Making Assessments While Taking Repeated Risks:
A Pattern of Multiple Response Pathways

Timothy J. Pleskac
Avishai Wershbale
Michigan State University

Word count (including abstract): 13,029

Author Note

Timothy J. Pleskac, Department of Psychology, Michigan State University.

A grant from the National Science Foundation (0955410) supported this work. We thank David McFarlane for help in programming the BART and Alex Jendrusina for help in data collection. We are grateful for the input from Todd Gureckis, Matt Zeigenfuse, James Grand, Thomas Wallsten, and the Laboratory for Cognitive & Decision Sciences. We also thank Patrycja Zdziarska and Mitchell Uitvlugt for assistance in preparing the manuscript.

Correspondence concerning this article should be addressed to Timothy J. Pleskac, Department of Psychology, Michigan State University, East Lansing, MI, 48824. E-mail: tim.pleskac@gmail.com
Abstract

Beyond simply a decision process, repeated risky decisions also require a number of cognitive processes including learning, search and exploration, and attention. In this paper we examine how multiple response pathways develop over repeated risky decisions. Using the Balloon Analogue Risk Task (BART) as a case study, we show that two different response pathways emerge over the course of the task. The assessment pathway is a slower, more controlled pathway where participants deliberate over taking a risk. The second pathway is a faster, more automatic process where no deliberation occurs. Results imply the slower assessment pathway is taken as choice conflict increases and that the faster automatic response is a learned response. Based on these results, we modify an existing formal cognitive model of decision making during the BART to account for these dual response pathways. The slower more deliberative response process is modeled with a sequential sampling process where evidence is accumulated to a threshold, while the other response is given automatically. We show that adolescents with conduct disorder and substance use disorder symptoms not only evaluate risks differently during the BART, but also differ in the rate at which they develop the more automatic response. More broadly, our results illustrate a challenge to cognitive models of judgment decision making to move from modeling decision making as the result of a single response pathway to multiple response pathways that change and develop over time.

Keywords: dual process, risk taking, sequential risk taking, conduct disorder, sequential sampling model
Making Assessments While Taking Repeated Risks:

A Pattern of Multiple Response Pathways

Risks like texting while driving, eating unhealthy food for a quick lunch, or smoking a cigarette, are typically not one-shot choices. Rather, these are choices that are made repeatedly over the course of a day, a week, a month, and so on. Repeated decisions put different demands on a cognitive system than one-shot choices. They, for example, allow for and even may require learning (Busemeyer & Stout, 2002; Denrell, 2007; Pleskac, 2008; Wallsten, Pleskac, & Lejuez, 2005) as well as search and exploration (Daw, O'Doherty, Dayan, Seymour, & Dolan, 2006). They also put different demands on attention and memory (Barron & Erev, 2003) and may even change the decision process itself (Jessup, Bishara, & Busemeyer, 2008). In this paper, we examine how multiple response pathways develop while people make repeated decisions.

Judgments and decisions under uncertainty are often described as being made with one of two information-processing systems: System 1 or System 2 (Evans, 2008; Kahneman, 2003; Mukherjee, 2010; Reyna, 2004; Sloman, 1996; Stanovich & West, 2000) (cf. Gigerenzer & Regier, 1996; Keren & Schul, 2009; Kruglanski & Gigerenzer, 2011). System 1 makes judgments that are fast, automatic, effortless, associative in nature, and undemanding on computational capacity. System 2, in comparison, makes judgments that are slower, controlled, rule-based, and demanding of computational capacity. Cognitive theory tells us that the development of an automatic response process—one characteristic of System 1—often requires an appreciable amount of repeated and similar actions (Schneider & Shiffrin, 1977; Shiffrin & Schneider, 1977). Thus, making repeated decisions should be sufficient for the development of an automatic response.
Neurocomputational accounts of decision making provide further evidence for multiple systems (Daw, Niv, & Dayan, 2005; Frank & Claus, 2006; Frank, Cohen, & Sanfey, 2009; Glascher, Daw, Dayan, & O’Doherty, 2010). One system is described as habitual or reflexive (Dickinson, 1985). It sequentially builds an association between an alternative and its expected reward via a reward prediction error. These prediction signals are typically localized to the striatum (Frank & Claus, 2006; O’Doherty, 2004). At the same time, a second more goal-directed system also drives decisions. This system, using the orbito frontal cortex, appears to maintain recent event contingencies between events and outcomes (Frank & Claus, 2006) or even a more detailed model of the decision situation (Daw et al., 2005).

Despite the evidence for multiple response systems, many formal models of risky decision making posit a single attention demanding controlled response pathway. One reason perhaps for this focus is that the theories and models are tested using single-shot gambles as critical tests of a particular theory (e.g., Birnbaum, 1999; Brandstatter, Gigerenzer, & Hertwig, 2006; Busemeyer & Townsend, 1993; Tversky & Kahneman, 1992). Thus, an automatic response is not allowed to develop, and even if a more automatic response did develop, it would be difficult to identify it from a more deliberate response. However, this assumption of a single controlled response pathway remains even as more complex decision making tasks are studied. For example, consider the application of cognitive models to laboratory-based gambling tasks like Bechara, Damasio, Damasio, and Anderson’s (1994) Iowa Gambling Task and Lejuez et al.’s (2002) Balloon Analogue Risk Task (BART). These tasks typically require participants to make a large number of risky choices with trial-by-trial feedback. They also typically reveal decision making deficits in populations with neurological and clinical disorders. Cognitive models, in turn, have been used to better identify and quantify the processes and deficits during
these tasks (e.g., Busemeyer & Stout, 2002; Pleskac, 2008; Wallsten et al., 2005; Yechiam, Busemeyer, Stout, & Bechara, 2005). These models based on judgment and decision making principles give a good account of the data, they do so by assuming and testing for a single controlled deliberative response pathway.

We sought to directly test this assumption of a single attention demanding controlled response pathway and investigate if there is evidence of a second more automatic response pathway. To do so, we focused on the BART, which is a specific instance of a larger class of tasks that include the devil task (Slovic, 1966), the angling risk task (Pleskac, 2008), and the Columbia card task (Figner, Mackinlay, Wilkening, & Weber, 2009). During the BART, participants are presented with a computerized balloon and two response buttons. One button inflates the balloon by a small amount and puts money in a temporary bank, while the other button ends the trial and transfers any money earned during that trial from the temporary bank to a permanent bank. Like an actual balloon, this one will eventually explode if pumped too many times.¹ When an explosion occurs, it terminates the trial and all money in the temporary bank is lost.

There are several reasons for using the BART to address the question of multiple response pathways. One reason, as already mentioned, is that the decision process during the BART and a broader range of sequential risk taking tasks appears to be well described by a

¹ The BART is programmed so that on each balloon trial the computer selects a random number between 1 and 128. This number determines how many successful pumps can be made before the balloon explodes. Thus, the optimal policy assuming a person is fully informed about the statistical structure of the BART is to take between 64 and 65 pumps (Pleskac, Wallsten, Wang, & Lejuez, 2008).
computational model called the Bayesian sequential risk-taking (BSR) model (Pleskac, 2008; Wallsten et al., 2005). The BSR posits a single response process pathway and provides a principled starting point for understanding how multiple response processes may be realized within the same task and in risky decision making in general.

A second reason for using this task stems from the suggestion that during sequential risk taking tasks different response pathways are used depending on how decision makers enter their choices (Figner et al., 2009). Figner et al. argue that decision makers use a more affective, spontaneous, and automatic decision process when they sequentially decide between taking a risk (pumping) and stopping (the standard response process in the BART). If, however, the task’s response mode is altered so that at trial onset decision makers must first estimate the number of risks they want to take and the computer then carries them out, then apparently the decision process is more deliberative/cold (Figner et al.). Our argument goes beyond this perspective and claims that even within the standard response entry procedure of the BART, there are multiple response pathways that develop during the course of the task. In particular, we argue that an automatic response pathway develops in the BART because of the high number of very similar responses (~900) per session of 30 balloon trials. This hypothesis is consistent with cognitive theory that suggests automatic cognitive processes develop during tasks that have an appreciable amount of repeated and similar actions (Schneider & Shiffrin, 1977; Shiffrin & Schneider, 1977).

Finally, the BART is of interest because risk taking behavior in the BART and similar tasks correlate with actual risk taking behavior outside a laboratory setting (Aklin, Lejuez, Zvolensky, Kahler, & Gwadz, 2005; Bornovalova, Daughters, Hernandez, Richards, & Lejuez, 2005; Hoffrage, Weber, Hertwig, & Chase, 2003; Hopko et al., 2006; Lejuez, Aklin, Bornovalova, & Moolchan, 2005; Lejuez, Aklin, Jones, et al., 2003; Lejuez, Aklin, Zvolensky, &
Pedulla, 2003; Lejuez, Simmons, Aklin, Daughters, & Dvir, 2004; Pleskac, 2008). This relationship raises the question of whether risk takers may differ not only in the risks they take during the BART, but also in the response pathways they use in taking sequential risks. To this end, we report a re-analysis of Crowley, Raymond, Milulich-Gilbertson, Thompson, and Lejuez (2006) where adolescents with conduct disorder/substance use (CD/SU) completed the BART. This re-analysis shows a critical difference between the groups in terms of their development of the automatic response pathway. It also illustrates how modeling choice and response times from these laboratory-based gambling tasks, something that is seldom if ever done, can provide a deeper cognitive understanding of decision deficits among clinical populations (though see Ratcliff, Thapar, & McKoon, 2001; 2004 for the implications in perceptual decision making).

**Bayesian Sequential Risk Taking (BSR) Model**

As a means of grounding our hypotheses about the development of multiple response pathways, it is useful to describe the BSR model in greater detail. According to the BSR, participants use three distinct processes when completing the BART: a learning process, a reward evaluation process, and a response selection process (Pleskac, 2008; Wallsten et al., 2005).

**Learning from experience.** The BSR model assumes that in completing the BART, decision makers evaluate the expected gains from each pump by weighting the potential gains from pumping the balloon with the probability of reaching those gains. The challenge is that decision makers are told very little about the task: they are only informed that somewhere between the first pump and when the balloon fills the screen the balloon will explode. Our past modeling work has revealed that the vague instructions result in decision makers relying on a learning process to infer the probability of the balloon not exploding. This learning process is
best described as a Bayesian learning process where decision makers (incorrectly) treat the balloon explosions as a stationary sampling-with-replacement process (Pleskac, 2008).

