Making Assessments While Taking Sequential Risks

Avishai Wershbale (wershbal@msu.edu) & Timothy J. Pleskac (pleskact@msu.edu)
Department of Psychology,
Michigan State University,
East Lansing, MI 48824

Abstract

This study utilized inter pump response times on a laboratory gambling task, the BART, to examine cognitive aspects of response selection during sequential risky decision making. Findings suggest a response procedure that utilizes multiple levels of processing. Amount of task exposure as well as the distance to the goal both affect the rate at which assessments are made, with task exposure decreasing assessment rate, while target distance increases assessment rate. Several alternative models are fit to the data, to determine if the behavioral results can be informative of a model that more accurately reflects differences in processing.

Keywords: Psychology, Decision Making, Mathematical Modeling, Cognitive Decision Theory

Introduction

People take sequential risks every day. Drivers repeatedly choose to talk on the cell phone, or text while driving each time they get in their car. People every day choose to eat at their favorite fast food establishment. Smokers of all ages repeatedly choose to smoke a cigarette. In all of these situations, our choice is not a one-shot deal rather our choices occur many times sequentially over the course of a day, a week, a month, etc. Often we are even presented with the same or similar choices on multiple occasions. Sometimes these choices may even change as a function of time or even as a function our previous choices (Busemeyer & Pleskac, 2009).

Despite the many instances of sequential risks in the real world, most of the laboratory analogs of risky decision making only ask participants to make one single choice (e.g., Hertwig, Barron, Weber, & Erev, 2004; Kahneman & Tversky, 1979). However, a number of laboratory-based gambling tasks have now been developed that require people to take sequential risk, such as the Iowa Gambling Task (Bechara, Damasio, & Anderson, 1994); the Balloon Analogue Risk Task (BART; Lejuez et al., 2002); or the Angling Risk Task (Pleskac, 2008). These tasks also appear to have some construct validity with real world risk taking. Risk taking in the BART, for example, correlates with smoking and drug abuse, as well as safety issues including seatbelt usage and safe sex (Lejuez et al., 2002; Lejuez, Aklin, Zvolensky, & Pedulla, 2003; Pleskac, Wallsten, Wang, & Lejuez, 2008).

The overall decision making processes for these sequential risk taking tasks are also reasonably well understood. In fact, the processes have been formalized in terms of cognitive models (Busemeyer & Stout, 2002; Wallsten, Pleskac, & Lejuez, 2005). However, the processes postulated within the components of the model have received less direct attention. In this study, we focus on the BART and test some of the processing implications of its respective cognitive model. In particular we use inter pump times to better understand the response selection process that decision makers use to take sequential risks during the BART.

Balloon Analogue Risk Task (BART; Lejuez et al., 2002)

During the BART participants are presented with a computerized balloon, a pump button, and a stop button. Pressing the pump button inflates the balloon, and also puts money in a temporary bank. Balloons in the BART explode after a randomly predetermined amount of pumps are made. An explosion terminates the trial and all money in the temporary bank is lost. Clicking the stop button, transfers the temporary points into the permanent bank and also terminates that trial. The participant’s goal in the BART is to earn as much money as they can, but they need to take into account the chance that the balloon could pop. Participants typically complete 30 independent balloon trials.

Participants are given no information about the design of the task, other than the basic reward and punishment rules, and the fact that the balloon will explode before it fills the entire screen. It is up to the participant to determine the amount of risk they are willing to incur for a given reward. Though the BART may seem simple at first, there are many processes that take place within the course of completing the task. The processes have been formally defined in the Bayesian Sequential Risk Taking model (BSR; Wallsten et al., 2005).

Bayesian Sequential Risk Taking (BSR) Model

We briefly review the general processes of this 4 parameter model (for a formal derivation see Pleskac, 2008; Wallsten et al., 2005). The BSR model consists of three sub processes. The first process is a reward evaluation process. During this process, participants select a target pump number to pump the balloon towards. The target selected is a function of the participant’s subjective value of a reward and their perceived chance that an explosion will occur for any pump. Participants who value the reward more will increase their target amount of pumps to be made, while an increase in the perceived chance of an explosion occurring decreases this pumping target.

