Journal of Economic Psychology 73 (2019) 15–33

Contents lists available at ScienceDirect

Journal of Economic Psychology

journal homepage: www.elsevier.com/locate/joep

Risky health choices and the Balloon Economic Risk Protocol

Kim Fairley[a,*], Jacob M. Parelman[b], Matt Jones[c], R. McKell Carter[c,d]

[a] Department of Economics, Leiden University, the Netherlands
[b] Annenberg School for Communication, University of Pennsylvania, United States
[c] Department of Psychology and Neuroscience, University of Colorado Boulder, United States
[d] Institute for Cognitive Science, University of Colorado Boulder, United States

ARTICLE INFO

JEL classification:
C91
C93
D81
D91

Keywords:
Risk
Beliefs
BART
Ecological validity
Healthy decision-making

ABSTRACT

We describe a risk protocol that combines the rigor of economic studies of risk with the ecological validity of tasks from psychology. Despite a wealth of experimental contributions on risk preferences, stemming from a variety of elicitation tasks, the external validity of standard measures of risk is questionable. In this study we focus on a risk task – the Balloon Analogue Risk Task (BART) – which is highly successful in predicting health-related risk behaviors such as alcohol use, drug use, smoking, unprotected sex, driving without a seatbelt, and stealing. The BART is not commonly used by economic scholars because of concerns that participants may not adequately comprehend uncertainty associated with the task and because of the resulting difficulty in relating participants' choices to standard risk models. To answer these concerns and build on associations with real world risk, we designed a modified BART, which we will refer to as the Balloon Economic Risk Protocol (BERP). In this protocol, participants observe the distribution of pop points prior to the task to create a more consistent knowledge base. We then use a belief elicitation technique to produce a user-generated prior distribution of balloon pops. Using these measures, we compare participants' behavior to the expected-value optimum to provide a link to standard models of risk. In accordance with past economic literature, we found that participants' BERP-generated risk preferences revealed mild risk aversion on average, and correlated with a self-report questionnaire on drinking, drug use, and smoking behavior.

1. Introduction

We address the ecological validity of risk measures in economics and show that our protocol, which uses building blocks from economics and psychology, is consistent with general observations of risk aversion and relates to health-related risk behaviors. Economics seeks to produce rigorous models that characterize such risk behaviors (Arrow, 1951; von Neumann and Morgenstern 1944; Tversky and Kahneman 1992), and so has developed numerous tasks designed to elicit risk preferences in people (Charness & Gneezy, 2012; Dohmen et al., 2011; Eckel & Grossman, 2002; Gneezy & Potters, 1997; Holt & Laury, 2002; Weber et al. 2002). This has allowed researchers to characterize individuals' variable appetite for risk. Nonetheless, despite a wealth of experimental contributions, risk preferences stemming from a variety of elicitation tasks are, at best, weakly correlated (Pedroni et al., 2017; Menkhoff & Sakha, 2017; Eckel & Wilson, 2004) and subsequently the ecological validity of these measures is uncertain. This study primarily focuses on this latter aspect of ecological validity.

* Corresponding author at: Faculty of Law, Institute of Tax Law and Economics, Department of Economics, Steenschuur 25, 2311 ES Leiden, the Netherlands.

E-mail address: k.fairley@law.leidenuniv.nl (K. Fairley).

https://doi.org/10.1016/j.joep.2019.04.005
Received 18 May 2018; Received in revised form 10 April 2019; Accepted 13 April 2019
Available online 25 April 2019
0167-4870/

Recently, the reasons underlying poor ecological validity of standard risk tasks have been the subject of a debate. Harrison (2015) argued that "out-of-sample prediction should be one criterion we use to evaluate our efforts, but not the sole metric or even the most interesting metric" (p. 124), another group of researchers (Eckel, 2016; Friedman, Mark Isaac, James, & Sunder, 2014; Trautmann, 2016) disagree and point to the wholesale failure of estimated risk parameters to predict behavior out of sample. According to Friedman et al. (2014), the main explanation of why risk methods have been unsuccessful in predicting behavior out of sample is the way that risk is measured in the economics literature: "the lottery paradigm with a focus on variation in outcomes does not relate to concepts of risk that people typically have in mind when making risky decisions". We seek to address those concerns by using a non-standard risk task, which does not rely on a lottery framework.

In this study we adapted the Balloon Analogue Risk Task (BART; Lejuez et al., 2002) to address concerns raised regarding its nature and procedures, and provide better out-of-sample reliability and external validity. In the standard BART, participants can pump a simulated balloon and accumulate earnings with every balloon pump. It is a task that can be characterized as a game and not as a standard lottery paradigm, giving it more similarity to the way participants conceptualize risk. Participants are aware that the balloon can pop after any pump, and that this event would result in the loss of any accumulated tokens. They secure their earnings by cashing in (i.e., deciding to stop pumping the balloon). In past studies, their earnings were saved to a bank which held all saved earnings from previous trials. Participants' willingness to tolerate risk was operationalized as the average number of pumps before cashing in (labeled as "adjusted pumps" in Lejuez et al., 2002).

Although the BART is not commonly used by economic scholars, this task is highly successful in predicting health-related risk behaviors in the clinical domain. The metric of average number of pumps before cashing in, in the BART, has been found to correlate with alcohol use, drug use, smoking, unprotected sex, driving without a seatbelt, and stealing (Aklin, Lejuez, Zvolensky, Kahler, & Gwadz, 2005; Hopko et al., 2006; Lejuez et al., 2003a, 2003b, 2002). Risk behavior elicited from the BART has also enabled researchers to distinguish psychopathic patients from healthy controls (Hunt, Hopko, Bare, Lejuez, & Robinson, 2005), cocaine users from heroin users (Bornovalova, Daughters, Hernandez, Richards, & Lejuez, 2005), and adolescents with serious conduct and substance disorders from healthy adolescents (Crowley, Raymond, Mikulich-Gilbertson, Thompson, & Lejuez, 2006). In contrast, a study attempting to relate a standard economic risk and ambiguity task with health-related behaviors in adolescents (Sutter et al., 2013) has proved unsuccessful. There is at least one example of a study (Anderson & Mellor, 2008) that showed a significant negative relationship between risk aversion, as measured by a Holt and Laury lottery design (Holt & Laury, 2002), and health-related behaviors such as drinking, smoking and seatbelt non-use, but these results have not been replicated. In contrast, numerous associations between the BART, and health-related risk behaviors have been successfully replicated (Aklin et al., 2005; Hopko et al., 2006; Lejuez et al. 2003a, 2003b, 2002).