According to the model, decision makers have a belief in the probability $q_h$ that balloon $h$ will not explode on any given pump. This belief is modeled with a beta distribution over $q_h$. The beta distribution is a function of two parameters $a_h > 0$ and $b_h > 0$. The mean of the beta distribution is used as the estimate of decision makers’ beliefs in the chances of balloon $h$ not exploding. The mean is found with the following formula

$$\hat{q}_h = \frac{a_h}{a_h + b_h} \quad (1)$$

The variance is

$$\delta_h = \frac{a_h b_h}{(a_h + b_h)^2(a_h + b_h + 1)} \quad (2)$$

and indexes decision makers’ uncertainty in those beliefs for each balloon.

The beta distribution is a conjugate distribution of the binomial, so the Bayesian updating process is straightforward (Gelman, Carlin, Stern, & Rubin, 2003; Kruschke, 2010). After balloon $h$, $b_h$ is increased by 1 if it exploded ($b_{h+1} = b_h + 1$); otherwise $b_{h+1} = b_h$. The parameter $a_h$ is incremented by the total number of pumps that did not end in an explosion. If the balloon exploded then $a_{h+1} = a_h + c_h - 1$ where $c_h$ is the total number of pumps on the previous balloon. If the balloon did not end with an explosion, then $a_{h+1} = a_h + c_h$. The learning component has two free parameters $\tilde{q}_1$ and $\delta_1$. The parameter $\tilde{q}_1$ indexes the degree of initial optimism that decision makers have in their beliefs that the balloon will not explode. Participants with high $\tilde{q}_1$ will tend to make more pumps. The reason why will become more apparent in the description of the

---

$^2$ In fitting the model we actually treat $a_1$ and $b_1$ as free parameters. The parameters $\tilde{q}_1$ and $\delta_1$ can be found from Equations 1 and 2.
reward evaluation process. The parameter $\delta_i$ indexes the degree of initial uncertainty decision makers have in their beliefs. Decision makers with high uncertainty will be more sensitive to observed pumps and explosions in their updating process.

**Reward evaluation.** At the start of each balloon trial, decision makers evaluate their potential options of pumping and stopping according to the expected gain on trial $h$ for each pump opportunity $i$,

$$v_{h,i} = (\hat{q}_h)^i (ix)^{\gamma^+}. \quad (3)$$

The value $(\hat{q}_h)^i$ is the probability that balloon $h$ will not explode after $i$ pumps, and $x$ is the reward for a successful pump. Lower values of $\gamma^+$ indicate less sensitivity to changes in payoffs, and higher values indicate greater sensitivity. Participants are assumed to target a pump $G_h$ that maximizes expected payoffs. This target pump is the maximum of Equation 3, which is

$$G_h = -\frac{\gamma^+}{\ln(\hat{q}_{h})}. \quad (4)$$

Equation 4 illustrates how participants with different values of $\gamma^+$ will behave differently during the BART. Participants with greater sensitivity to payoffs (larger values of $\gamma^+$) will have larger target values $G_h$, and so they will typically choose to pump more on a given balloon. Equation 4 also illustrates how different mean beliefs $\hat{q}_h$ from the learning process shapes behavior with higher values (greater optimism) leading to larger targets.

**Response selection.** At each pump opportunity $i$, decision makers probabilistically choose between pumping and stopping. This choice is based on an assessment of the distance from the targeted pump, $d_h(i) = i - G_h$. The response rule assumes that the probability of choosing to pump $r_h(i)$ on balloon $h$ at pump opportunity $i$ strictly decreases with decreasing
distance. Decision makers are assumed indifferent between pumping and stopping at $G_h$. We use a logistic response rule to capture these properties

$$\Pr[\text{Pump}_h(i)] = r_h(i) = \frac{1}{1 + \exp[\beta d_h(i)]},$$

where $\beta$ is a free parameter representing how consistently participants follow their targeted evaluation. Lower values of $\beta$ are indicative of participants being sensitive to other information besides their targeted reward pump and thus will be more variable in their pumping behavior.

**Single or Multiple Response Pathways?**

The response rule in Equation 5 illustrates the single response pathway assumption of the BSR: On every single pump, decision makers are assumed to carry out a distance-to-target assessment. Presumably this distance assessment is a controlled, attention demanding, capacity limited process. The single response pathway is no doubt a parsimonious description of the decision process and similar assumptions are made in other laboratory-based gambling tasks (Bishara et al., 2010; Busemeyer & Stout, 2002; Yechiam et al., 2006). However, given the results in other areas of decision making that implicate multiple response pathways, one might question the plausibility of this single response pathway. Moreover, given the high number of repeated choices made during the BART, one would expect that the development of an automatic pump response is likely to occur (Schneider & Shiffrin, 1977; Shiffrin & Schneider, 1977).

A mixture of deliberative and automatic responses is also consistent with maze learning experiments with rats. In these experiments, rats were observed to pause at choice points and orient themselves toward potential options (Tolman, 1938, 1948). This behavior was called vicarious trial and error (VTE; Muenzinger, 1938; Tolman, 1938). Tolman (1948) characterized VTE activity as “not just one of responding passively to discrete stimuli, but rather one of the active selecting and comparing of stimuli” (p. 200). Neural evidence links VTEs to hippocampal
activity in rats, suggesting that deliberative, planning-type processes are occurring during VTE activity (Hu & Amsel, 1995; Hu, Xu, & Gonzalez-Lima, 2006; Johnson, van der Meer, & Redish, 2007). These observations lead to a general hypothesis about decision making during the BART:

**Assessment hypothesis.** Human participants engage in distance-to-target assessments on pump opportunities during the BART. Diverging from the BSR model (Pleskac, 2008; Wallsten et al., 2005), we suggest that this assessment occurs only on a select few pump opportunities while on the other pump opportunities participants make more automatic pump responses. In the studies below we examined this hypothesis in several different ways. Behaviorally, we tested this hypothesis by comparing response times (RTs) during the BART to RTs from a baseline BART. During the baseline BART, participants were told to simply pump the balloon until it explodes. Thus, RTs in the baseline BART should reflect baseline-pumping speed and not be mixed with more deliberative processing times. We also tested this hypothesis with a model comparison. The model comparison evaluated the ability of a computational model assuming dual response pathways to account for the data with models that assume a single response pathway.

We also had two specific non-exclusive hypotheses about how the rate of assessments changes both between balloon trials and within balloon trials. One property, we suggest, of non-assessed pumps is that they are learned automatic responses. Schneider and Shiffrin’s (1977) theory of automatic processes suggests that learning an automatic response takes “appreciable and consistent training” (p. 2). We also know that VTE behavior also appears to change with experience, with VTE activity occurring more often during early maze learning trials and then tail off with experience (Tolman, 1948). These properties motivate our second hypothesis:
Assessment by balloon hypothesis. If an assessment pump is a controlled, deliberative response while a non-assessed pump is a learned, automatic response, then we should see the rate of assessment pumps decrease with BART exposure. This hypothesis is consistent with the cognitive literature on automatic and controlled processes (Schneider & Shiffrin, 1977; Shiffrin & Schneider, 1977) as well as the animal learning literature on VTE behavior. We examined this hypothesis by first empirically identifying pump opportunities as assessments that were substantially longer than RTs seen in the baseline BART. Then, we calculated the rate of these assessment pumps as a function of balloon trial expecting to see the assessment rate monotonically decrease over successive balloon trials.

Finally, we expect assessment rates to change within a balloon trial. Referring back to VTE behavior, at choice points where the animal experienced greater conflict between alternatives (e.g., more difficult discrimination) the rate of VTE behavior was larger (Goss & Wischner, 1956; Tolman, 1938, 1948). Reasoning by analogy, we expect a similar pattern of results as summarized in our third hypothesis:

Assessment by pump hypothesis. The BSR model assumes that participants engage in a distance-to-target comparison in choosing between pumping and stopping. Based on the dependence of VTE behavior on conflict, we expect that assessments would be more likely to occur as choice conflict increases. According to the BSR model, choice conflict occurs as participants close in on his or her target towards the end of a sequence of pumps. To examine this hypothesis, we again used the empirically classified assessment pumps and examined their rate of occurrence over the course of a given balloon trial.
In our first study, we examine these three hypotheses in a large sample of undergraduate participants who completed 30 balloon trials of the standard BART. Critically, before completing the BART, they completed a baseline BART where they were instructed to simply pump the balloon until it explodes. Thus, the RTs from the baseline BART help estimate what pumps during the BART are closer to baseline pumping speed, indicating a more automatic pump response. In anticipation of our data, we find support for our hypothesis that during the BART two different pump responses develop: a more controlled, less frequent pump (assessment) response and a more automatic pump response. The more controlled response assessment occurs more often in early balloons and is more likely to occur within a balloon trial as participants near their stopping points. We use the results to modify the BSR creating a dual response BSR (drBSR) that explicitly models the two response pathways. Finally, we examine the clinical implications of dual response pathways in a re-analysis of Crowley et al. (2006). This paper originally reported that compared to controls, adolescents with serious conduct disorder and substance problems (a) made more pumps and (b) had longer RTs. The drBSR model provides a deeper understanding of this pattern of behavior isolating the difference in RTs to a difference in the rate of assessments with CD/SU showing a delay in developing an automatic response.

**Empirical Characterization of Multiple Response Pathways**

In this study, we investigate the evidence for multiple response pathways in the BART. This study uses data from two experiments. Both experiments sought to systematically examine RTs during the BART. For this reason, participants completed both the BART described earlier and a baseline BART. The baseline BART allowed us to estimate baseline-pumping speed and to characterize the different response process pathways that emerge. Very similar and consistent
patterns in terms of the time course of responses emerged across both studies, so we have merged the two datasets.

**Methods**

**Participants.** A total of 212 participants (104 in dataset 1 and 108 in dataset 2) were recruited from the undergraduate research participation pool from the psychology department at Michigan State University. The average age of the participants was 19.7 (SD = 2.6) with 52% females. Participants were given course credit for their participation and were also able to earn an additional monetary reward based on their performance during the experiment. This monetary reward was given to the participants as an incentive to maximize their winnings.

