Aligning Data Collection with Multi-dimensional Construct Definitions: The Example of Behavioral Tasks for Measuring Risk-taking Behavior

Carl W. Lejuez, Dean, College of Liberal Arts & Sciences, University of Kansas

In their seminal work, Jessor and Jessor (1977) defined risk-taking as “behavior that is socially defined as a problem, a source of concern, or as undesirable by the norms of conventional society and the institutions of adult authority, and its occurrence usually elicits some kind of social control response” (p. 33). Focusing more explicitly on the potential consequences or outcomes of such behavior, risk-taking has also been conceptualized as behavior that involves some potential for harm or negative consequence to the individual, but that may also result in a positive outcome or reward (Byrnes et al. 1999; Leigh 1999). Further, the propensity to take risks exists on a continuum, with some risk-taking being adaptive, and only more extreme levels being maladaptive (Bornova et al. 2009). The availability of well-developed behavioral methodologies used to study risk-taking behavior and to ensure quality data that are reliable, valid, and meaningful are crucial to the advancement of our understanding of the processes underlying the development and maintenance of risk behaviors (Koffman & Kaplan, in press).

Decision-making Risk-taking Tasks
The Iowa Gambling Task

The original gold standard measure of risk taking is the Iowa Gambling Task (IGT; Bechara et al. 1994). The IGT is a decision making task originally developed to examine decisional processes associated with neuropsychological impairment (e.g., Rogers et al. 1999a). At the beginning of the task, the participant is given $2,000, is instructed to maximize earnings over the course of 100 decision-making trials, and is provided with four decks of cards on the computer screen. As described by Bechara et al. (2001), the decks are labelled A, B, C, and D at the top end of each deck. All cards are identical, and each card is associated with hypothetical payoffs or losses (although versions with real financial contingencies are available). Cards from decks A and B pay an average of $100, but also contain cards with higher losses, while cards from decks C and D pay an average of $50, but losses are smaller. Accordingly, 10 draws from decks A and B (the “disadvantageous” decks) lead to a net loss of $250, while 10 draws from decks C and D (the “advantageous” decks) lead to a net gain of $250 (Bechara et al. 1994; Buelow and Suhr 2009).

During the task, the participant clicks on a card from any of the four decks. Once the card is selected, the computer makes a sound similar to that of a slot machine. The selected card appears as either red or black, indicating whether money was lost or gained, and the value of the
reward or loss appears at the top of the
screen. Following this feedback, the card
disappears and the participant selects an-
other card. Each deck of cards is pro-
grammed to have 60 cards (30 red and 30
black), although the participant is una-
ware of how many cards of each type are
in each deck. Losses are equally frequent
in each deck. Several dependent variables
from the IGT indicate RDM, but the most
widely reported indices are the number
or percentage of disadvantageous choices
over 100 trials, with larger values repre-
senting greater riskiness.

While there is value in considering
risk taking this way, there are concerns
about the challenge in being able to re-
peat the task given the participants goal
is to “figure out” the risky options and
avoid them. Thus, once this is figured
out, regardless of how long that has
taken, there is no way to return that same
participant to the naïve state in relation to
the goals of the task. Researchers have at-
ttempted to address this issue, but it re-
 mains a challenge for studies that require
repeated measures (Almy, Kuskowski,
Myers, & Luciana, in press). The nature of
the task also presents a conceptual chal-
lenge as the task’s focus on risky decision
making means that it only provides one
narrow way to explore risk taking. In-
deed, understanding the maladaptive as-
pcts of risk taking from a decision mak-
king perspective is at the same time in-
credibly valuable for understanding one as-
pct of risk taking but also too narrow to
be considered a comprehensive assess-
ment. Given that risk taking is so multi-
dimensional, the argument is not that the
IGT should be discarded given its nar-
rowness, but instead that it should be cat-
egorized clearly for the aspects of risk
taking it does represent well and there
should be simultaneous efforts to de-
velop tasks that are targeted at the other
dimensions of the construct. The remain-
der of this paper seeks to discuss this very
goal.