Despite these successes, the BART is not rigorously defined and suffers from multiple implementation flaws. Some of our concerns were also raised by others in meta-analyses of economic risk tasks (Charness, Gneezy, & Imas, 2013; Deck, Lee, Reyes, & Rosen, 2013). We first describe these criticisms of the BART and then present a modified BART executed within a standardized protocol, which we will refer to as the Balloon Economic Risk Protocol (BERP), designed to address them, while we aim to preserve the BART's ecological validity.

The first criticism we address is the participants' understanding of uncertainty in the BART, stemming from an absence of objective probabilities related to balloon pops. As a result, participants have no clear understanding of the balloon's underlying risk profile. In the original study (Lejuez et al., 2002), participants were told only that the balloon could potentially grow to the full size of the computer screen, but this information was never visually displayed or explained in terms of number of pumps. Participants also likely use some subjective mental model of balloon pop points that does not match the uniform distribution of pop points from the BART. These factors make it difficult for participants to form any prior regarding the distribution of balloon pumps. We address this shortcoming by providing participants the opportunity to observe the distribution of balloon pop points.

In order to help participants establish a prior and facilitate their understanding of the risk they would face in the BART, we began our protocol by showing participants 16 balloons being pumped until they popped. This procedure follows previous work illustrating the power of learning by experience versus learning by description (Hertwig, Hogarth, & Lejarraga, 2018). To test if participants had learned from the balloon pop observation, we employed a novel belief elicitation task (similar to Harrison, Martínez-Correa, Todd Swarthout, & Ulm, 2017) that asked them to report their beliefs regarding the maximum balloon size and to bet on where they expected balloon pops to occur. This provided a subjective measure of participants' understanding of the risk they faced in the experiment and an opportunity to assess whether participants pumped balloons in a manner that was optimal according to their beliefs.

Second, we sought to address concerns about the BART's incentive structure, which included a lack of incentive compatibility, deception, potential portfolio effects, and no disincentive associated with popping. For instance, in one study (Lejuez et al., 2003a), participants were told that if they accumulated earnings above an undisclosed amount, they would receive two movie tickets. At the end of the study however, all participants were rewarded the movie tickets 'to prevent disappointment'. Another study (Lejuez et al., 2003b) told participants their earnings would be compared to those of other participants and, depending on their standing, they would receive a corresponding monetary reward. In reality, individual scores were never compared to others and all participants were given the same reward. Furthermore, in all studies the earnings were based only on trials wherein participants had cashed in, leaving out trials in which individuals popped and earned nothing. Finally, participants saw a running total of accumulated earnings throughout the experiment, which could have affected their game play throughout the course of the study. We sought to make the task consistent with established best practices (Kagel & Roth, 2016; List et al., 2011).

The incentive scheme in the BERP is non-deceptive and monetarily based. Incentives consisted of the accuracy, assessed by the quadratic scoring rule (Artinger, Exadaktylos, Koppel, & Sääksvuori, 2010), of participants' belief rating in balloon pops, and one random draw of all the trials they played in the BERP (including popped balloons). Regarding this last component, participants were made explicitly aware of the possibility that bonus earnings could be 0 if the computer randomly drew a popped balloon. As a consequence of the random-lottery incentive scheme chosen – drawing a random balloon trial from all decision trials for each participant at the end of the experiment – participants' earnings were not tracked for each and every trial nor were they shown as a running total (bank) to participants during the actual experiment, as was the case in standard procedures of the BART (Lejuez et al., 2002). Our procedure helps to avoid hedging strategies as a result of wealth effects (a participant might respond differently on the basis of amount won on a previous decision). A potential risk of using a randomization device choice is that participants treat a random-lottery experiment as a single choice problem, and if participants violate the reduction axiom, participants' true risk preferences are not elicited (Kaivanto & Kroll, 2012; Starmer and Sugden, 1991; Cubitt, Starmer, & Sugden, 1998). In the BERP, participants make one unique choice (where they want to cash in the balloon) and each trial is independent from the other (the balloon has a new pop point), making it very unlikely that results may be biased by violation of the reduction axiom.

Lastly, we address an inability to formally establish individuals' risk preferences in the BART. In this study, we sought to link behavior in the BERP to common risk-preference findings from other tasks often used in experimental economics. Previous work employed balloons with the same minimum balloon size of 0 pumps – the balloon can pop on the first pop –, but with several maximum balloon sizes – 8 pumps, 32 pumps and 128 pumps – and importantly only the largest range had a relationship with health-related risk behaviors (Lejuez et al., 2002). The optimal behavior, in terms of maximizing expected value, related to a maximum balloon size of 128, is to pump to 64 (see BERP's risk profile in the methods section for further explanations). However, participants' mean behavior on cashed trials – which is the common practice to elicit individuals' risk behavior in the BART – was 30.5 pumps for men and 25 pumps for women (Lejuez et al., 2002). This behavior is far from optimal and thus implies quite extreme risk aversion. This issue is not discussed in Lejuez et al. (2002) or any follow-up studies (Lejuez et al., 2003a, 2003b). This deviation from widely replicated findings on mild risk aversion for risk preferences (Crosetto & Filippin, 2016) may be related to the previous concern that participants in past tasks have not understood how pump mechanics related to popping and can be addressed by allowing participants to observe balloons popping as described above. We then compare participants' number of pumps on cashed trials in the BERP to the optimal behavior, in terms of maximizing expected value, to properly assess their risk preferences and compare our behavioral results to other risk tasks.