The second process in the model is a response selection process. This process describes how participants use their target to determine whether to select a pump or a stop response at each pumping opportunity. In particular, the probability of pumping is assumed to be a function of the participant's current distance to the pumping target that they derived in the reward evaluation component. According to the model, the probability of pumping decreases as the current number of pumps taken approaches and passes the target number of pumps. Some participants are more consistent in pumping to their targets. A topic of interest in this paper is if in fact participants appear to be making a distance calculation, as the model seems to assume, at every pump opportunity.

The final process in the BSR model describes how participants learn from their experience. This process describes how participants arrive at their belief of the probability that the balloon will explode (used in the reward evaluation process described previously). The model assumes participants use a Bayesian learning process to integrate their prior beliefs with the observed data from each balloon trial (# pumps and if it exploded or not). Their new belief is used to evaluate rewards and select a target during the next balloon trial.

The model has been formally specified and tested (see Pleskac, 2008; Pleskac et al., 2009; Wallsten et al., 2005) with previous studies being by and large centered on the reward evaluation process and learning components of the BART. Little focus, however, has been allocated to the response selection component of the model. Recall the BSR model implies that on every pump opportunity of every balloon participants engage in some sort of distance-to-target calculation. If they are far from the target they are almost certain to pump and as they approach the target they become more and more likely to stop pumping. This raises the question whether participants perform a distance calculation at every pump opportunity? Our hypothesis is that instead of performing this calculation on every pump opportunity, there are instead two different types of pumping behavior being utilized. One pump type is a relatively automatic pump, while on other pump opportunities, decision makers pause and take an assessment of how far they have gone and how far they want to go. To test this hypothesis we examined the inter-pump times (the amount of time taken between responses).

Assessments

Cognitive psychologists have long known that due to limitations in working memory capacity an increase in the amount of information to process leads to an increase in the time it takes to process that information (Atkinson, Holmgren, & Juola, 1969; & Schneider & Schiffrin, 1977). This means that an action that occurs following a complex calculation should have a slower response time then if that action were preceded by an easier calculation. This cognitive principle implies the distance calculation the BSR assumes to take place when selecting a response should take some observable amount of processing time over and above motor time.

However, we also know that these sequential risk taking situations require choices on multiple trials. This high exposure to the task and task structure may lead to a routinization of the decision making process (Betsch, Haberstroh, Glöckner, & Fiedler, 2001) and perhaps even eventually approaching the automaticity properties of a habit (Aarts, Verplanken, & van Knippenberg, 1998). This routinization of decision making would imply less and less demands on working memory and thus lead to fairly quick inter-pump times.

Our hypothesis though is that there is some mix between fairly routine almost automatic pumps and other pumps where the decision maker pauses to take an assessment of where they are in the balloon trial. Our hypothesis is very much motivated by analogous findings from the animal learning literature where rats make a series of sequential decisions while traversing a maze. In particular, rats when learning a maze will at some decision points pause and appear to orient themselves toward potential options (Tolman, 1938 & Tolman 1948). Then after orienting themselves make a decision. This behavior has been termed vicarious trial and error (VTE), and has several interesting characteristics (Gallistel, Fairhurst, & Balsam, 2004). It was found that these VTEs occur fairly frequently during the early learning trials, and decreases with exposure to the task. This decrease in VTE's means that after enough exposure to the task environment, rats upon reaching a decision point (a fork where they have to go either right or left), eventually stop orienting themselves towards both potential options before making a decision, and instead simply immediately take the correct turn. It was also shown that this decrease in VTE's takes a non-linear shape.

These results prompt the question whether our postulated assessments follow the same pattern as VTEs. To test this we examined inter-pump times. Our hypothesis was that the inter-pump intervals in which a distance calculation was performed should take longer than those intervals in which no calculation was performed. And the inter-pump times for non-calculation intervals should not differ from baseline pumping speed.

To determine baseline inter-pump times, participants first completed a task in which only one option is presented to them: a pump option. Participants were instructed to pump each balloon as quickly as possible until they exploded. Participants neither received nor lost money for these trials. The inter-pump times from this were averaged together to estimate a baseline inter-pump time for each participant, as well as the standard deviation of their baseline pumping speed.

An assessment pump was operationalized as any pump for which the respective inter-pump time was 3 standard deviations or greater than the mean baseline time. Our hypotheses are as follows.