**BART.** The BART was programmed in E-Prime (Schneider, Eschman, & Zuccolotto, 2002). During the BART, a computerized balloon is shown on the screen. Each time participants press the ‘v’ key on the QWERTY keyboard (labeled with a “P” for pump) the balloon can inflate and if it does participants earn 10 points. However, the balloon can also explode and if it does participants lose the points earned on that balloon trial and begin a new trial. To stop and collect the points, participants press the ‘n’ key on the keyboard (labeled with an “S” for stop). Doing so ends the trial and transfers the points earned on that trial to a permanent bank. Participants are instructed to use a single index finger to press either key. After each trial, a fixation cross appears on the center of the screen for 1 second and then the next balloon trial begins.

The explosion points were set to insure that on average the optimal number of pumps to take was 64 (assuming complete knowledge on the participant’s part).\(^3\) Briefly, fifteen integers between 1 and 128 were randomly generated. These random numbers determined the explosion

\(^3\) This algorithm was developed by David McFarlane.
points for fifteen balloon trials. The explosion points for the remaining 15 balloon trials were set equal to $128 - X + 1$ where $X$ was the vector of 15 randomly generated explosion points. Then all 30 explosion points were randomly sorted across all 30 balloon trials.

**Baseline BART.** The baseline BART is a version of the BART created to measure average RTs for pumping behavior on the BART for each participant. During the baseline BART, participants were instructed to pump every balloon until it explodes. Only the pump button was operational. Participants only saw the balloon and no other information such as the amount of points earned and number of pumps made. The computerized balloon was programmed exactly as the BART (described above).

**Procedure.** Procedures for the two experiments were nearly identical. After completing the informed consent form, participants completed ten trials of the baseline BART. In both studies, participants completed thirty trials of the standard BART and thirty trials of what is called the AutoBART, which asks participants to type in their targeted pump and the computer pumps the balloon for them (Pleskac et al., 2008). The order of the standard and AutoBARTs was counter balanced. The purpose of including the AutoBART was to examine the different properties of the two tasks. We also collected measures of self-reported risk taking in the first dataset and individual difference measures in executive processes in the second. Results of these individual difference measures and cross-task comparisons were inconclusive and will not be presented here.\(^4\) The total experiment took just over an hour. For every point they earned

\(^4\) The average number of pumps taken on non-exploding balloons in the AutoBART or the adjusted score was $31.2 (SD = 16.1)$. This is significantly fewer pumps than participants took in the manual BART (see Results for Adjusted BART scores), $t(208) = 5.96, p < .0001$, and replicates similar results from Figner et al. (2009).
participants were paid 0.0002 cents. On average, participants earned an additional $2.69 ($SD = $0.88) in addition to the course credit they obtained.

**Results**

One participant did not complete the BART leaving 211 participants. Due to an experimenter error, one subject from the first dataset and 41 from the second dataset did not complete the baseline BART. For our initial analyses comparing the distribution of RTs between baseline and BART we focus on these 169 subjects. For all subsequent analyses, we use the entire dataset by using the average baseline pumping speed from the 169 subjects as an estimate for the remaining 42 subjects.

The average number of pumps made per successful balloon trial or adjusted BART score was 36.1 ($SD = 17.2$). This performance is well within the range of pumping behavior from past studies (Lejue, Aklin, Zvolensky, et al., 2003; Lejuez et al., 2002; Pleskac et al., 2008). In total, participants made on average 954 ($SD = 19$) pumps across the 30 balloon trials. Thus, participants made a high number of repeated and consistent actions laying the groundwork for the development of a more automatic pump response.

**RT analyses.** The top panel of Figure 1 displays the average RT for balloons that did not explode as a function of the pump number, which was normalized according to the number of pumps taken on that trial. The plot suggests that initially there was a noticeable delay in RTs. Looking only at the first pump, the average RT was 540 ms ($SD = 381$). Then, the RTs get quite fast, taking on average 226 ms ($SD = 124$), until the participant reaches a point where they want to stop at which point the RT between the last pump and their stop choice slows to on average 583 ms ($SD = 311$). For reference, the sold straight line indicates the average RT from the
baseline BART and the dashed line is a threshold value used later in the results. The threshold is three standard deviations above the average RT from the baseline BART.

A different view, however, emerges if individual balloon trials are examined. The bottom 3 panels of Figure 1 display the RTs of three different participants at three different points in the 30 balloon trials. They reveal the danger in averaging. On individual balloon trials, there are a series of pumps for which the RT is noticeably slower with the remaining RTs appearing to occur at a baseline pumping speed. The average RT from the baseline BART and the threshold value are the respective estimates from the individual participants.

**Cumulative distribution function comparisons.** To test if and how the RTs we observed in the BART are different from baseline pumping speed, we compared the empirical distribution of RTs from the BART to the distributions from the baseline BART in terms of their cumulative distribution functions (CDF). Figure 2 displays the empirical cumulative distributions across subjects for the two tasks. A Kolmogorov-Smirnov test (K-S test) of the two distributions reveals that they are significantly different from each other, \( D = 0.11 \) (\( p < .001 \)). At the individual level, for all 169 participants whom completed the baseline BART had significantly different distributions of RTs between the BART and baseline BART.

According to the assessment hypothesis, during the standard BART, there are two different pump types: (1) a sizable proportion of fast automatic pumps that are not different from baseline pumping and (2) a distinct set of slower pumps. Consistent with this hypothesis, the CDFs of the RTs reveal a large number of fast pumps that were at baseline pumping speed during the BART. However, near 180ms the two CDFs split revealing a small proportion of pumps that are distinctly slower than baseline pumping speed.
Ex-Gaussian comparisons. We sought to parametrically examine the difference between RTs of the standard and baseline BART. To do so, we fit an ex-Gaussian distribution to both distributions using maximum-likelihood methods (Lacouture & Cousineau, 2008; Van Zandt, 2000). The ex-Gaussian is the convolution of a Gaussian and exponential distribution and is a common distribution to model RTs (Hohle, 1965; Luce, 1986; Ratcliff, 1979; Ratcliff & Murdock, 1976). The probability density function is found with the following expression

$$g(t) = \frac{1}{\tau} \exp \left( \frac{\mu}{\tau} + \frac{s^2}{2\tau^2} \right) \Phi \left( \frac{t - \mu - \frac{s^2}{\tau}}{s} \right).$$

The parameters $\mu$ and $s$ control the mean and standard deviation of the normal distribution component of the ex-Gaussian. The free parameter $\tau$ comes from the exponential distribution and largely controls the right tail of the distribution.\(^5\) The assessment hypothesis implies the distribution of RTs from the BART should have a significantly larger $\tau$.

Before fitting the ex-Gaussian for each subject we removed any pump from the baseline BART or the BART with a RT greater than 4 standard deviations above the subject’s mean RT for that task. On average this removed 0.83% of the pumps from the baseline BART and 1.11% from the BART. We use this trimmed dataset for the remaining descriptive analyses. Table 1 lists the average maximum likelihood estimates (MLE) of ex-Gaussian distribution parameters for the baseline and BART distribution of RTs. The values show that in accord with the Assessment Hypothesis the BART had a significantly larger $\tau$ (Cohen’s $d = 0.39$). This indicates that participants had a larger proportion of slow RTs in the BART. Note there was also an indication of a small shift in the distribution as $\mu$ was also significantly different between groups.

\(^5\) The mean of the ex-Gaussian is $\mu + \tau$ and the overall variance is $s^2 + \tau^2$. The skewness is $\tau^2$ with a larger $\tau$ producing a more skewed right tail.
However, the size of this shift is quite small (Cohen’s $d = 0.07$). We interpret this set of results as further support for the assessment hypothesis that there are two distinct response types during the BART: a slow more controlled pump response (assessment) and a fast more automatic pump (non-assessment). Next we empirically classify pumps as assessments to examine the assessment by balloon and assessment by pump hypotheses.

**Empirical classification of assessments.** We created an empirical model-free method to classify observed pumps as either an assessment or a non-assessment pump. Using the observed RTs from the baseline BART, we classified a pump as a distance-to-target assessment if its RT was three standard deviations slower than the individual’s baseline pumping rate. The average RT in the baseline BART was $185.5$ ms ($SD = 72.7$) and the average within subject standard deviation was $99.6$ ms. Thus, on average, we classified a pump as an assessment if it was above $185.5 + 3 \times 99.6 = 484$ ms. This threshold is shown in the top panel of Figure 1. All classifications were done on the individual level with each subjects’ own baseline pumping speed. For the 42 participants who did not complete a baseline BART, we used the average mean and standard deviation of baseline pumping speed to classify pumps. We reran all analyses with a threshold set at two and four standard deviations, and conclusions remain the same.6

This classification revealed that only a small proportion of those pumps (13.4%) had RTs above our assessment threshold so that participants on average made 4.8 (SD = 7.4) assessments per balloon during the BART on non-exploding balloons. Two pump opportunities did stand out

6 Using the 3 standard deviation classification scheme, 5 of the 211 participants were identified as having no assessment pumps. For 1 participant every pump was classified as an assessment. These classifications appear to be driven largely by an overall mean difference in baseline pumping.
in this analysis. One such pump was the first pump opportunity. Across participants and balloon trials this pump was classified as an assessment 53% of the time. We suspect part of the reason for the slower first pump is due to two trial onset issues. We did have a fixation slide between balloon trials, but it was not self-paced resulting in some potential trial onset contamination in the first pump. A second reason is that according to the BSR, participants first identify a target pump before pumping the balloon. This process should also lead to contamination in the first pump opportunity. We review these issues further in the Discussion, but for these reasons we do not consider the first pump opportunity as a special assessment pump opportunity. The second opportunity that stood out was the pump opportunities when participants decided to stop pumping. Using our empirical method of classification the last opportunity would be classified as an assessment 57% of the time. In the model we develop later, we handle this by assuming stop decisions only come during an assessment.

**Assessment by balloon hypothesis.** We used the empirically classified assessment pumps to examine assessment rates as a function of balloon number on non-exploding balloons. Figure 3 plots the average assessment rate for each balloon number. The circles are the data, and the black line is the average predicted assessment rate derived from a modified BSR model described later. Consistent with the assessment by balloon hypothesis, assessment rates were highest on early balloons and diminished across balloon trials. A non-linear regression showed that this data was well described by a decreasing exponential function, $R^2 = .93$ and $p < .001$.