Besides the common practice to take only cashed balloons into account, we will also consider two different risk metrics as a robustness test. The first is participants' cashed tokens on their <u>first</u> cashed balloon. This allows us to assess participants' risk preferences without the influence of their history of game play – also referred to as serial autocorrelation – which might influence their appetite for risk over time. Secondly, we correct participants' tokens on cashed balloons using their beliefs regarding balloon pop points as an alternative risk metric. That is, a participant who, on average cashed in after 30 pumps, but believed the balloon would pop after 25 pumps, could be, subjectively, classified as risk loving with a high risk perception, whereas another participant who cashed in after 40 pumps and believed the balloon would pop after 45 pumps is rather risk averse with a low risk perception. Finally, to check for consistency within participants, we also elicit risk and ambiguity parameters with a multiple-choice list procedure (Sutter et al., 2013).

In designing the BERP we aimed to utilize the ecological validity of the standard BART in predicting participants' health-related risk behaviors, while adapting the protocol of the task in such a way that participants have a better understanding of the risk they face in this game, and also adhering to best practices in experimental economics.

We describe results from the BERP supporting our goals of producing a more rigorous BART-inspired risk task that retains ecological validity. First, we show that participants did not believe the distribution of balloon pops was uniform, yet most participants reported a maximum balloon size close to the actual maximum and the average of pop-point beliefs close to the actual distribution mean. Subsequently, we illustrate that participants' mean beliefs in pop points were related to their behavior in the task. The participants who pumped more, believed the balloon pop points occurred more frequently at higher levels. Second, we show that, consistent with the large literature on risk preferences, on average participants showed mild risk aversion. Third, we confirm the ecological validity of the BERP, as the higher participants pumped a balloon on average, the higher they scored on a self-reported questionnaire on problematic drinking, the more drugs they used, and the more likely they were to smoke. Finally, we show that neither risk nor ambiguity preferences elicited by a multiple-choice list procedure correlated with risky health behavior.

In the next section we review some additional literature, followed by an outline of the methods, then the results in the third section, and finally a discussion in the last section of this paper.

2. Literature background

We adapt the standard BART by implementing a new protocol we refer to as the Balloon Economic Risk Protocol (BERP) while aiming to maintain ecological validity. Several economic studies have qualitatively and quantitatively compared the BART to other risk elicitation tasks or developed a new experimental risk task based on BART features. This short literature background provides an overview of these studies and their findings. Subsequently we will explain how it has guided us in developing the protocol for the BERP and our choices to retain specific features, based on the ecological validity of the standard BART, in the version used in this study.

Deck et al. (2013) compared a single-shot BART – participants only played one balloon - with other risk tasks and found a small significant correlation with another dynamic task Deal or No Deal, but not with the other more static lottery paradigms. Compared to the other tasks, the subjects were quite risk averse in this version of the BART, but that conclusion might be (as Deck et al. note on p. 12) rather misleading as participants who popped their balloon could not be properly characterized.

More recently, Pedroni et al. (2017) compared six risk tasks, including the BART (operationalized as a balloon with a maximum size of 128 pumps as in Lejuez et al., 2002), and found extreme risk averse behavior (not a single participant could be indicated as risk seeking). Risk preferences elicited by the BART correlated significantly, but weakly, with the risk tasks referred to as adaptive lotteries (Spearman's $r = 0.11$), multiple price list (Spearman's $r = 0.11$) and the Columbia card task (Spearman's $r = 0.14$).

Charness et al. (2013) qualitatively reviewed several risk tasks and made numerous remarks regarding the BART. On the positive side, they point out that the context in which the game is played is easy to grasp. Related to this point, they state that simple methods, as opposed to complex risk methods which demand more understanding and mathematical sophistication, are most useful for trying to capture treatment effects and differences in individual risk preferences. On the other hand, they question whether the BART aligns with other economic risk tasks and other non-health related decision domains, like decisions from the financial domain, which are of particular interest to economists. Lastly, the fact that the BART requires multiple trials on a computer is seen as a drawback when time plays a role or for field work when access to computers is limited.

Apart from these reviews, there is also one particular task in economics, which is very similar to the BART. In 2013, Crosetto and Filippin introduced the Bomb Risk Elicitation task (BRET) as a new intuitive risk task in economics. The main features of the BRET are very similar to the BART. The BRET replaced balloon pumps and a potential pop with boxes and a potential bomb. In contrast to the BART, the BRET was converted to one single trial: hit the 'stop' button to let the program stop turning boxes automatically, with delayed 'bomb' feedback. Although the automatic program might have already opened a box with a bomb before the participant decided to stop, that information is only revealed after participants decided to stop opening further boxes. Apart from the delayed outcome feedback, participants were well aware of the maximum number of 100 boxes in the study by Crosetto and Filippin (2013). A visual display indicated the number of boxes opened by participants, the number of remaining boxes unopened and the total monetary value collected so far, conditional on not having opened a box with a bomb. In the BART, participants know how many times they pumped the balloon, but not how many pumps are remaining as the maximum balloon size is unknown (although they could assess this value during the balloon pop observation).

In a later study, Crosetto and Filippin (2016) compared the BRET with three other well known risk tasks and found that all tasks were behaviorally uncorrelated. 73.50% of all participants who played the BRET could be classified as risk averse, which was just a bit lower compared to the other risk tasks. Crosetto and Filippin (2013) did not aim to replicate the BART's success of predicting participants' health-related risk behaviors. It is therefore unknown whether behavior in the BRET might be associated with health-related risk behaviors.

After having focused on issues with the standard BART and the alternative BRET, we now focus on factors that might contribute to the high ecological validity of the BART in explaining risky health behaviors. It is likely that the essential game-like features of the BART align with the behaviors it attempts to capture. In the BART, participants experience diminishing returns and increasing hazard. Namely, with each successive pump, the amount that could potentially be lost increases and the expected gain decreases. The idea of a one-shot decision, which is very common in economic risk tasks, probably does not line up with the reality of many risky health decision-making contexts. Even if people a priori establish a risk threshold, it is known that people, for instance in the context of alcohol abuse, often violate their self-imposed limit (Muraven et al., 2005a, 2005b).