Hypothesis 1: Assessments will only occur periodically throughout a given balloon trial. The remaining trials will reflect relatively routine almost automatic choices due to the frequency of their occurrence.

Hypothesis 2: Another testable prediction comes out of the hypothesis that the assessment points found in the BART reflect the same type of learning as the VTE's in rodent maze learning. As in the VTE's we expect that over balloon trials assessments will decrease in a non-linear fashion so that the observed assessment rate will decrease as exposure to the BART increases.

Hypothesis 3: Within a balloon trial, as participants approach their targeted stopping point their assessment rate should increase. This hypothesis is derived from the following reasoning. First, we assume that as participants pump they form an association between pump opportunities and the difficulty of making a choice (to pump or not). For example, participants will tend to associate the 4^{th} pump trial with an easy decision (pump), but later pump opportunities (e.g., 48^{th} opportunity) will present the participant with a more difficult choice. The prediction that follows from this hypothesis is that participants should be more likely to make an assessment on later pump opportunities for a given balloon.

Finally, we were interested whether the actual magnitude of the inter-pump times on routine pumps (non-assessed) would change over the course of pumping any given balloon. If we think of reaching the target pump as the main goal and the assessment points as sub-goals in reaching the target then we form a goal hierarchy. We know from goal activation models (Altmann & Trafton, 2002) that respondents often track their distance from previous sub goals and this leads to a slowing in response times as they progress. One might expect that participants are somehow implicitly tracking their distance from the last assessment. This would imply that inter-pump times of non-assessed pumps should increase as the distance from the last assessment increases. Next we test these predictions using data from two studies. Then we propose modifications based on these results to the BSR model.

Methods

The data examined comes from two experiments conducted in the Laboratory for Cognitive Decisions at Michigan State University during spring of 2008 and spring of 2009. These studies were designed to look at the effect of individual differences in various executive functions on BART performance. Both a standard version of the BART and a response time BART were included in these studies as well as a number of other executive function tasks, however we will limit descriptions of the tasks to just those relevant to this paper. Participants were college age undergraduates. A total of 104 students participated in the 2008 study and 108 in the 2009 study. There were no substantial differences between the two studies so we report their results together.

Baseline BART

The first task that every participant completed in both data sets was a baseline BART. The baseline BART is a simplified version of the BART that was created to measure average response time for pumping behavior. This version has only a pump button and participants are instructed to pump each balloon until it explodes. The balloons in the baseline BART were programmed to explode with the same statistical distribution that the normal BART balloons utilize. Participants completed ten trials of the baseline BART in order to establish a baseline non-fatigued measure of pumping motor time.

Manual BART

The regular BART task that we used is based on the task used in previous studies (see Lejuez et al., 2002; Pleskac et al., 2009). The task consists of a virtual balloon that is inflated by pressing a button. Participants were awarded 10 points for each successful pump. The popping point for each trial was randomly chosen out of 128, and pairings for each random trial were included to assure the same optimal distribution as in the original BART paper (Lejuez et al., 2002). Each trial ends with either a popped balloon (participants earn no points), or the participant clicking the stop button in which case the participant keeps all of the earned points for that balloon trial. Either way a fixation cross then appears in the center of the screen to prepare the participants for the onset of the following trial. To obtain more accurate response time data, our version of the manual BART was programmed in E-Prime 2.0. Furthermore, participants entered their pump and stop choices with separate keyboard buttons.

Results

Behavioral

In both studies, participants' risky behavior was consistent with past studies. On non-exploding balloons, they pumped an average of 39 (SD =16.03) and 34.3 (SD = 18.01) pumps

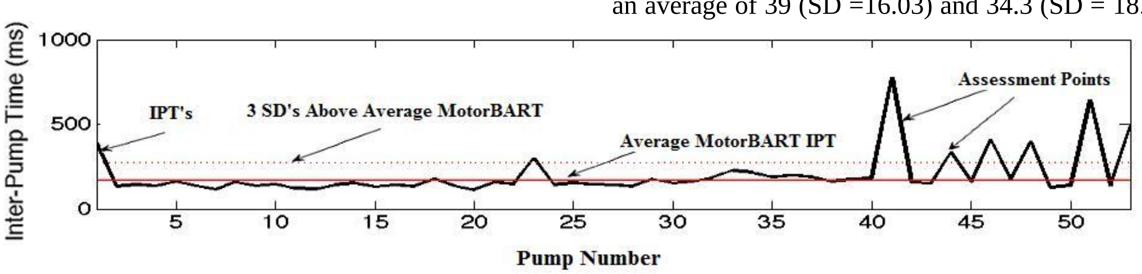

Figure 1: Inter Pump Time on a Single Trial for Participant 55 in study 2.

