error theory might be sufficient to explain both the observed risk aversion with the adjusted scores and the systematic individual differences. Indeed, for bias-free DMs maximizing expected value ($\gamma^+ = 1$ in Equations 2 and 6), the predicted adjusted ART scores will fall as response consistency (β) decreases (or error increases; see Equation 3). To show this prediction, Table B1 lists the predicted adjusted ART scores for three bias-free expected value maximizers ($\gamma^+ = 1$) in the catch ‘n’ keep, sunny condition. The calculations were done using Equations 3, 6, B1, and B3. As β decreases, the predicted adjusted ART score also falls. However, the predicted standard deviation of the number of casts made on the stop rounds increases. Thus, the models predict that if an error theory is sufficient to explain the data, then the adjusted ART scores and the standard deviation of casts taken on stop trials should be negatively correlated. This is not the case for the dataset presented here. There was a significant positive correlation between the adjusted ART scores and standard deviation of casts taken on rounds ending in a stop for all four conditions, $r_{(\text{keep,sunny})} = .44$; $r_{(\text{keep,cloudy})} = .57$; $r_{(\text{release, sunny})} = .74$; $r_{(\text{release,cloudy})} = .75$, thus indicating that response variability is not sufficient and other cognitive processes are necessary to explain the data.

Table B1
Expected Value Maximizers for Three Levels of Response Variability (β) Playing Catch ‘n’ Keep on Sunny Days

β	Predicted adjusted ART score	Predicted *SD* of casts taken on stop rounds
0.3	58.0	4.7
0.5	52.7	6.9
0.1	34.8	11.9

Note. ART = Angling Risk Tasks.

Appendix C

Simultaneous Estimation of the Value and Weighting Function for the Bayesian Sequential Risk-Taking Model

The data set from the entire experimental session is needed to simultaneously estimate both the value and weighting function, because the weighting and the value function for each individual can be estimated for each tournament only if their respective parameters are constrained across different stochastic representations (i.e., across catch ‘n’ keep and catch ‘n’ release; for a proof see Pleskac, 2004). The most reasonable constraint is setting γ^+ and α equal across the nonstationary and stationary models for the two release laws within a weather condition. None of the other parameters are constrained across tournaments. This constraint is summarized in Table C1, under Model Framework 2. Model Framework 1 is the set of models without the weighting function that has already been estimated for the four tournaments. Each framework has 16 total free parameters that account for all 120 rounds in the entire experimental session, with an average of 3,360 choice trials.

The frameworks are not nested and have an equal number of parameters, so comparisons of the maximum log-likelihood (ML) of the data for the experimental session given each framework can only serve descriptive purposes. The results in Table C2 show that Model Framework 1 provided a better fit than did Model Framework 2 for a majority of the participants when the value function allows both diminishing sensitivity to outcomes and increasing sensitivity to outcomes ($\gamma^+ > 0$).

Table C1
Constraints Imposed on the BSR Model to Estimate the Weighting Function

Variable	Model Framework 1		Model Framework 2	
	Cloudy	Sunny	Cloudy	Sunny
Learning	Yes (4)	No	Yes (4)	No
Evaluation				
Value function	Yes (2)	Yes (2)	Yes (1)	Yes (1)
Weighting function	No	No	Yes (1)	Yes (1)
Response				
Bias	Yes (2)	Yes (2)	Yes (2)	Yes (2)
Payoff sensitivity (γ^+)	Yes (2)	Yes (2)	Yes (2)	Yes (2)
Total parameters	16		16	

Note. Model Framework 1 is the set of models without the weighting function. Model Framework 2 is the set constrained to estimate the weighting function. On cloudy and sunny days, both catch ‘n’ keep and catch ‘n’ release tournaments were held. For example, Model Framework 1 has a value function and has 2 γ^+ parameters in the cloudy conditions and 2 in the sunny conditions. BSR = Bayesian sequential risk-taking model.

An alternative constraint is to allow only diminishing sensitivity ($0 < \gamma^+ \leq 1$). With this constraint, Model Framework 1 was a better fit for only a few more participants than Framework 2 (see Table C2).

Table C2
Comparison of Models With Different Constraints to Estimate a Weighting Function, With Varying Constraints on γ^+

	Increasing and decreasing sensitivity to outcomes ($0 < \gamma^+$)		Decreasing sensitivity to outcomes ($0 < \gamma^+ \leq 1$)	
Model	Mean ML	Number of DMs best fit	Mean ML	Number of DMs best fit
Baseline	−362.01	0	−362.01	0
Model Framework 1 (without weighting function)	−270.32	69	−343.98	38
Model Framework 2 (with weighting function)	−280.76	2	−282.27	33

Note. ML = maximum log likelihood; DM = decision maker.

However, the mean ML for Model Framework 2 is greater than the mean ML for Framework 1. This indicates that for some individuals, Model Framework 1's fit is quite poor as compared with Framework 2. A better sense of this for each participant was ascertained by calculating the absolute difference between ML for Model Framework 1 and Model Framework 2. For the 38 individuals for whom Model Framework 1 (no weighting) was a better fit, the average absolute difference was 6.01 (in log-likelihood space). But for the 33 individuals for whom Model Framework 2 (weighting) was a better fit, the average deviation was 200.65. In other words, when Model Framework 2 (with the weighting function) was a poor fit, it was much less a poor fit than was Model Framework 1 (without the weighting function). These results indicate that, based on fit alone, the weighting function can contribute to the explanatory power of the model, but only when γ^+ is constrained to show only diminishing sensitivity.

Received July 10, 2006
Revision received June 25, 2007
Accepted September 29, 2007 ■