We also carried out a similar analysis at the individual level for the 205 participants with meaningful variation in assessments by using this particular classification scheme of three standard deviations above baseline RT (see Footnote 4). As a first test, we used the Goodman-Kruskal Gamma rank order correlation ($\Gamma$) to test how many individuals showed a significant
monotonic decrease in assessment rate across balloon trials. The average $\Gamma$ across subjects was $-0.30$ ($SD = .32, Mdn = -0.30$). Forty four percent of the subjects ($N = 91$) had significant negative correlations (84% in total had negative $\Gamma$ correlations), and only 3% ($N = 6$) had significant positive correlations; for 74% of the participants, a decreasing exponential function provided a significantly better fit than a linear function. Thus, at the aggregate and individual level there is support for the assessment by balloon hypothesis implying that not only is there a mix between controlled, deliberate pumps and fast, automatic pumps, but also that the latter is a learned response that increases in rate over exposure.

**Assessment by pump hypothesis.** The assessment by pump hypothesis posits that as subjects approach their targeted pump the rate of assessments should increase. Unfortunately subjects’ targeted pumps are latent in the standard BART. Consequently, we focused on balloon trials that ended in a stop and used the stop point as an estimate of the targeted pump. Because the number of pumps changes from balloon to balloon, we cannot use pump number itself to examine this hypothesis. Instead, we calculated how many non-assessed pumps were taken between each of the empirically defined assessment points. Dividing that number by the total number of pumps taken on that balloon tells us the proportion of non-assessed pumps that fall between any two assessment points. According to the assessment by pump hypothesis, the proportion of non-assessed pumps preceding each assessment should decrease over the course of the balloon trial. This implies that as participants approached their target they experienced more conflict between pumping and stopping. Figure 4 shows that the relationship between the proportion of non-assessed pumps taken and the ranked assessment point (i.e., 1st, 2nd, 3rd,…) is consistent with this prediction. However, there appears to be a small departure from this prediction when comparing the 1st and 2nd assessment points. The deviation is due to the slower
first pump discussed earlier. In general, the relationship between the proportion of non-assessed pumps preceding each assessment is well described by a J-shaped power curve, $R^2 = .90$ and $p < .001$. The dark line shown in the figure is the predicted function generated from Monte Carlo simulations of the modified BSR model run the individual level.

At the individual level, we also used a Goodman-Kruskal $\Gamma$ to investigate the negative correlation between the percentages of non-assessed pumps that fall before each assessment. Participants needed two or more assessments for the non-parametric evaluation leaving 193 out of 211 participants. The average $\Gamma$ was -.50 ($SD = .5, Mdn = -.64$). Sixty-eight percent were significantly negative (86% in total were negative), and only 3% had significant positive $\Gamma$ correlations. In terms of functional form, at the individual level 71% had within trial assessments rates that were better described with a power relationship rather than a linear relationship. Together these results support the assessment by pump hypothesis implying assessments become more likely as participants face more and more choice conflict between pumping and stopping.

**Dual Response BSR Model (drBSR)**

These descriptive results imply that the BSR assumption of a deliberate distance-to-target assessment at every pump opportunity is incorrect. Our past data implies there are several properties that the model captures including reward processing (Wallsten et al., 2005), learning (Pleskac, 2008), and the target selection process (Pleskac et al., 2008). For these reasons, we sought to keep these basic properties but modify the model to better account for the different response pathways. Figure 5 illustrates how we have modified the overall architecture of the BSR model. A first stage was added where on balloon trial $h$ and pump opportunity $i$ there is a probability $s(h, i)$ that participants engage in a distance-to-pump assessment and with probability $1 - s(h, i)$ participants make an automated pump. Based on the empirical results showing the
probability of an assessment is both a function of balloon number (Figure 3) and pump opportunity (Figure 4), we used the following logistic function

\[ s(h, i) = \frac{\exp[\phi(\frac{i - \lambda}{\bar{h}})]}{1 + \exp[\phi(\frac{i - \lambda}{\bar{h}})]}. \]  

(7)

The free parameter \( \phi \) determines how sensitive assessment rates are to the balloon trial \( h \) and pump opportunity \( i \). The free parameter \( \lambda \) indexes participants’ bias to engage in a non-assessed pump.

Another limitation of the BSR is that it is silent in terms of RTs. To address this limitation, we modeled the assessment choice as a sequential sampling process instead of the static logistic rule in Equation 5. A sequential sampling process models choice as a dynamic process where during an assessment participants sequentially sample and accumulate evidence for pumping and stopping (Luce, 1986; Ratcliff & Smith, 2004; Townsend & Ashby, 1983). We used a standard drift-diffusion process as a specific model of this process where evidence for pumping is evidence against stopping (see top branch of Figure 5) (Ratcliff & Smith, 2004; Busemeyer & Diederich, 2009).

The drift representing the strength of the evidence for pumping was a function of the distance from the targeted pump, \( d_h(i) = (G_h - i)^\alpha \) where \( G_h \) is the goal for balloon \( h \) determined with Equation 4 of the BSR model. The parameter \( \alpha \) is a scaling parameter that was set to .1 for all of the analyses. According to the drift diffusion model the probability of a pump is

\[ n_h(i) = P[Pump_h(i)] = \frac{\exp[4d_h(i)\theta / \sigma^2] - \exp[4d_h(i)(\theta - 2)/\sigma^2]}{\exp[4d_h(i)\theta / \sigma^2] - 1}. \]  

(8)

The parameter \( \sigma \) is a free parameter that works much the same way as \( \beta \) in Equation 5 indexing the degree to which decision makers consistently consider the distance to the target during deliberation. As \( \sigma \) gets smaller, decision makers more consistently consider the distance to the
targeted pump in their choice. The model adds an additional free parameter $\theta$ that is a choice threshold. If the evidence reaches $\theta$, then a decision to pump is made. If it reaches $-\theta$, then a decision to stop is made. As $\theta$ gets larger, decision makers need to collect more evidence and thus make slower decisions that more closely track the distance to the respective target. The bias parameter $z$ describes the bias at each assessment the decision maker has towards pumping or stopping. For simplicity we set $z = 0$ for no bias in the model. The probability of a stop is $P[\text{Stop}_n(i)] = 1 - P[\text{Pump}_n(i)]$. The probability density function for a given RT is then given by the finishing time PDF for a drift diffusion model,

$$f(\text{RT}|\text{Pump}_n(i)) = \frac{1}{p[\text{Pump}_n(i)]} \pi \left( \frac{2\theta}{\sigma} \right)^{-2} \exp \left[ \frac{d'(i) \cdot (\theta-z)}{\sigma^2} \right] \sum_{k=1}^{\infty} k \cdot \sin \left[ \frac{k\pi (\theta-z)}{2\theta} \right] \cdot \exp \left\{ -t \left[ \frac{d'(i)^2}{\sigma^2} + \left( \frac{k\pi\sigma}{2\theta} \right)^2 \right] \right\}$$

(9).

The function corresponding to a stop choice can be found by replacing $(\theta-z)$ with $(\theta+z)$ and $P[\text{Pump}_n(i)]$ with $P[\text{Stop}_n(i)]$. For derivations see Busemeyer and Diederich (2010). We model the RT for a non-assessed, automatic pump as a draw from an ex-Gaussian distribution (Equation 5). The $\tau$, $\mu$, and $s$ are free parameters describing the RT distribution for non-assessed pumps.

In sum, the model has 10 free parameters to describe choice and RTs from (on average) 954 (SD = 10) pumps made by participants. Table 2 provides a description of the parameters.

The drBSR was developed to account for what appears to be two distinct pump types during the BART: an automatic pump (non-assessed) and a more controlled, deliberate pump (assessed). The behavioral data supports this development. There are two related remarks to be made about the model and its process level account of a mixture of the non-assessed and assessed pumps and their associated RTs. First, the model does not rely on some a priori classification of pumps as assessments such as we used with the earlier empirical analysis. Instead, the probability of either
of the pump types is directly modeled. Second, the drBSR implies that not all fast RTs are non-assessments (and vise versa) rather there is a distribution of RTs for assessment pumps and a distribution of RTs for non-assessments. It is a mixture of these distributions that we are modeling. Next, we use the drBSR to examine the degree to which this more complex model is necessary or whether competing models that assume a single process can give a better account of the data.

**Alternative Single Process Models**

The drBSR models the observed pumps as a mixture of non-assessed and assessed pumps. This structure reveals an alternative means to test the assessment hypothesis via model comparisons with models that assume pumps are only assessments or non-assessments.

**Automatic-only model.** The automatic-only model assumes no deliberative cognitive processing on any pump opportunity. According to the model participants have a fixed propensity to pump. In other words, the automatic-only model is essentially the non-assessment branch of the drBSR model. To incorporate a propensity to pump, we adapted Wallsten et al.’s (2005) statistical baseline model so that every pump opportunity with probability $r_{auto}$ participants pump and with probability $1- r_{auto}$ they stop. The estimate of $r_{auto}$ is the relative frequency of observed pumps over all pump opportunities taken during the BART. At each pump opportunity, the observed RT (for pump or stop) is distributed as an ex-Gaussian with free parameters $\tau_{auto}$, $\mu_{auto}$, and $s_{auto}$. Thus, in total the model has 4 free parameters.

**Assessment-only model.** The assessment-only model assumes that at every pump opportunity participants engage in a deliberative distance-to-goal assessment. It is the assessment branch of the drBSR model. In other words, the model is the BSR model modified to have a drift diffusion response rule (Equation 8) rather than the logistic rule (Equation 5). At every pump
opportunity participants engage in a sequential sampling process where they deliberate over the
distance to the goal by collecting evidence until a threshold is reached at which point a decision
to pump or stop is made accordingly. The model has five free parameters in total: two learning
parameters ($q_1$ and $\delta$), a reward sensitivity parameter $\gamma$, deliberation variability $\sigma$, and a choice
threshold ($\theta$).

**Model Comparisons**

We fit all three models to the choice and RT data using maximum likelihood methods. The Appendix describes the more general likelihood function for the drBSR model and a
description of the fitting routine. The likelihood functions for the single process models can be
derived from this more general function. We compared the models at the individual level with
the Bayesian Information Criterion (BIC; Raftery, 1995; Wagenmakers, 2007),

$$BIC = -2 \cdot LL + k \cdot \ln(n),$$

which penalizes models with more parameters. Models with smaller BIC’s are preferred. BIC
differences of 10 or more are taken as very strong evidence for the corresponding model.