in studies 1 and 2, respectively. The average baseline inter-pump time in both studies was 181.38 ms (SD = 58.4) and 178.47 ms (SD = 54.22). The average within-subject standard deviation of inter-pump times on the baseline task was 53.91 ms (SD = 32.14) and 55.38 ms (SD = 40.66).

Recall we defined an assessment as any pump during the BART for which its respective inter-pump time exceeded 3 standard deviations above the baseline inter-pump time. With that definition, 11.1% of the pump opportunities from study 1 and 16.35% from study 2 would be classified as assessments. This yields an average assessment rate of 4.3 and 5.6 assessments per trial respectively. A plot of inter-pump times from a single subject and balloon trial is shown in Figure 1.

In terms of Hypothesis 2, to test whether there was a change in assessment behavior as participants become more familiar with the task, we regressed assessment rate onto trial number. Assessment rate was calculated by dividing the number of assessments in a given trial by the length of that particular trial. The results of this regression showed the same pattern for both studies, which is a decrease in assessment rate as trial number increases. Averaging across participants, the data seemed to best fit a logarithmic decreasing curve, with study 1 significant at $R^2 = .876$ and $p < .001$, and study 2 significant at $R^2 = .935$ and $p < .001$ (Figure 2). Thus, assessment rate was high on the first few trials (approximately a 40% assessment rate on average) and then decreased at an increasing rate as participants experienced more balloon trials.

Hypothesis 3 focused on whether the probability of an assessment changes within a balloon trial. Pump number itself cannot be used as the predictor for this regression, due to the fact that the length of each trial is entirely dependent on the participant's own pumping behavior. Instead a count of how many non-assessment pumps was taken between each assessment point, and then that count was divided by the length of that trial. This number is the proportion of pumps within that trial that preceded each assessment point (excluding the initial pump opportunity). Assessment points that are immediately followed by another assessment point were only counted as a single assessment. These proportions were then averaged across trials and across participants to give us a proportion score. A regression was run with assessment point number as the predictor and the proportion score as the dependant measure. Results of this regression also had a logarithmic fit, with $R^2 = .652$ and $p < .001$ for study 1, and $R^2 = .675$ and $p < .001$ for study 2. This is similar to the shape as for trial number, but the interpretation is that assessment rate increases as a factor of pump number.

Along with determining the factors that influence the probability of a distance to target assessment occurring, it is also important to identify characteristics of the non-assessed pumps as they approach the next assessment point. To characterize the changes in response times of the non-assessed pumps, Goodman-Kruskal Γ rank order correlations were ran to determine if there is generally an

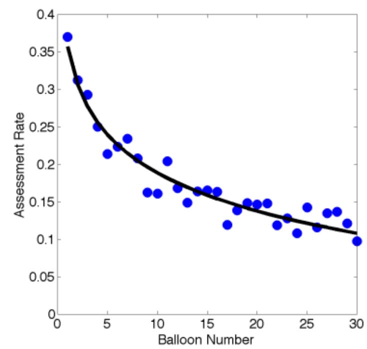

Figure 2: Change in Assessment Rate by Trial Number

increase, decrease, or no change in the response times the non-assessed pumps that were made between assessment points (hypothesis 4). A Γ coefficient was calculated for every string of non-assessed pumps that occurred. When the gamma coefficients were averaged (2615 coefficients in study 1 and 2887 coefficients from study 2), the results indicate an increase in inter-pump times time for the non-assessed pumps as they approach the next assessment point, with an average gamma coefficient of .108 for study 1 and .135 for study 2. While small, this result indicates a small but systematic increase in inter-pump times the further one gets from an assessment.

These results suggest that choice behavior during the BART is a bit more complex than that which is depicted in the BSR model. In particular, we have shown that in two studies on some trials participants pause and perhaps take an assessment of their situation. On the other trials the inter-pump times are quick enough to suggest a routine or perhaps even an almost automatic pumping behavior. Next we examine how to best modify the BSR model to incorporate these findings.