**Risk-Taking Propensity Tasks**

In considering other key dimensions
of risk taking, an obvious starting point is
the movement away from examining risk
taking as solely maladaptive and instead
as a continuum, ranging from maladap-
tive in different ways at its extremes and
adaptive in moderation. Such an ap-
proach makes intuitive sense with one re-
ceiving few rewards when taking no risks
whatsoever and too many negative con-
sequences when extreme risks are taken,
thereby leaving risk taking in moderation
as an ideal goal in most situations, and
the outcome measure most relevant be-
ring risk taking propensity (RTP). As our
research group began to develop a task to
do this very thing, we found a simple ap-
proach that had been around for some
time that could serve as a starting point.

**Slovic’s Devil Task**

The first behavioral task designed
and used to assess RTP was Slovic’s Devil
Task (1966). Although it was not used to
assess risk-taking related to substance
use in the extant literature, review of the
Devil Task is important for its historical
significance and relation to current, com-
monly used RTP behavioral measures.

In the Devil Task originally imple-
mented by Slovic (1966), participants
were seated before a panel of ten small
knife switches and told that nine of the
switches were “safe” and that the tenth
was a “disaster” switch, with it being im-
possible to distinguish which was the dis-
disaster switch. Each participant was informed that a switch could only be pulled once (i.e., sampled without replacement), thus the likelihood of pulling the disaster switch increased with each subsequent trial. The participant was then asked to pull one of the switches and, if the participant chose a safe switch, he/she was allowed to place one spoonful of candies into a glass bowl. The participant then had to decide whether to pull another switch or to stop and keep the candy he/she had already won. If the participant decided to continue but subsequently pulled the disaster switch, a buzzer would sound and the participant would lose everything he/she had already earned. The task ended when the participant either chose to stop and collect his/her winnings or pulled the disaster switch and lost everything. In the event that the participant pulled nine safe switches in a row, he/she was automatically forced to stop and take his/her nine spoonfuls of candy, as the only remaining switch necessarily was the “disaster” switch. Slovic (1966) argued that because both the probability and magnitude of one’s potential loss increases with the number of switches pulled, stopping performance on the task can be considered an index of risk-taking tendencies. A slightly altered version of the Devil Task includes a presentation with ten wooden boxes, where one box contains a devil. The potential reward for this version of the task are stickers (Hoffrage et al. 2003). A computerized version of the Devil Task using boxes also exists (Eisenegger et al. 2010).

The Devil Task is most notable for its historical significance. Despite its limited use, the task possesses multiple features that make it a useful measure for studying risk-taking. For example, the Devil Task discriminates between risk takers without involving learning (Hoffrage et al. 2003). Although few variants of the task have been tested, the task could be a useful vehicle for examining contextual influences on risk-taking such as the impact of varying either the number of switches and/or the magnitude of stakes. Because the task has clear face validity, it is likely useful without modification across developmental stages, and the short administration time makes it ideal for examining the impact of experimental manipulations such as stress or drug/alcohol administration.

Balloon Analogue Risk Task (BART)

The Devil Task provided a simple model of RTP. This simplicity gave rise to our development of a task that took the basic idea of measuring RTP but was done in a way that would allow for increasingly complex ways to study the type of complex risk behaviour seen in the real world. This work led to the development of the Balloon Analogue Risk Task (BART; Lejuez et al. 2002).

The BART is a computerized measure of RTP that models real-world risk behavior through the conceptual framework of risky behaviors in having both the potential for reward as well as the risk of harm (Leigh 1999; Lejuez et al. 2002). In the task, the participant is presented with a balloon and asked to pump the balloon by clicking a button on the screen. With each click, the balloon inflates .3 cm and money is added to the participant’s temporary winnings; however, if the participant pumps up the bal-
loon beyond its explosion point, the balloon explodes and he/she loses the money earned on that balloon. The task is depicted in the Figure below.