Across all three models, the drBSR had the lowest BIC score for 160 of the 211
participants (76%). The automatic-only model was never the best model for any of the
participants. The average difference between BICs ($\text{BIC}_{\text{automatic-only}} - \text{BIC}_{\text{drBSR}}$) was 421 ($Mdn =
382.8, SD = 309.7$). Thus, the data suggest a model that assumes some level of deliberative
cognitive processing.

The data for the 51 participants that were not best fit by the drBSR were, by elimination,
best captured by the assessment-only model. Across all 211 participants, the average difference
between BICs ($\text{BIC}_{\text{assessment-only}} - \text{BIC}_{\text{drBSR}}$) was -18,765 ($SD = 39,189.6$). The median value,
however, was 251.8. In other words, for 51 individuals the BIC scores pointed towards the
assessment-only model, but for a majority of participants the BIC scores pointed to the drBSR model.

Thus, the model comparison results provide further support for the assessment hypothesis. Moreover, they point to the drBSR as providing the best fit for a majority of participants. There are some individual differences in that the data for some individuals was best fit by the assessment-only model. This result is consistent with our earlier analyses using empirically classified assessment pumps, which identified some individual variability in the accordance with the assessment, assessment by balloon, and assessment by pump hypothesis.

In sum, we conclude that the drBSR is the more useful model. It provides a better description of the data from a majority of the participants. It can also capture data from participants who use a single response process (automatic or assessment-only). This is done with the bias parameter \( \lambda \) in Equation 7 where large positive values bias the participant to use the automatic branch and large negative values bias the participant to use the assessment branch.

The drBSR does a good job recreating the data. Recall the average adjusted BART score was 31.3 (\( SD = 13.2 \)). The average predicted adjusted BART score from the best fitting model parameters was 31.1 (\( SD = 14.6 \)) (see Pleskac, 2008 for the analytic solution for estimating adjusted BART scores from the model). The correlation between the predicted and observed adjusted BART scores was \( r = .92 \). To get a sense of how well the full model recreates the RTs, we simulated the model with 100 runs of each of the 211 subjects’ best fitting parameters playing the BART. The model was successful in recreating the observed RTs; the average RT across subjects and balloon trials was 240.6 ms (\( SD = 122.2 \)), while the average simulated RTs from the best fitting model parameters was 241.4 (\( SD = 131.1 \)). The correlation between the predicted and observed RTs was \( r = .99 \).
Figures 3 shows the drBSR model gives a good account of the changes in assessment rates across balloon trials. The prediction from the model was estimated from Equation 6 using the MLE parameter estimates and averaging across pump opportunities and participants. The drBSR also gives a good account of the change in non-assessed pumps between assessments shown in Figure 4. This last function was estimated from Monte Carlo simulations of each participants’ MLE parameter estimates playing the BART via the drBSR and then classifying each pump as an assessment or not using the same empirical classification method.

Discussion

Many risks we take are repeated and sequential in nature. Psychologically we know this property can promote the use of multiple response pathways. In this first study, we found support for multiple response pathway during the BART where one pathway appears to be a slower, more deliberate pump response and the other a faster automatic pump response. The rate of the slower pump responses decreases with task exposure suggesting the automatic pump response is a learned response (Schneider & Shiffrin, 1977; Shiffrin & Schneider, 1977). The slower more deliberate responses also occur more often towards the end of a balloon trial. These results

To calculate the average we weighted the assessment rate by the probability of the decision maker making it to pump $i$ on balloon $h$.

Consistent with the hypothesis that the automatic response develops with task exposure we also found that average assessment rate depended on whether participants completed the automatic BART first. Participants who had exposure to the BART via the AutoBART first in our studies had a lower percentage of assessments 11.6% ($SD = 13.9$) versus 19.6% ($SD = 21.3$), $t(209) = 3.21, p < .01$. This difference did not impact the general pattern of assessment rates between and within balloon trials.
suggest that these slower responses are a point in time where participants are planning their next steps, much like VTE behavior in early animal learning studies (Tolman, 1938, 1948).

This joint pattern of a larger assessment rate during early balloon trials and an increasing rate toward the end of a given trial also speaks against alternative explanations for the slower pump responses such as fatigue or disengagement due to boredom from the task. These alternative explanations would predict more frequent slow pumps with increasing balloon trials; however, the opposite pattern was observed. Moreover, the fact that the slower pumps appear to happen in a systematic fashion within a balloon trial as participants approach their stopping point also speaks against these alternative explanations. At the same time, the joint occurrence also speaks against a practice effect. While this explanation would be consistent with a decrease in the rate of slower pump responses across the balloon trials, again it would not explain why these assessments appear to occur more frequently at the end of a given balloon trial.

Taken together, these results imply a single response process pathway account is insufficient to describe behavior during the BART (Pleskac, 2008; Wallsten et al., 2005). We summarized these results with the drBSR model. The drBSR model (a) accommodates the dual response pathways (Figure 5) and (b) gives a process level account of both choice and RTs. We should note that this model is only an approximation of the response pathways during the BART. For example, despite the empirical data suggesting that the first pump tends to be delayed we chose not to model this pump as a special assessment pump. One reason is that there are contaminants that are potentially in these observed RTs including target pump estimation as well as trial onset issues. Future work should address this perhaps by using a self-paced inter-trial break. We could have also used a different sequential sampling model to model the choice process in the assessment pathway such as the linear ballistic accumulator (Brown & Heathcote,
2008) or the Poisson accumulator model (Townsend & Ashby, 1983). We suspect, though, they would produce very similar results. We discuss the broader implications of the empirical results and model in the General Discussion. Next we re-analyze existing data to examine the clinical utility of modeling decision making during the BART with multiple response pathways.

**Re-Analysis of Crowley et al. (2006)**

The BART and BART-like tasks have been used as an assessment tool to study risk taking by clinical populations in a controlled laboratory setting (Aklin et al., 2005; Bornovalova et al., 2005; Hoffrage et al., 2003; Hopko et al., 2006; Lejuez et al., 2005; Lejuez, Aklin, Jones, et al., 2003; Lejuez, Aklin, Zvolensky, et al., 2003; Lejuez et al., 2004; Pleskac, 2008). The tasks allows clinicians to investigate whether clinical populations who engage in risky behaviors do so in new situations without prior learning, peer pressure, intoxication, etc. They also allow a more controlled study of risky behavior. Crowley et al. (2006) used the BART for exactly this reason to study risk taking among adolescent patients who were currently in treatment for conduct disorder (CD) and substance use (SU; abuse or dependence). CD is marked by anti-social behaviors such as aggression and violation of the rights of others, often in an opportunistic or impulsive fashion (American Psychiatric Association, 2000). This set of behaviors led Crowley et al. to investigate whether the BART could measure an initial propensity among these individuals to take risks.

Crowley et al.’s (2006) results are summarized in Figure 6. They show that CD/SU adolescents took more pumps than controls, \( t(38) = 4.17, p < .001, d = 1.32 \). They were also on average significantly slower, \( t(38) = 2.18, p < .05, d = 0.69 \) (see Table 3 for summary statistics).\(^9,10,11\) If we take a single process perspective with the BART, then these results seem

---

\(^9\) We thank Dr. Thomas Crowley for providing us with the data from Crowley et al. (2006).
puzzling. Intuitively we would expect that if the primary difference between CD/SU and controls rests in terms of impulsivity (Moeller, Barratt, Dougherty, Schmitz, & Swann, 2001), then they would show greater risk taking and faster RTs. However, the drBSR tells us that the BART is not process pure: Multiple cognitive processes are involved in making decisions during the BART. Thus, we were curious if the drBSR could help reconcile the differences in risky choices, on the one hand, and RTs, on the other hand.

The drBSR suggests that one way the differences in the RTs could arise is if CD/SU engaged in more assessments. That is, the slower average RTs might not be indicative of slower pumping in general, but an increase in the proportion of a smaller subset of slow RTs. We know individuals with CD also have impairments in response inhibition (Nigg, 2000; Oosterlaan, Logan, & Sergeant, 1998). One aspect of this impairment is an inability inhibit interferences (Barkley, 1997), which should impact the development of an automatic response. The drBSR can also be used to better understand Crowley et al.’s (2006) suggested that adolescents with CD/SU differ in “initial propensity or ‘set,’ to take risks” (p. 176) and this is what the BART is measuring via the adjusted BART scores. This would suggest CD/SU should differ in their initial belief parameter $\theta_1$. Finally, on a generalizability note, we were curious how the drBSR and the multiple response pathway account would generalize to a different implementation of the BART and a younger age group.

Methods

10 The same conclusion is reached with the log transform of the response times.

11 Note Crowley et al. (2006) only analyze the first 10 inter-pump times. Figure 6 is based on the average RTs across all pumps excluding the initial pump response on each balloon (see Methods).
Participants. It is useful to briefly describe the sample (for more details see Crowley et al., 2006). The \( N = 20 \) CD/SU adolescents were recruited from a day-treatment program for youths with serious conduct and substance use problems. The average age of the participants was 16.1 (\( SD = 1.1 \)), 17 of whom were males. Participants needed to be between the ages of 14 to 18, have an IQ \( \geq 80 \), have serious antisocial problems including at least some CD symptoms, substance use or dependence on one or more non-nicotine substance, and test negative for at least seven days for the nine-most common substances. The \( N = 20 \) controls were recruited from areas via zip codes where the patients from the treatment program had come. The controls were required to have similar age and IQ ranges, with no history of substance-related problems, no prior arrests or convictions, and no positive test for common substances one week before testing.

Materials. Both controls and patients completed the Lejuez et al.’s (2002) version of the BART. This version is roughly identical to ours, but differs in three aspects relevant to our analysis. First, participants entered responses using a mouse click on a computerized pump button and stop button. Thus, we were able to investigate the degree to which our findings from our version of the BART that used keyboard button responses generalize to other response entry methods. Second, there was no delay between balloon trials, so that once the balloon exploded or the bank was updated on stop trials, the next balloon trial began. As a result, there is likely an even greater rate of RT contamination on the first pump. Supporting this conclusion, nearly all first pumps in this dataset would be classified as an assessment using our empirical classification method. For this reason in our analyses we do not use the first observed RT. Third and finally, the Lejuez et al. (2002) BART uses the same set of explosion points across participants. The optimal solution given knowledge of the stochastic structure of the task for the pump sequence is still to pump 64 times on every balloon; however, the fixed explosion points introduce small
systematic deviations into balloon trial data because on some balloons nearly everyone exploded the balloon. These deviations become especially apparent in the between-balloon-trial analyses.