Proposed Changes to BSR Model

One possible way to account for these observed pauses in the BSR model is to modify the response selection process. Figure 3 illustrates this proposed change. The idea is that in the original response selection process participants either pumped or stopped and each response followed an assessment. Instead we propose that not every pump opportunity involves a distance to target assessment. That is with some probability participants stop to make an

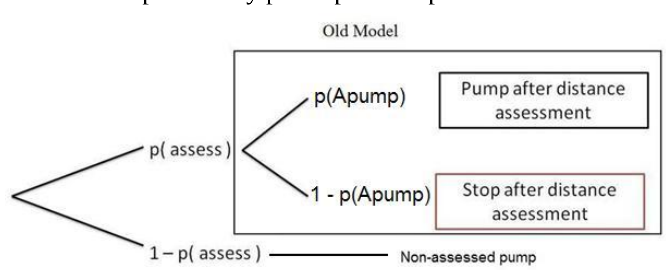

Figure 3: One possible modification to the BSR model.

assessment and only then do they choose between pumping and stopping. On other trials they make what we call a non-assessed pump which has an inter-pump time that is close to baseline. Several different functional forms of this modification were examined to determine if the inclusion of non-assessed behaviors improves the model fit.

The changes to the model are based off of the predictions from the new hypotheses. We tested four modifications. The first model assumed a static assessment rate, μ, which estimates the probability of making an assessment before the response is selected.

The second model assumed assessment changed as a function of balloon trial (h),

$$P(assess) = \frac{1}{1+\exp(\mu(h-\lambda))}. \quad (1)$$

Where λ is a biasing factor that controls the starting point of the assessment rate and μ now controls the rate of change in the assessment rate.

The third model incorporates the idea that assessments change as a function of pump opportunity i,

$$P(assess) = \frac{\exp(\mu(i-\lambda))}{1+\exp(\mu(i-\lambda))} \quad (2)$$

The parameters serve much the same role as in Equation 2, but now assessment rate changes as a function of pump opportunity and λ controls the starting assessment rate for each balloon.

The fourth model incorporates both an assessment rate that changes as a function of balloon trial h and pump opportunity i,

$$P(assess) = \frac{\exp(\mu(i/h-\lambda))}{1+\exp(\mu(i/h-\lambda))} \quad (3)$$

Preliminary Model Fitting & Comparisons

The modifications were incorporated into the BSR model. Then each model was fit at the individual level with maximum likelihood procedures using the Nelder-Meade numerical optimization routine. Several different starting points were used to try and guard against local maxima issues. A Bayesian Information Criterion (BIC) was calculated for each model, and was used to determine the best fitting model overall (where lower BIC means better fit). The BIC is a goodness of fit measure which penalizes models for the number of parameters they have.

Table 1: Average BIC scores for Alternative Models

	Ave (Std) BIC	Ave (Std) BIC
Baseline	740.32 (358.85)	758.86 (428.15)
Static Assessment Rate	733.06(360.15)	746.54 (434.73)
Assessment rate changes as a function of balloon trial	718.74 (325.78)	734.20 (397.82)
Assessment rate changes as a function of pump opportunity	728.57 (358.84)	747.88 (442.52)
Assessment rate is a function of both trial and pump opportunity	741.62 (338.5)	733.60 (410.01)

The models were also compared against a statistical baseline model. The baseline model simply uses the observed proportion of assessed pumps, non-assessed pumps, and stops over the thirty balloons and estimates the likelihood of the data by utilizing those proportions.

The average BICs are shown in Table 1. They show that the best fitting model is one in which assessment rate changes as a function of balloon trial. It is of note that there is some variability in this conclusion at the individual level. In particular, while nearly all the participants exhibited some form of assessment behavior, there was individual variability in the relationship between assessment rate and balloon trial. Some (~40%) showed a very weak relationship between balloon trial and assessment rate. For these individuals, the constant assessment model and the baseline model provided better fits.