Furthermore, the BART is an evolving, continuous and dynamic decision-making task. In dynamic situations, decisions are incremental and increase over time (Figner & Weber, 2011). The risk of an initial action – e.g. consuming alcohol – is relatively low, but increases – the risk of getting intoxicated – over repeated choices. This increasing risk dynamic aligns with many risky health scenarios in which the decision maker will first experience a positive outcome, as the probability for a negative outcome is relatively low in the beginning, which in turn reinforces the likelihood of accepting additional risk, hence contributing to the difficulty of stopping when the risk increases to a point the decision maker would, without the initial positive outcomes, stop (Figner, Mackinlay, Wilkening, & Weber, 2009).

Although the BRET is much more intuitive than an average lottery task, participants are not opening boxes themselves and outcome feedback is delayed in a context which is less uncertain than the BART. Therefore, the risk dynamic we described above in relation to risky health choices does not hold in the BRET. There is no experience of a positive outcome with turning boxes as the participant does not know if she already faced a bomb and there is less difficulty to stop turning boxes if there is no experience of subsequent positive outcomes, which, if present, has been shown to reinforce risk taking (Figner et al., 2009). Moreover, immediate outcome feedback is likely an important feature of maintaining consistency with many risky health choices. This is supported by studies showing that tasks with immediate outcome feedback trigger stronger physiological responses, as measured by skin conductance, than tasks which delay outcome feedback after the task is completed (Figner & Murphy, 2011). Finally, strictly speaking, the BART and the BERP do not capture risk as strictly defined by clear probabilities and their outcomes. We chose to let participants form a prior based on the balloon pop observation, instead of providing the distribution a priori. We expect that this greatly reduces the uncertainty as compared to the original BART, but corresponds to the reality of many risky health scenarios in which decision-makers have priors, but not exact probabilities of corresponding health outcomes.

Not all risky health choices provide immediate health feedback. For instance, lung cancer and cirrhosis of the liver might appear decades after the onset of smoking and problematic alcohol use. In this study, when we discuss the risk dynamics associated with risky health choices, we refer to negative outcomes such as alcohol intoxication, drugs overdose, and shortness of breath to name just a few immediate to short-term effects of substance use and smoking.

Altogether, the BART thus differs from economic lottery risk tasks in that it elicits dynamic risk taking with immediate balloon pop feedback, in a context not precisely defined by probabilities, but in which the participant understands that excessive risk taking leads to diminishing returns and greater hazards.

Therefore, despite concerns described by Charness et al. (2013) and Deck et al. (2013) over the repeated decisions in the BART, we did not convert the BERP to a single shot decision and/or remove the immediate outcome feedback as with the BRET. We believe these elements subsequently elicit increasing risk dynamics (Figner & Weber, 2011) and physiological affect (Figner & Murphy, 2011) that make the standard BART successful in predicting out-of-sample health behavior. Further details regarding the adaptations made in the BERP are described in the Methods section below.

3. Methods

3.1. The BERP

We include the BART as part of a protocol that provides careful instruction, explicit observation of the balloon pop distribution, and a belief elicitation task designed to measure participants' understanding of the balloon pop probabilities.

3.1.1. Instructions

In the standard BART, participants pump a balloon and accumulate money on every balloon pump. In our study, with every balloon pump participants accrue tokens. One token corresponds to $0.20. This conversion is conveyed to participants during the instructions. In the instructions, we refer to this game as the Balloon pump game to participants.

Ordinarily, participants are not informed about the maximum balloon size and are only told that the balloon can pop at any time. Thus, they are, with respect to number of pumps, completely unaware of the underlying risk profile of the balloon. In the BERP, as part of our instructions, participants first learned the essential features of the game before we displayed 16 balloons, one at a time, growing before they would pop. The respective pop points [1, 4, 8, … , 60, 64], presented in a random but fixed order across all participants, covered the full distribution of balloon pop points in steps of 4 (after the first step).

A complete transcript of the instructions can be found in Appendix A.

3.1.2. Belief elicitation procedure

After observing the series of balloons popping, we asked participants to report their beliefs regarding the maximum balloon size (how many pumps the balloon could maximally handle) in this experiment. Then we elicited their beliefs regarding the balloon's pop point by asking participants 'where do you expect pop points to occur, relative to the maximum possible balloon size?'. Participants could indicate their beliefs by distributing at least 100 points on a scale that indicated the balloon's size from its smallest size to its maximum size in 20 equal bins (Fig. 1). We carefully explained to participants that the more points they distributed to a specific bin, the more balloons they expected to pop at a certain balloon size (expressed as a percentage of their reported maximum balloon size). Participants could easily drag bars up and down, by holding down the computer mouse, and dragging the cursor from left to right, the bars would automatically rise to facilitate the drawing of a distribution. Participants were told they would be paid for the accuracy of their belief rating.

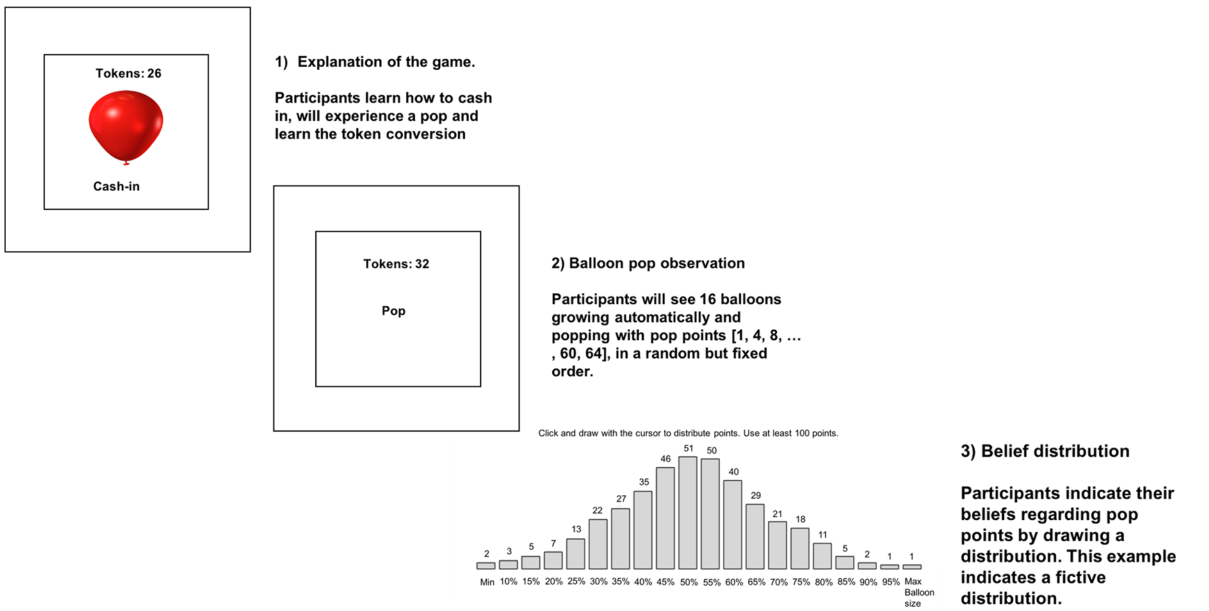

Fig. 1. Schematic overview of the important elements of the BERP.