Results

Our re-analysis is in two parts. In the first part of the analysis, we compare the empirical distribution of RTs for the CD/SU adolescents to the controls to examine if the latter were overall slower or if their RT distributions showed the signature pattern of different rates of assessments established in the first study. In the second part of the analysis, we use the dsBSR to quantitatively describe the differences between the two groups during the BART.

RT distribution analyses. Figure 7 displays the empirical cumulative distribution functions for each group. A K-S test of the two distributions reveals they are significantly different from each other, \( D = 0.23, p < .001 \). The CDFs reveal that, unlike the conclusions drawn from comparing mean RTs, the two groups showed a similar magnitude and proportion of fast RTs. However, there is a noticeable break where the CD/SU adolescents appear to have a larger proportion of longer RTs. Table 4 summarizes the MLE parameter estimates from fits of the ex-Gaussian distribution to the RTs from each individual.\(^{12}\) They show that CD/SU participants had a significantly larger \( \tau \) implying a longer right tail in the RT distributions for these participants. These results support our hypothesis that CD/SU adolescents were not slower across pump opportunities, but instead consistent with taking more assessments just slower on a small proportion of trials.

\(^{12}\) As in the first study, we removed any RT that was 4 standard deviations above the subject’s mean RT. This removed on average 1.1% of pumps.
the BART. We fit the model with maximum likelihood methods using the entire set of observed pumps (see Appendix). The model reproduces the data well. Figure 6 shows a comparison between the observed and model predictions for the adjusted BART score and the RTs. The parameter estimates are given in Table 5.

Consistent with the differences in the RT distributions, the dRBSR reveals differences in the assessment rates between the CD/SU adolescents and controls. CD/SU participants had a significantly lower bias ($\lambda$) indicating a greater likelihood to engage in an assessment. They also had a higher sensitivity parameter ($\phi$) indicating their assessment rates were more sensitive to balloon trial and pump opportunities. To better understand this pattern of parametric differences, we empirically classified assessment pumps using the threshold of 484 ms from the the undergraduate sample in the first study. Figure 8 plots the change in assessment rate between balloon trials as well as the model fits. It shows that CD/SU adolescents had a higher initial rate of assessments (as the $\lambda$ comparison implies) and showed a larger change with balloon number in the assessment rate (as the $\phi$ comparison suggests).

One concern that arises from examining Figure 8 is that the difference between the CD/SU and controls may be driven more by the controls engaging in a relatively low constant rate of assessments. To test this, we compared the parameter estimates between the CD/SU patients and our undergraduate sample (Table 2). The CD/SU sample had a significantly lower bias to engage in a non-assessed pump ($\lambda$) ($u = -3.04, p < .01$) and their assessment rate was more sensitive to task exposure ($\phi$) ($u = 3.13, p < .01$). These differences are evident by comparing the assessment rates for each balloon trial from the undergraduate study, which are
replotted on the right of Figure 8. Consistent with the parameter estimates, the assessment rates for the CD/SU adolescents are above the undergraduate sample.\textsuperscript{13}

We also examined the degree to which assessment rates changed within a balloon trial (assessment by pump hypothesis). Figure 9 illustrates that both groups exhibited a similar pattern with a greater rate of assessments as they approached the point where they stopped. Although the model predictions imply a difference between the two groups in the within-trial assessment rates, there was not a reliable difference in the empirically classified data.

The drBSR also provides a better understanding for the differences in the adjusted BART scores (Figure 6). All else being equal, if individuals have a greater assessment rate (like the CD/SU adolescents), then they would be more likely to stop pumping because they enter the assessment process more often and thus have on average lower adjusted BART scores. However, not all else remained equal. The parameter estimates in Table 5 reveal a marginal difference between CD/SU adolescents and controls in their initial belief that the balloon would not explode on any given pump during the BART ($\hat{q}_1$). The CD/SU adolescents also showed significantly smaller deliberation variability ($\sigma$). Both of these differences are sufficient to produce the observed differences in the adjusted BART scores and provide a deeper understanding as to why the CD/SU adolescents appear to be risk seeking.

According to the drBSR the difference in the initial beliefs led CD/SU adolescents to set larger initial target pumps (Equation 3). Using the MLE parameter estimates, the median initial target for CD/SU adolescents was 59 ($M = 76, SD = 72$) compared to 18 ($M = 30.3, SD = 36.4$)\textsuperscript{13}

\textsuperscript{13} Compared to the controls from Crowlet et al. (2006), the undergraduate sample also had a significantly lower bias to engage in a non-assessed pump ($\lambda$) ($u = -2.54, p = .01$) and their assessment rate was more sensitive to task exposure than the controls ($\phi$) ($u = 3.64, p < .001$).
MAKING ASSESSMENTS WHILE TAKING REPEATED RISKS

for controls ($U = 3.15$, $p < .01$). These initial beliefs anchored subjects in terms of their experiences and their subsequent updated beliefs. So that across all balloons, CD/SU adolescents typically set far larger targets across all balloons ($M = 58.5$, $Mdn = 59.9$, $SD = 27.0$) than the controls ($M = 28.1$, $Mdn = 23.2$, $SD = 21.8$) ($u = 3.39$, $p < .001$).

To further test the role this difference in initial beliefs played, we simulated the drBSR model using the parameter estimates for the CD/SU adolescents. However, we substituted their initial beliefs about the likelihood the balloon would not explode with the average of the beliefs of the controls ($q_1 = .93$). As expected, the substitution was sufficient to drop the predicted adjusted BART scores for the CD/SU to approximately the level of the controls ($M = 26.0$, $Mdn = 26.1$, $SD = 8.3$) (see Table 3 for the summary statistics of the observed data). If we do the same thing and simulate the drBSR model using the parameter estimates of the controls, but substitute their prior belief with the average estimate for CD/SU adolescents ($q_1 = .98$) the predicted adjusted BART score also increases but not quite to the level of the controls ($M = 29.8$, $Mdn = 29.9$, $SD = 8.6$).

Differences in deliberation variability can also give rise to the differences in the adjusted BART scores. The reason why is best understood by imagining two participants with the same targeted pump, but different deliberation consistency $\sigma$. For the participants with smaller $\sigma$, the target will have greater impact in the evidence accumulation process and thus before reaching the target will be less likely to choose to stop when making an assessment. To test out this explanation, we again simulated the drBSR model using the parameter estimates of the CD/SU

\[ 14 \text{ To do this we adjusted the } a_1 \text{ parameter to match the mean. Technically, this parameter substitution also impacts the initial uncertainty } \delta \text{ as the mean and variance of the beta distribution are not independent, but our calculations show the impact was minimal.} \]
adolescents, but this time substitute their deliberation consistency ($\sigma$) values with the mean value of the controls (Table 4). Doing so drops the predicted adjusted BART scores of the CD/SU adolescents to the level of controls ($M = 23.6$, $\text{Mdn} = 22.7$, $SD = 10.1$). Analogously, setting the deliberation consistency parameter for the controls equal to the mean value from the CD/SU adolescents increases the predicted adjusted BART scores rise up, but not to the level of the CD/SU adolescents ($M = 29.0$, $\text{Mdn} = 27.9$, $SD = 10.2$).

**Discussion**

Our re-analysis of Crowley et al. (2006) demonstrates that the assessed and non-assessed dual response pathway process generalizes to different variation of the BART and different age groups. The drBSR also provides a deeper cognitive understanding of the differences between CD/SU adolescents and their matched controls.

Crowley et al. (2006) originally reported that on average CD/SU adolescents were slower in pumping during the BART, but did not offer an explanation as to why. If we only focus on the mean RT, then this result would indeed seem inconsistent with the idea that CD/SU are more impulsive. The drBSR, however, suggests the mean difference is the result of CD/SU adolescents engaging in a higher rate of assessments during early balloon trials suggesting they were slower to develop the automatic pump response. This result is consistent with CD/SU adolescents having impairments in response inhibition resulting in an inability to inhibit interference from other aspects of the task like the potential to stop and collect a reward and thus delaying the development of an automatic response (Barkley, 1977; Nigg, 2000; Oosterlaan et al., 1998). This result certainly identifies an intriguing area for future research suggesting cognitive control can play a critical role in the development of more automatic response processes during decision making. It also suggests that these repeated decision making tasks may be well suited for
Making Assessments While Taking Repeated Risks

studying not just multiple response pathways, but the development of the response pathways. We will return to this in the General Discussion.

The drBSR also identifies two logically distinct cognitive sources for the observed difference in the adjusted BART scores between CD/SU and controls: (a) a tendency to set more optimistic initial beliefs about rewards rather than a punishment ($\hat{q}_1$); and (b) to more consistently deliberate about the targeted reward ($\alpha$). The difference in the initial beliefs is consistent with Crowley et al.’s (2006) suggestion that the BART reveals an initial propensity to engage in risky behavior. The drBSR goes beyond the behavioral analysis isolating the cognitive mechanism underlying this propensity and explain how it results in a trajectory for risk taking throughout the task. The difference in the consistency in which the two groups deliberated about the target with the CD/SU showing more consistent deliberation about the target was unexpected. These individuals are characterized as having more dominant reward systems (Quay, 1993; Shapiro, Quay, Hogan, & Schwartz, 1988), which may, in turn, lead to a greater focus on potential rewards.

General Discussion

An impasse has emerged in modeling decision processes. On the one hand, multiple response processes are often suggested to describe a variety of decisions (e.g., Daw et al., 2005; Frank & Claus, 2006; Frank et al., 2009; Kahneman, 2003; Sloman, 1996). On the other hand, most formal models of decision making posit a single response process (Birnbaum, 1999; Brandstatter et al., 2006; Busemeyer & Stout, 2002; Busemeyer & Townsend, 1993; Tversky & Kahneman, 1992; Wallsten et al., 2005).\textsuperscript{15} Using the BART as a test case of repeated decision processes that the decision maker engaged in before evaluating the prospects in terms of

\textsuperscript{15} The original derivation of prospect theory also contained editing and dominance detection processes that the decision maker engaged in before evaluating the prospects in terms of
making, we found evidence suggesting that participants developed two different response pathways over the course of the task. One pathway is a slower, controlled, attention demanding, capacity limited process where participants make a distance-to-target assessment to choose between pumping and stopping. The other response is a fast, automatic process that develops with experience in the task. Consequently, slower assessment responses occur more often during balloon trials and then diminish at an increasing rate. Moreover, we showed that much like VTEs from the animal learning literature (Tolman, 1938, 1948), these slower more deliberate responses occur at an increasing rate as decision makers approach their stopping point for a given balloon suggesting the slower distance-to-target assessment are engaged as choice conflict increases.