The explosion point can be varied both across and within studies. Lejuez et al. (2002) used three different balloon colors, with each color associated with a range of 1-8, 1-32, or 1-128 pumps, respectively. The probability that a balloon would explode was arranged by constructing an array of \( N \) numbers. The number 1 was designated as indicating a balloon explosion. On each pump of the balloon, a number was selected without replacement from the array. The balloon exploded if the number 1 was selected. For example, for the blue balloon ranging from 1-128, the probability that a balloon would explode on the first pump was 1/128. If the balloon did not explode after the first pump, the probability that the balloon would explode on the second pump was 1/127, on the third pump, 1/126, and so on up until the 128th pump, at which the probability of an explosion was 1/1 (i.e., 100%). According to this algorithm, the average break point would be the midpoint of the range, in this case 64 pumps. Before the balloon pops, the participant can press “Collect $$$” which saves his or her earnings to a permanent bank. If the balloon pops before the participant collects the money, all earnings for that balloon are lost, and the next balloon is presented.

RTP on the BART is defined as the adjusted average number of pumps on unpopped balloons (Bornovalova et al. 2005; Lejuez et al. 2002), with higher scores indicative of greater RTP. In the original version of the task, each pump was worth $.05 and there were 30 total balloons. Participants were provided this information, but were not given information about the breakpoints. This instructional set allowed for the examination of participants’ initial responses to the task and to changes as they experienced the contingencies related to payout collections and balloon explosions. Results of the original study also established
that the relationships between key outcome variables and BART scores were most evident at the largest balloon range of 1-128, which has largely been used in subsequent studies.

The BART has well-established reliability across a range of samples. Split-third reliability has been examined by comparing scores across the first block of 10 balloons, middle block of 10 balloons, and final block of 10 balloons on the task. The reliability estimates typically indicate strong correlations (>0.7) among the blocks (Lejuez et al. 2010; Lejuez et al. 2002). Extended test-retest reliability has been indicated with the task presented twice across a two-week period, with a nonsignificant increase across administrations (T2–T1Δ = 1.2 adjusted average pumps) and a reasonably robust test-retest correlation (T1/T2 \( r = .77 \); White et al. 2008).

It is notable that relationships of performance on the BART to self-report measures of disinhibition are inconsistent. Findings indicate modest, though significant, relationships with sensation seeking (\( r = -.20 \)), but typically nonsignificant relationships with self-report and other behavioral measures of impulsivity (Bornovalova et al. 2005; Lejuez et al. 2007; Meda et al. 2009).

The BART is considered to be one of a small number of gold-standard measures of risk-taking (Harrison et al. 2005). Within adolescent studies, RTP as measured by the BART is related to adolescent self-reported engagement in real-world risk-taking behaviors including substance use behaviors and delinquency/safety behaviors (Aklin et al. 2005; Fernie, Goudie, & field, 2010; Lejuez et al. 2003b). Greater levels of RTP on the BART have been found in adolescent ever-smokers as compared to never-smokers as well as in adolescents with conduct and substance problems when compared to matched controls (Crowley et al. 2006; Lejuez et al. 2005). Performance on the BART-Y (a youth version; Lejuez et al. 2007) correlates significantly with a variety of real-world risky behaviors. Specifically, greater RTP on the BART-Y is associated with increased frequency of substance use, gambling, delinquent behaviors, and risky sexual behavior (Aklin et al. 2005; Hamilton, Felton, Risco, Lejuez, & MacPherson, 2014; Lejuez et al. 2007; Lejuez et al. 2005; Macpherson 2010).

**Prediction of Future Risk Behavior**

Of particular importance is the extent to which the BART can be used to predict future risk behavior. As noted above, existing data often shows a correlation between risk taking on the task and current levels of real world risk behavior, and a handful of studies have further shown that changes in risk-taking behavior in the real world over a period of time correlate with changes in risk taking on the BART over the same period of time (MacPherson et al. 2010). However, there appears to be no evidence of risk taking on the BART at one time point predicting future risk taking behavior. The absence of prediction data considered together with reasonably solid evidence of cross-sectional relationships between BART suggest that while the BART may not be a task to predict who will be risky in the future, its does a competent job of serving as a proxy of existing risk behavior in the real world. Towards this end, we have completed several studies that leverage this opportunity to understand how risk
taking is impacted by a range of external factors in the real world by manipulating those factors in a controlled laboratory environment. 