We paid participants for the accuracy of their beliefs by applying the quadratic scoring rule for a multiple-choice setting for subjective probability distributions (Harrison et al., 2017; Artinger et al., 2010).[1] The number of points a participant distributes to a specific bin on the scale relative to all points used for their distribution $p = (p_1, \dots, p_n)$ is formally expressed as p_i $(1 \le i \le n)$. Each p_i thus refers to the participant's expected probability with which pop point i occurs. The 'correct' event, indexed as *j*, was established at the end of the experiment for every participant by randomly drawing one balloon pop point from all the balloons they played during the experiment. Finally, each participant's score was calculated with the quadratic scoring rule:

$$Q_j(p) = (\alpha + \beta) - \beta(1 - p_j)^2 - \beta \sum_{i \neq j} {p_i}^2$$

Participants start with an individual endowment of $\alpha + \beta$ (both set to 10 tokens in our experiment) and are penalized for not assigning maximum probability to the event that occurs and placing a positive probability on events that do not occur. If a participant places all points on the correct event *j*, they win 20 tokens. If they place all points on other events other than *j*, they win 0 tokens. They win a positive amount between 0 and 20 tokens if they place some points on the correct event *j*. Like other studies (Qiu and Weitzel, 2016), we did not explain the formula to participants, but explicitly told them that the more points they distributed to their belief regarding the correct pop point, the more they would be financially rewarded.

Beliefs were only elicited once, as part of the instructions, before the game started. After the belief elicitation, participants had ample time to ask final questions before the game would start. Please see Fig. 1 for a schematic overview of the important elements in the BERP.

3.1.3. BART

Finally, we discuss several important game features of the BART and how we implemented these as part of our BERP.

Several studies have varied the maximum balloon size used in the BART, e.g. maxima of 8 pumps (Lejuez et al., 2002), 12 pumps (Fukunaga, Brown, & Bogg, 2012; Rao et al., 2008; Schonberg et al., 2012), 32 pumps (Lejuez et al., 2002) and 128 pumps (Lejuez et al., 2002; Crowley et al., 2006; Aklin et al., 2005). In this study the maximum balloon size is 64.

The balloon can pop at any time between the first pump and the maximum balloon size. This means that the balloon can pop on the first pump and will pop with certainty at the maximum balloon size. On each trial the pop point is randomly drawn from a uniform distribution between one and the maximum balloon size. Participants can 'cash in' their accumulated tokens before the balloon's pop point.

Each trial starts with a small simulated balloon, which can be pumped at any pace participants prefer by pressing on a key for a single balloon pump or holding down the space bar for continuous balloon pumps, with a growth rate of 0.1 s per balloon pump. Balloon pumps are accompanied by a token counter placed above the growing balloon. When the balloon pops, the token counter will display 0 tokens and the trial ends. When the participant decides to cash in, a green square surround both the balloon and the token counter to indicate a 'win' on this trial.

3.2. Economic risk and ambiguity

In order to compare behavior in the BERP to economic measures of risk and ambiguity, we elicited risk and ambiguity preferences via a multiple-choice list procedure, as per Sutter et al. (2013). In both the risky and ambiguous settings, participants chose on each trial between a gamble and a sure payoff, with the latter increasing in size after each choice.

Risk and ambiguity were operationalized via an Ellsberg procedure (Ellsberg, 1961). In our version there were two (actual) bowls, which were both filled with 10 paper slips. One bowl, the risky prospect, was filled with exactly 5 orange and 5 white paper slips. The other bowl, the ambiguous prospect, was filled with an unknown composition of white and orange paper slips. Beforehand, participants chose either white or orange as 'their' color, with this choice determining the specific paper slip color that would lead to a win for them. It is an important feature that participants could choose their own winning color in order to (correctly) eliminate any suspicion that the likelihood of a win/loss from the ambiguous bag may have been predetermined (Wakker, 2010).

For both risky and ambiguous prospects, participants saw a total of 20 choices each. Each choice was between gambling on the draw of a winning colored paper slip from the bowl or receiving the sure payoff. A winning colored paper slip would earn participants 5 tokens, or else nothing. The sure payoff increased from 0.25 to 5 tokens in 20 equal steps of 0.25 tokens. A risk neutral participant should switch once, from gambling to the sure payoff after 10 rows. We did not force participants to switch once however and this allows us to perform robustness analyses on participants who switched once versus multiple times. For both choice lists, we counted the number of gambles each participant was willing to engage in. The more they were willing to gamble, the less participants were interpreted to express risk/ambiguity averse behavior. The order of presenting the full choice lists for the risky and ambiguous prospects was counterbalanced across participants to control for any potential order effect.

After the experiment, the computer randomly drew one choice from either the risky or ambiguous prospect. If participants chose to receive the sure payoff in the randomly selected choice, they were paid the sure amount and the task ended. If they decided to gamble in that selected choice, they were allowed to draw one paper slip, without looking, from the corresponding risky or ambiguous bowl. If the drawn paper slip matched the color they selected at the beginning of the experiment, they would win 5 tokens. If the paper slip color did not match their selected color, they would win nothing.

[1] Harrison et al. (2017) provide empirical support for not needing to adjust beliefs in light of participants' risk aversion, if one assumes expected utility, when eliciting subjective belief distributions over continuous events.