Finally, we showed that understanding the different impacts of controlled and automatic response pathway can bring a deeper understanding to decision deficits observed in CD/SU adolescents.

**Dual Response Pathway BSR (drBSR)**

In light of this apparent mix of fast, automatic pumps and slow, more deliberate pumps, we modified the single response pathway BSR model (Pleskac, 2008; Wallsten et al., 2005) to better account for the different response pathways (see Figure 5). The drBSR maintains the primary assumptions of the BSR model. At the onset of each balloon, decision makers evaluate the possible payoffs associated with each pump option and the probability of successfully reaching that pump (Equation 3). From these evaluations a targeted pump is selected that maximizes the expected payoffs (Equation 4). However, instead of engaging in distance-to-target assessment at each pump, the drBSR has a second pathway so that only on some pump expected subjective payoffs, so it may also be understood as a multiple response pathway theory (Kahneman & Tversky, 1979).
opportunities do decision makers engage in a slow, more deliberate distance-to-target assessment. This assessment process is modeled with a drift diffusion process where participants accumulate evidence regarding their distance from the targeted pump to form a preference to pump or not. If the preference reaches an upper boundary, then decision-makers pump the balloon; otherwise they stop. The further decision makers are from the target, the faster preference accumulates toward pumping. On the remaining pump opportunities, decision-makers engage in a fast more automatic pump and the RT for that pump is drawn from an ex-Gaussian distribution. According to the model—and as the data suggests—the different response pathways are used at different points in the task. The slower assessment pathway occurs more often during initial exposure and is more likely to occur as conflict increases. The faster more automatic pump develops with experience in the task. This more complete cognitive model not only accounts for the choice and assessment process, but also the distribution of RTs in the BART

**Multiple Response Pathways**

Often two-system accounts of decision making have focused largely on showing how respondents use different response pathways in one-shot decisions between conditions or between tasks (Evans, 2008; Kahneman, 2003; Sloman, 1996; Stanovich, 1999). We have shown that in making repeated risky choices a mixture of slow more deliberate responses and fast more automatic responses develops and modeled this with the drBSR. An alternative modeling framework might use a multi-system reinforcement learning architecture where two systems coexist and simultaneously drive decision making (Daw et al., 2005). This model would have a habitual-like system that associates expectancies with each alternative and a goal directed system that learns contingencies between the outcomes and events associated with each alternative. One
could imagine the habitual-like system is well aligned with the non-assessed pump, while the goal directed system is aligned with the distance-to-target assessment. A limitation of this multi-system reinforcement learning architecture is that the framework does not make RT predictions and as a result relies heavily on choice patterns to determine when decision makers are relying on one system or the other to make a choice (though see Keramati, Dezfooli, & Piray, 2011). The multi-system reinforcement learning architecture does, however, appear in some cases to map well onto different neural systems that guide decision making (Daw et al., 2005; Glascher et al., 2010). Future work may benefit from merging these two multi-system/multiple response pathway accounts of decision making to give a more complete picture of the neural and computational processes.

The drBSR illustrates one way RTs can be modeled during sequential risk taking tasks. Our work echoes Bower’s (1959) use of a random walk model—one of the earliest applications in psychology using random walk models to model choice and RTs—to describe VTE behavior of rats performing a standard T-Maze (see also Chapter 4 in Atkinson, Bower, & Crothers, 1965). According to Bower’s model, when rats reach a choice point in a T-Maze, they orient towards one branch of the maze and infer the possible outcomes via visual, auditory, olfactory, or spatial cues. Then, with some probability, the subjects approach the stimulus; otherwise they turn and considers the other alternative in the same manner. This process continues until rats approach the stimulus.

The assessment branch of the drBSR operates in much the same way, but in the model the cue is an internal, cognitive factor (e.g., the pump target). It is certainly the case that much like a rat in the maze other factors in the task environment impact the deliberation process. This raises the intriguing question of how the task environment impacts risk taking and if its impact
depends on whether decision-makers are in the slow more deliberate assessment pathway or the faster, more automatic, non-assessed pathway. To the degree that the BART models real world risk taking situations, answering this question may help us better understand why people take repeated risks and may, in fact, suggest ways to slow them down and make them think about the risks they are taking.

The drBSR also goes beyond the simple random walk model of VTE behavior in that it models the development of a second more automatic response. We expect that this property of multiple response pathways is present in the larger class of sequential risk taking tasks of which BART is one specific instance (Figner et al., 2009; Pleskac, 2008; Slovic, 1966). The degree to which multiple response pathways develop in other repeated risky decision tasks like the Iowa Gambling Task (Bechara et al., 1994). Certainly Damasio’s (1994) somatic marker hypothesis suggests such a development.

The drBSR can serve as one possible blueprint for modifying cognitive models of decision making (e.g., Busemeyer & Townsend, 1993; Ratcliff & Smith, 2004; Usher & McClelland, 2004) to include multiple response pathways. In human decision making, extremely fast or extremely slow responses are often either discarded or modeled directly as contaminant trials (e.g., Ratcliff & Tuerlinckx, 2002). Often these contaminants are quite rare given the design of the decision-making task. We have, however, shown that during sequential risky

---

16 We have subjected the data from Pleskac’s (2008) sequential risk taking tasks to a similar set of analyses and found a similar pattern of data in terms of assessments and assessments changing with balloon and pump. One issue with these data is that Pleskac had a 200 ms fixed delay between decisions making obscuring the RTs a bit for the analyses.
decision making where choices are made repeatedly over very similar options, a mix of slower more deliberate and faster more automatic responses occurs.

**Clinical Utility of Modeling Multiple Response Pathways**

Our results imply that during the standard response procedure used in the BART as well as other similar sequential risk taking tasks (Figner et al., 2009; Pleskac, 2008; Slovic, 1966) an automatic response can develop. Understanding the properties of this process is important. Although initially these sequential risk taking tasks might have described as simple measures risk attitudes (e.g., Hoffrage et al., 2003; Lejuez et al., 2002; Slovic, 1966), the drBSR and the earlier BSR illustrate that these tasks are better conceptualized as behavioral tasks that isolate critical cognitive processes used during risky decision making (Pleskac, 2008). Such a view can help explain how multiple cognitive processes interact to produce what appears to be an inconsistent pattern of data like simultaneously taking a higher number of risk and appearing to do it slowly as is the case with the CD/SU adolescents from Crowley et al. (2006). It is difficult to account for these results with a single construct like impulsivity. Yet, the drBSR offers a different view attributing the differences in observed risk taking to the differences in the assessment process and differences in the RTs to the development of an automatic response process. This interpretation also identifies new more basic-level questions like the role of cognitive control in developing these automated decision processes like those that appear during the BART, but more broadly in other areas of judgment and decision making (Evans, 2008; Frank et al., 2009; Kahneman, 2003; Mukerjhee, 2010; Reyna, 2004; Sloman, 1996).

**Conclusion**

In this paper, we examined how multiple response pathways develop as decision makers make repeated risky decisions during a sequential risk taking task. One pathway is a slower,
more controlled pathway where participants deliberate over taking a risk or stopping and collecting their reward. The second pathway is a learned more automatic process where little deliberation occurs. This automatic process appears to be a learned response such that slower more deliberate distance-to-target assessments occurs early in task exposure. An assessment is also more likely to occur as choice conflict between the safe and risky option increases. The drBSR modifies the existing single response pathway model to account for these dual response pathways and gives a good account of not only the choice and assessment behavior, but also RTs. Finally, we showed that understanding the different impacts of controlled and automatic response pathways has clinical importance in that the drBSR model distills performance differences with conduct disorder/substance using adolescents into two different processing components: how they evaluate risks and the degree to which they enter the slower-more deliberate assessment response pathway.
References


MAKING ASSESSMENTS WHILE TAKING REPEATED RISKS

Experimental and Clinical Psychopharmacology, 13(4), 311-318. doi: 10.1037/1064-1297.13.4.311


Luce, R. D. (1986). *Response times: Their role in inferring elementary mental organization.* New York, NY: Oxford University Press.


MAKING ASSESSMENTS WHILE TAKING REPEATED RISKS


Table 1

*Mean, median and SD of ex-Gaussian parameters for the baseline and standard BART reaction times.*

<table>
<thead>
<tr>
<th></th>
<th>Baseline BART</th>
<th></th>
<th>BART</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>M</td>
<td>Mdn</td>
<td>SD</td>
<td>M</td>
<td>Mdn</td>
</tr>
<tr>
<td>µ</td>
<td>0.137</td>
<td>0.13</td>
<td>0.046</td>
<td>0.133</td>
<td>0.12</td>
</tr>
<tr>
<td>S</td>
<td>0.022</td>
<td>0.016</td>
<td>0.026</td>
<td>0.019</td>
<td>0.016</td>
</tr>
<tr>
<td>τ*</td>
<td>0.049</td>
<td>0.03</td>
<td>0.05</td>
<td>0.097</td>
<td>0.06</td>
</tr>
</tbody>
</table>