**Reward density and category.** The BART has one of the few studies examining the effects of varying cash reward magnitudes on RTP, as a function of individual differences in impulsivity (Bornovalova et al. 2009). Specifically, Bornovalova and colleagues manipulated the magnitude of reward/loss value, examining differences in BART score at 1, 5, and 25 cents per pump, as a function of trait impulsivity and sensation seeking. As the magnitude of monetary reward/loss value increased, risk-taking on the task decreased. However, when examined separately among individuals with high impulsivity/sensation seeking as compared to individuals low in these traits, the negative relationship between reward/loss magnitude and riskiness appeared most prominent among individuals low in impulsivity/sensation seeking. Conversely, individuals high in impulsivity/sensation seeking showed little change in riskiness as reward/loss magnitude increased. These findings illustrate that higher reward/loss magnitudes convey less risk-taking on the task, particularly for individuals low in impulsivity-related traits, suggesting that researchers should consider the value of reinforcers that are employed in behavioral risk-taking paradigms when interpreting results.

**Peer Influence.** Reynolds (2014) examined the effect of peer influence on BART RTP. Results of this study indicated that while no differences existed between the groups at baseline, RTP on the BART was significantly greater at a second experimental session for individuals who had peers encouraging their risk-taking as compared to individuals who completed the second session alone or with peers solely present, but not encouraging. In a second study, Cavalca et al. (2013) found that youth who were smokers (a proxy for higher levels of real world risk taking) were more influenced by peer pressure to be risky than youth who were not smokers. These data suggest that risk taking and the influence of peer pressure on risk taking can be assessed using the BART.

**Anxiety.** One particularly challenging question in the risk taking literature is the impact of anxiety. While anxiety is more generally thought to be negatively related to risk taking (e.g., Butler & Matthews, 1987), other evidence suggests that risk behaviors increase among those with anxiety in situations in which an anxiety-inducing stimulus is present (Kashdan & Hofmann, 2008). While this dichotomy makes intuitive sense it is difficult to test tease apart the disposition vs situation impact on risk taking of those with elevated anxiety. To address this issue, Reynolds et al. (2013) utilized an anxiety induction with a community sample of adolescents and compared performance on the BART in a high and low stress state (see the top of the Figure below for a diagram of the study). The high stress state included informing participants that they would be giving a speech at the end of the study and in the interim, they watched study confederates giving the speech and getting a difficult response from the audience; the low stress state included watching pleasant videos. Participants completed the BART before the low or high stress induction and again after the induction to examine the impact of the induction on risk taking as...
a function of anxiety level. Participants were partitioned based on their level of anxiety specific to social phobia (resulting in a high and low social phobia group). As shown in the bottom half of the Figure below, an interaction of group (low social phobia versus high social phobia) by condition was observed. Those in the high social phobia group had significantly increased risk-taking behavior from the control to the experimental session. Moreover, while the report of negative affect was similar for both groups in the low stress condition, the high social phobia group reported significantly higher self-reported distress during the high stress condition. This suggests that a task such as the BART can be useful in isolating the impact of anxiety inducing stimuli, and that for those showing greater impact of that stimuli a corresponding increase in risk taking behavior is observed.
Conclusion

In conclusion, data plays an important role in our understanding of real world risk-taking behavior. Ensuring the quality of that data requires an understanding of the rationale for the tasks developed and used. It also requires a clear sense of what useful existing data or new behavioral tasks can provide and where they fall short. The IGT and the BART assess risk taking in different ways, and together they may provide a strong comprehensive picture. Nonetheless, there is little data from any task suggesting strong longitudinal predictions. Thus, at least for now it seems clear that isolating risk taking in a controlled laboratory setting and providing the opportunity to manipulate key variables thought to impact risk behavior in the real world should be the focus on experimental studies. We have begun to conduct this work with the BART and have reviewed some early studies above, but considerable work including studies that bring in genetic factors and neural assessment (Bjork & Pardini, 2015; Gu, Zhang, Luo, Wang, & Broster, in press) as well as a wider range of environmental factors will be crucial to any further progress.

References


seeking. Pharmacology, Biochemistry, and Behavior, 93, 258-262.