3.3. Questionnaire

After the decision-making tasks, both the BERP and the economic risk and ambiguity tasks, participants were routed to a questionnaire, which was programmed with and hosted on Qualtrics. The questionnaire consisted of 12 parts and took an average of 15 min to complete. The following surveys were administered in their respective order to participants.

First, we measured potential hazardous drinking with the Alcohol Use Disorders Identification test (AUDIT, Saunders et al., 1993). In accordance with the literature, participants' responses to this 10-item measure were summed in order to assess individuals' drinking behavior.

Following the AUDIT, we administered the Brief Sensation Seeking scale (BSSS; (Hoyle, Stephenson, & Palmgreen, 2002), which is a brief, validated version of the original Zuckerman sensation seeking scale (Zuckerman et al., 1964). We scored this measure following the normal convention of summing all (reverse scored) items.

The third part consisted of the 10-item measure based on the Centers for Disease Control Youth Risk (US Department of Health and Human Services; Centers for Disease Control (CDC) 2004). Specifically, participants had to respond with yes or no to engagement in risky behaviors, like gambled for real money, smoked a cigarette, stolen anything from a store, ridden in a car without wearing a seatbelt, over the past 12 months. We summed the 10 items, but also analyzed each behavior separately as a dichotomous variable as the risky behaviors differed quite substantially.

Subsequently, we measured participants' drug use, by examining polydrug use (Grant, Contoreggi, & London, 2000), which is defined as the number of drug categories tried over the past 12 months: marijuana, stimulants, cocaine, hallucinogens, opiates, sedatives and 'other'.

Following drug use, we asked questions selected from the HIV risk taking behavioral scale (Ward, 1990) regarding sexual behavior. We did not follow the exact questions of the original scale as we thought some questions were too intrusive for this study of a population with an expected typical HIV base rate. We primarily focused on more general aspects of behaviors around sexual intercourse. For example, number of sexual partners last year, condom use and whether sexual partners discussed potential risks of STD transmission.

The following measures were all standard questionnaires which were administered and scored according to the original studies. Respectively, participants answered questions for social risk (Weber et al., 2002), impulsivity as elicited via the Barratt impulsivity scale (Barratt, 1985), peer pressure inventory for misconduct (Clasen & Brown, 1985), resistance to peer influence (Prinstein et al., 2011) and ambiguity tolerance (Budner, 1962).[2]

Finally, we asked demographic questions such as age, gender, and education and offered participants an opportunity to describe their game strategy in some open-ended questions.

3.4. Procedures

The laboratory experiments took place at the Center for Innovation and Creativity at the University of Colorado Boulder. The experiment was programmed with PsychoPy (Peirce, 2007). Our experimental sample consisted of 156 participants (44% female, mean age 20.79 years). Session sizes varied between 4 and 10 participants. This experiment was part of an fMRI experiment wherein one participant in every session made decisions while lying in an MRI scanner. Both behavioral and MRI participants received extensive instructions together before we placed the MRI participant in the magnet. After the decision-making portion of the study, the MRI participant was again seated in the behavioral laboratory to complete the lottery measures of risk and ambiguity and the questionnaire on real-life risky decision-making with the behavioral participants.

A session took about 2 h for behavioral participants. On average, participants earned $26.38. Earnings were determined based on a fixed fee for participation ($10 per hour) and bonus payments for three elements in this experiment: belief elicitation regarding the balloon's pop point, one random trial from the decision-making in the BART, and one random decision in the risk or ambiguity lottery. Payments were calculated and given to participants at the end of the study.

As part of this research project, we also designed a social version of the BERP and in an accompanying paper, we will describe social modulations of risky decision-making and their underlying neural correlates. In this manuscript we solely focus on behavior in the individual BERP setup and its relationship with real-life risky decision-making. Participants played a block of the BERP setup by themselves (12 decisions) before they were exposed to any other variations of the BERP. Throughout the remainder of the experiment, participants would play the BERP by themselves if no partner was available to play a social game. In this study, we analyzed, per participant, all individual BERP trials as a characterization of their individual risk preference.

All the results we present in this paper are robust against type of participant, MRI or behavioral, and if we restrict participants' decision-making to the initial block of individual play.

3.5. Optimal behavior in the BERP

In this paragraph, we explicitly define optimal behavior in the BERP before we relate it to participants' decision-making in the BERP in the results section. In the BERP, there is a clear distinction between the probability the balloon will pop at every balloon

[2] The social related questionnaires - social risk, peer pressure, resistance to peer influence - are not discussed in this paper as these will be used for the social adaptations of the BERP.

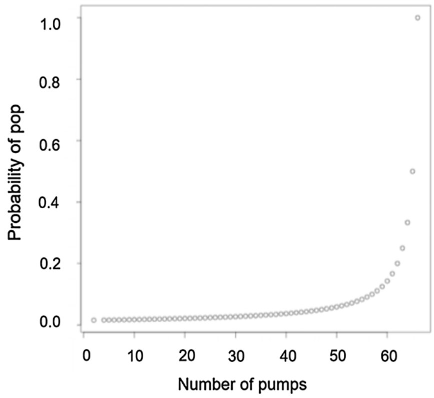

Fig. 2. Hazards of balloon pop points.

pump within one trial and the average pop point across all balloons participants play during the experiment. As the balloon pop point is drawn from a uniform distribution between 1 and 64, each number has an equal chance to represent a balloon's pop point. Therefore, across all balloons the cumulative distribution function is linear, as each balloon pop has a probability of $\frac{1}{n}$ to be randomly drawn. The objective mean pop point is thus 32.5. During one trial however, the hazard of a balloon pop, meaning the probability of a pop on the next pump for a balloon that has not popped yet, grows steadily from $\frac{1}{n}, \frac{1}{(n-1)}, \frac{1}{(n-2)}$ and so forth. If the balloon has not yet popped at the 62nd pump, the chance of popping is 50% at the subsequent pump, and at the 64th pump the balloon will pop with certainty (see Fig. 2 for this function).