# p < .05 with Wilcoxon sign rank test on differences

* p < .05 with a repeated measures t-test.
Table 2

*Descriptions, means, medians and standard deviations of drBSR parameters.*

<table>
<thead>
<tr>
<th>Description</th>
<th>$M$</th>
<th>$Mdn$</th>
<th>$SD$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$q_1$</td>
<td>.94</td>
<td>.98</td>
<td>.12</td>
</tr>
<tr>
<td>index of initial belief that the balloon will not explode. Larger values of $q_0$ lead to larger targeted pumps, larger drift rates in the drift diffusion process, and thus more pumps being made on average.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>log($\delta$)</td>
<td>-9.62</td>
<td>-9.76</td>
<td>2.55</td>
</tr>
<tr>
<td>measure of uncertainty participants have in their initial belief about the likelihood of the balloon exploding. Larger values indicate more uncertainty and thus more sensitivity to observed pump data.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\gamma$</td>
<td>.74</td>
<td>0.69</td>
<td>0.32</td>
</tr>
<tr>
<td>measure of reward sensitivity. Higher values lead to larger targeted pumps, larger drift rates in the drift diffusion process, and thus more pumps being made on average.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\sigma$</td>
<td>0.90</td>
<td>0.80</td>
<td>0.71</td>
</tr>
<tr>
<td>deliberation consistency and as $\sigma \rightarrow 0$ choice behavior becomes more deterministic.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\alpha$</td>
<td>0.75</td>
<td>1.23</td>
<td>0.49</td>
</tr>
<tr>
<td>scales the distance from the goal into a drift. Set to .1 for all analyses.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\theta$</td>
<td>0.75</td>
<td>1.23</td>
<td>0.49</td>
</tr>
<tr>
<td>the choice threshold in the assessment process. It measures the quantity evidence a decision maker must collect in order to make a decision between pumping and stopping. As $\theta$ gets larger decision makers need to collect more evidence and thus make slower decisions, but also more deterministic decisions.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$z$</td>
<td>0.75</td>
<td>1.23</td>
<td>0.49</td>
</tr>
<tr>
<td>the bias parameter in the assessment process. It describes the bias at each assessment decision makers have towards pumping or stopping. We set this at 0 (no bias) for all analyses.</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
A sensitivity parameter controlling how sensitive decision makers’ assessment rates are to experience in the task. Lower values indicate less sensitivity to the pump opportunity and balloon trial.

A measure of the decision maker’s bias to engage in a non-assessed pump. Higher values indicate a larger bias to engage in a non-assessment.

The mean of the Gaussian component of the Ex-Gaussian distribution used to model the automatic, non-assessed pump RTs. It largely determines the central tendency of the ex-Gaussian.

The standard deviation of the Gaussian component of the Ex-Gaussian distribution used to model the automatic, non-assessed pump RTs. It largely determines the variability of the ex-Gaussian.

The parameter of the exponential component of the Ex-Gaussian distribution used to model the automatic, non-assessed pump RTs. It largely determines the positive skew of the distribution.

<table>
<thead>
<tr>
<th>Symbol</th>
<th>Description</th>
<th>Value 1</th>
<th>Value 2</th>
<th>Value 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\phi$</td>
<td>A sensitivity parameter controlling how sensitive decision makers’ assessment rates are to experience in the task. Lower values indicate less sensitivity to the pump opportunity and balloon trial.</td>
<td>0.21</td>
<td>0.08</td>
<td>0.32</td>
</tr>
<tr>
<td>$\lambda$</td>
<td>A measure of the decision maker’s bias to engage in a non-assessed pump. Higher values indicate a larger bias to engage in a non-assessment.</td>
<td>25.78</td>
<td>24.17</td>
<td>15.52</td>
</tr>
<tr>
<td>$\mu$</td>
<td>The mean of the Gaussian component of the Ex-Gaussian distribution used to model the automatic, non-assessed pump RTs. It largely determines the central tendency of the ex-Gaussian.</td>
<td>0.15</td>
<td>0.14</td>
<td>0.05</td>
</tr>
<tr>
<td>$s$</td>
<td>The standard deviation of the Gaussian component of the Ex-Gaussian distribution used to model the automatic, non-assessed pump RTs. It largely determines the variability of the ex-Gaussian.</td>
<td>0.02</td>
<td>0.02</td>
<td>0.01</td>
</tr>
<tr>
<td>$\tau$</td>
<td>The parameter of the exponential component of the Ex-Gaussian distribution used to model the automatic, non-assessed pump RTs. It largely determines the positive skew of the distribution.</td>
<td>0.04</td>
<td>0.02</td>
<td>0.06</td>
</tr>
</tbody>
</table>
Table 3

Mean, median, and standard deviations of the adjusted BART scores and response times from Crowley et al. (2005).

<table>
<thead>
<tr>
<th></th>
<th>Controls</th>
<th>CD/SU</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$M$</td>
<td>$Mdn$</td>
</tr>
<tr>
<td>Adjusted BART score</td>
<td>24.0</td>
<td>23.6</td>
</tr>
<tr>
<td>Response Time (ms)</td>
<td>242</td>
<td>231</td>
</tr>
</tbody>
</table>
Table 4

Mean, median, SD of ex-Gaussian parameters fit to the RTs for the controls and CD/SU patients in Crowley et al. (2006).

<table>
<thead>
<tr>
<th></th>
<th>Controls</th>
<th></th>
<th></th>
<th>CD/SU</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$M$</td>
<td>$Mdn$</td>
<td>$SD$</td>
<td>$M$</td>
<td>$Mdn$</td>
<td>$SD$</td>
</tr>
<tr>
<td>$\mu$</td>
<td>0.16</td>
<td>0.18</td>
<td>0.04</td>
<td>0.17</td>
<td>0.15</td>
<td>0.09</td>
</tr>
<tr>
<td>$s$</td>
<td>0.02</td>
<td>0.02</td>
<td>0.01</td>
<td>0.03</td>
<td>0.02</td>
<td>0.02</td>
</tr>
<tr>
<td>$\tau^*$</td>
<td>0.08</td>
<td>0.07</td>
<td>0.07</td>
<td>0.17</td>
<td>0.12</td>
<td>0.11</td>
</tr>
</tbody>
</table>

* = $p < .05$ with Mann-Whitney U test
### Table 5

*drBSR MLE Parameters for Crowley et al. (2006)*

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Controls (M, Mdn, SD)</th>
<th>Conduct Disorder/Substance Use (M, Mdn, SD)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Prior belief ($\hat{q}_1$)</td>
<td>.93 (.96, .10)</td>
<td>.98 (.99, .02)</td>
</tr>
<tr>
<td>Prior Uncertainty ($\ln(\delta)$)</td>
<td>-9.97 (-9.83, 2.35)</td>
<td>-10.67 (-10.90, 2.18)</td>
</tr>
<tr>
<td>Reward Sensitivity ($\gamma$)</td>
<td>.56 (.67, .37)</td>
<td>.55 (.64, .26)</td>
</tr>
<tr>
<td>Deliberation consistency</td>
<td>1.43 1.50 .90</td>
<td>1.02 0.90 0.50</td>
</tr>
<tr>
<td>Ex-Gaussian ($\mu$)</td>
<td>.19 .20 .03</td>
<td>.19 .16 .09</td>
</tr>
<tr>
<td>Ex-Gaussian ($s$)</td>
<td>.02 .02 .01</td>
<td>.02 .02 .02</td>
</tr>
<tr>
<td>Ex-Gaussian ($\tau$)</td>
<td>.03* .02 .03</td>
<td>.07* .05 .08</td>
</tr>
</tbody>
</table>

*p < .10 using Mann-Whitney U test.*
Figure 1. RTs as a function of pump number. The top panel shows the average RTs across subjects and balloons as a function of normalized pump number. The bottom three panels show the RTs for representative balloon trials for three different subjects on three different balloon trials. The solid straight line indicates the average baseline BART RT for the average participant (top panel) and for the individual participants (bottom 3 panels). The dotted line identifies three standard deviations above the baseline BART RT, which was used to define assessments.
Figure 2. The empirical cumulative distribution functions across subjects for the standard and baseline BARTs.
Figure 3. Average assessment rate as a function of balloon trial. The solid line is the predicted assessment rate predicted by the BSR-PAM fit to each individual.
Figure 4. The proportion of non-assessed pumps taken preceding each assessment in a given balloon averaged across subjects and balloon trials. Note the number of observations that contributed to each data point decrease monotonically by assessment number with the number of observations at assessment number 1 being $N = 205$ and the number of observations at assessment number 20 being $N = 45$. 
Figure 5. The dual response pathway BSR model (drBSR).
Figure 6. Observed and model simulated data for the controls and CD/SU participants in Crowley et al. (2006). The left panel plots the adjusted BART scores for the two groups. The right panel plots the RTs for the two groups. The summary statistics for the observed data are in Table 3.
Figure 7. Empirical CDFs of RTs for control and CD/SU patients in Crowley et al. (2006)
Figure 8. Assessment rate as a function of balloon trial. The left panel plots the function for controls and CD/SU. The right panel re-plots the data from the first study shown in Figure 3 demonstrating that the CD/SU adolescents differed from this sample of young adults. The lines are the predicted assessment rate functions derived via simulations of the drBSR model.
Figure 9. The proportion of non-assessed pumps taken preceding each assessment in a given balloon averaged across subjects and balloon trials.
Appendix

The drBSR model was fit to each individual’s data using maximum likelihood methods. The likelihood of the data is given by the following expression

\[
L = \prod_{h=1}^{30} \prod_{i=1}^{I_h} \{s[h, i]r[h, i']f[t|pump_h(i)][1 - s(h, i)]g(t)\}^{c_{h,i}} \times \{s[h, i][1 - r_{h, i'}]f[t|stop_h(i)]\}^{(1 - c_{h,i})}.
\]

\(I_h\) is the number of responses (pump or stop) the participant took on balloon \(h\). The variable \(c_{h,i}\) is a 1 if the participant made a pump on balloon \(h\), pump \(i\) otherwise it is 0. The variable \(t_{h, i}\) is the observed response time on balloon \(h\), pump \(i\). Equation 6 gives the expression for \(s(h, i)\), Equation 7 gives the expression for \(r_h(i)\), and Equation 8 gives the expression for \(f[t|pump_h(i)]\) and can also be used to find \(f[t|stop_h(i)]\).

A closed form solution of the likelihood is not available, so the maximum was estimated with numerical optimization methods. We used Nelder and Mead’s (1965) downhill simplex routine (available in Mathwork’s Matlab) combined with a grid search technique to estimate the maximum. During the estimation process the parameter space was divided into plausible sectors and then we randomly selected starting values from these sectors. This was done 18 times. The maximum likelihood parameter estimates from the 18 different starting points was identified and then tested with simulations to check for correspondence between observed and predicted data.

In fitting the model, we used all the observed response times and choices in study 1. In the re-analysis of Crowley et al. (2006) we also used all the observed response times and choices. However, due to the first pump being always much larger than any other pump time (between 1 and 2 s) we excluded this RT from the fitting.