Discussion

This study aimed to better understand the response selection process in sequential risk taking situations. Using the BART as an analog to these situations, we found that nearly all participants engaged in a behavior we call assessment. That is, within a given sequence of risky choices, generally decision makers would make very quick choices, but periodically (about every 4 to 5 pumps) they would take very long pauses. We interpret this behavior as a time of assessment when the decision maker gauges how many risks they have taken and how many more risks they plan to take. We also found the following behavioral properties of an assessment rate.

First, across balloon trials the assessment rate was on average higher for early balloon trials and then diminished at an increasing rate. This idea is consistent with previous decision making literature with rodents, showing that as task exposure increases, learning takes place, which leads to automated decision evaluations (Tolman, 1938 & Tolman 1948). The second, property of assessment rate, was the change within a single balloon trial, where assessment rate increases towards the end of the trial. One possible explanation is that the assessment rate increases relative to the level of perceived risk. It would be interesting to see if this is reflected in self reported risk taking measures. Lastly, there was a small but significant increase in inter-pump times between assessments. This suggests an increasing taxing of cognitive resources, which may be due to a buildup of interference, so that eventually the participants need to reassess their location relative to their pump target

Assessments and the change in assessment rate over balloon trials may be analogous to the distinction between exploration and exploitation in sequential decision making (Schumpter, 1934 & Holland, 1975). This idea of exploration versus exploitation holds that in order to maximize gains, one should initially explore the structure of the environment to create a good approximation of the distribution of rewards. Once a good approximation of the environmental structure is obtained, then one should begin exploiting it in a manner that maximizes their gains.

Assessments in sequential risk taking may afford the decision maker with an opportunity to explore different risk options and then exploit the options.

Acknowledgments

We thank David McFarlane for his help in programming the BART in E-Prime.

References

Aarts, H., Verplanken, B., & van Knippenberg, A. (1998). Predicting behavior from actions in the past: Repeated decision making or a matter of habit? *Journal of Applied Social Psychology*, *28*, 1355-1374.

Altmann, E. M. & Trafton, J. G. (2002). Memory for goals: An activation-based model. Cognitive Science, 26, 39-83

Atkinson, R. C., Holmgren, J. E., & Juola, J. F. (1969) Processing time as influenced by the number of elements in a visual display. *Perception & Psychophysics, 6*, 321-326

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

Betsch, T., Haberstroh, S., Glöckner, Haar, T., & Fiedler, K., (2001). The effects of routine strength on adaptation and information search in recurrent decision making. Organizational Behavior and Human Decision Processes*, 84*, 23-53.

Busemeyer, J. R., & Pleskac, T. J. (2009). Theoretical tools for understanding and aiding dynamic decision making. *Journal of Mathematical Psychology*, *53*(3), 126-138.

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

Gallistel, C. R., Fairhurst, S., & Balsam, P. (2004). The learning curve: Implications of a quantitative analysis. *PNAS*, *101*, 13124-13131

Hertwig, R., Barron, G., Weber, E. U., & Erev, I. (2004). Decisions from experience and the effect of rare events in risky choice. *Psychological Science, 15*(8), 534-539.

Holland, J. H. (1975), *Adaptation in Natural and Artificial Systems*. Ann Arbor, Ml: University of Michigan Press

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

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

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

Pleskac, T. J. (2008). Decision making and learning while taking sequential risks. *Journal of Experimental Psychology-Learning Memory and Cognition*, *34*(1), 167-185.

Pleskac, T. J., Wallsten, T. S., Wang, P., & Lejuez, C. W. (2008). Development of an Automatic Response Mode to Improve the Clinical Utility of Sequential Risk-Taking Tasks. *Experimental and Clinical Psychopharmacology, 16*(6), 555-564.

Schneider, W., & Shiffrin, R. M. (1977). Controlled and automatic information processing: I. Detection, Search, and attention. *Psychological Review*, *84*(1), 1-66

Schumpter, J. A. (1934). *The Theory of Economic Development*. Cambridge. MA: Harvard University Press.

Tolman, E. C. (1938). The determiners of behavior at a choice point. *Psychological Review*, *45*, 1-41

Tolman, E. C. (1948). Cognitive maps in rats and men. *Psychological Review, 55*, 189-208

Wallsten, T. S., Pleskac, T. J., & Lejuez, C. W. (2005). Modeling behavior in a clinically diagnostic sequential risk-taking task. *Psychological Review*, *112*(4), 862-880.