The expected value (EV) can be estimated by multiplying the number of tokens accumulated with every balloon pump x and the probability the balloon will not pop at any corresponding balloon size. Formally, this can be written as: $x * \frac{n-x}{n}$. In Fig. 3 the EV is plotted for each balloon size. This bow-shaped function illustrates that the maximum expected value is 16 tokens at 32 pumps. The choice to pump again carries a gain of 1 with probability (n-x-1)/(n-x) and a loss of x with probability 1/(n-x), for a net incremental EV of (n-2x-1)/(n-x). Therefore, pumping has positive incremental EV if x < (n-1)/2 = 31.5 and negative incremental EV if x > 31.5. So a 32nd pump (i.e., pumping again after 31 pumps) is rational, but a 33rd pump is not. We therefore define risk neutrality as mean cash in behavior between 31.5 and 32.5.

4. Results

All raw data and scripts can be downloaded from: https://github.com/SNaGLab/Risky-health-choices-and-the-Balloon-Economic-Risk-Protocol.git.

4.1. Beliefs

Before the experiment started and after the balloon pop observation, participants indicated their beliefs regarding balloons' pop point. This allowed us to infer whether participants established a prior in line with the objective mean pop point of 32.5 in the BERP.

We first inspected the shape of participants' drawn distribution. In Fig. 4, for each balloon size bin, we depicted a boxplot representing participants' distributed points to that particular balloon size. The figure is affected by some outliers who have ratios as high as 1 (which means this participant made a single point prediction by placing all points on a single balloon size bin).

Please see Fig. A1 in Appendix for a representation without significant outliers. On average, participants do not believe balloon pops stem from a uniform distribution as one can see from the shape of Fig. 4 and by the results of the Kolmogorov-Smirnov test

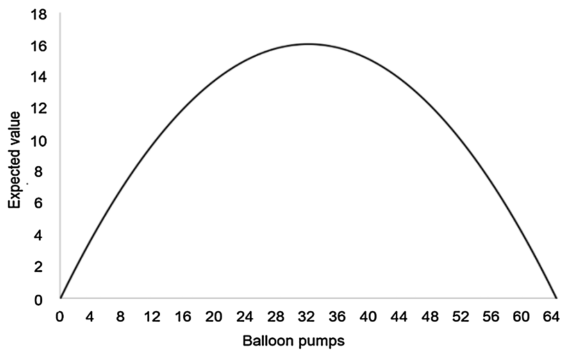

Fig. 3. The expected value as a function of all potential numbers of balloon pumps.

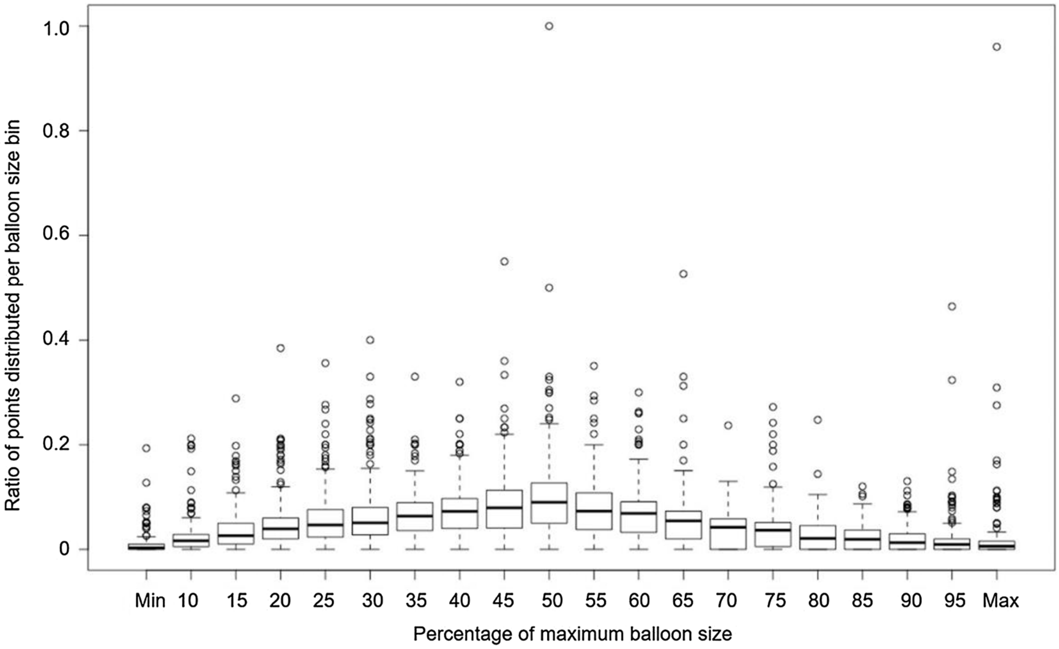

Fig. 4. Boxplots of participants' distributed points per balloon size bin.

against a uniform distribution (D = 0.168, $p = 0.007$).

Subsequently, we calculated the mean of each participant's belief distribution by tailoring the balloon size to every participant's maximum balloon size estimation.[3] In Fig. 5, a histogram shows participants' mean belief estimations regarding balloons' pop point. The mean of belief means is a balloon pop value of 32.63 (indicated by the blue dashed line in Fig. 5). The sample as a whole had a good understanding of the balloon's underlying risk profile, but some individual participants were far from the correct mean.

4.2. Risk preferences in the BERP

On average, participants pumped balloons 27.7 times before they decided to cash in (only cashed balloons are taken into account). This implies that the average risk behavior in the BERP can be characterized as risk averse as it is below the risk-neutral behavior of pumping between 31.5 and 32.5. However, the mean number of pumps on cashed-in trials varied considerably across participants (σ = 7.38 pumps, min = 9.08 pumps, max = 46.83). 69% of all participants (n = 109) can be classified as risk averse, 5% as risk neutral (n = 8) and 26% as risk seeking (n = 41). Although males pumped balloons slightly higher (28.06) than females (27.44), this difference was not significant ($p = 0.618$). Age ($p = 0.474$) and educational attainment ($p = 0.740$) also had no effect on mean cash-in behavior in the BERP.

One of the critiques of including all cashed balloons in these analyses, is that participants might alter behavior during the BERP as a function of history of balloon pops and cash-ins. In Fig. 6 we plotted four participants' history of play. These are selected to indicate several behavioral patterns we observed in the BERP. Participant 59 was not affected by the history of play and aimed to cash in 26 token on every trial. Participant 64 played quite random and decided to cash in at various levels, without experiencing a lot of balloon pops. Participant 51 pumped more after cashing in a balloon, while pumping less after a previous balloon pop, whereas participant 53 primarily responded to balloon pops by trying to pump higher after popping a balloon.

In order to extract participants' risk behavior in the BERP and mitigate the effects of play history, we took participants' cashed in tokens on the first balloon they cashed in.[4] We will refer to this variable as 'first cash' in subsequent Tables. On average, participants pumped their first cashed balloon 23.94 times, indicating that, on average, participants increased their risk over time, as the number of pumps for all cashed balloons is more than 3 pumps higher. Nevertheless, the ordering of participants' risk behaviors was quite similar between both measures. Risk as measured on the first cashed balloon in the BERP strongly correlated with all cashed balloons

[3] Most participants correctly indicated that the maximum balloon size was (close to) 64. A few indicated 100. If these individuals expected the balloon to pop at a high balloon size, their mean belief estimation regarding balloon pops could be higher than the actual maximum balloon size of 64 pumps. In Fig. 5 there are two such outliers.

[4] One could argue that participants' first cashed in balloon is affected negatively when they faced (a) pop(s) on previous balloon(s), and importantly that risk seeking participants will disproportionally face these pop events more often as compared to risk averse participants. Our robustness analyses showed that the number of pops that occurred before participants cashed their first balloon was not related to the mean cashed in tokens during the first balloon (Pearson's $r = -0.02$, p = 0.849) nor the overall mean of cashed balloons (Pearson's $r = 0.07$, p = 0.356), ruling out an effect of the number of pops experiences before the first cashed in balloon and risk preferences.

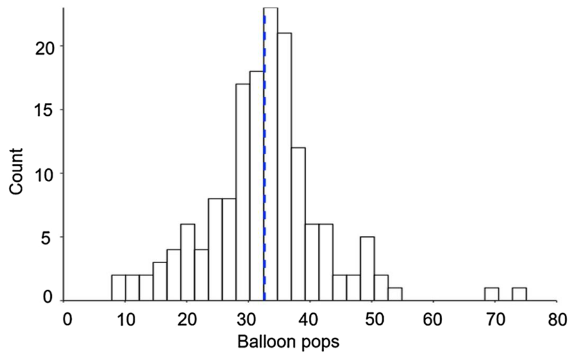

Fig. 5. Histogram of participants' mean belief for balloons' pop point (blue dashed line indicates the mean).

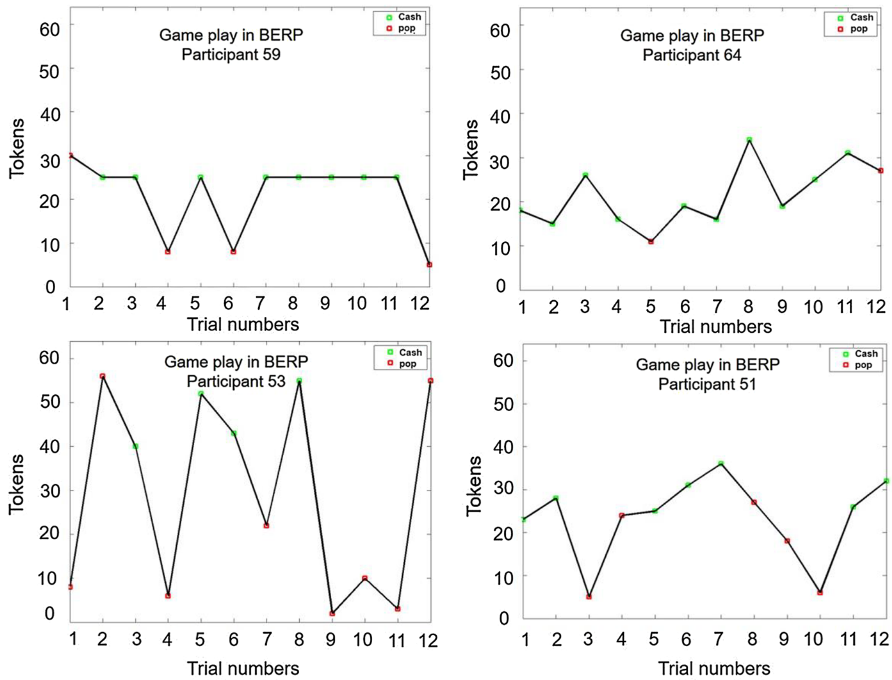

Fig. 6. Four participants' history of play as illustrations of various game play strategy.

(Pearson's $r = 0.74$, $p < 0.001$).

Finally, although participants did not perfectly learn the uniform distribution of pop points, their beliefs about the distribution of pop points can be used to estimate whether they on average held short of what they believed to be the mean pop point or continued to pump past it. Although Fig. 5 clearly showed that some participants have much lower or higher beliefs than the objective mean pop point of 32.5, these variations in belief related to subsequent pump behavior in the BERP. The higher a participant's estimation of where the balloon was most likely to pop, the higher that participant would pump the balloon on cashed balloons (Pearson's $r = 0.32$, $p < 0.001$, Fig. 7).[5]

Also with regard to first cash, participants' mean belief regarding balloon pop points correlated with cashed tokens on the first balloon (Pearson's $r = 0.28$, $p < 0.001$). We therefore adjusted individual's risk behavior in the BERP by taking individuals' belief in balloon pops into account. This variable 'risk-belief' was calculated by subtracting participant's mean belief, as extracted from their distribution, from their average cashed tokens on cashed trials in the BERP. The mean of risk-belief was -4.86 ($\sigma = 10.18$, min $= -46.81$, max $= 16.18$) indicating that, on average, participants cashed in below their belief regarding the balloon pop point (also illustrated by the majority of participants who fall below the 45 degree line in Fig. 7). Instead of comparing participants' mean cashed tokens to risk-neutral behavior, in terms of maximizing expected value, we considered participants risk averse when they

[5] If we adjust the maximum balloon size to 64 for *all* participants, we find a similar positive correlation between individual beliefs and pump behavior in the BERP, indicating that our result is driven by participants' subjective belief distribution, rather than their maximum balloon size estimation.

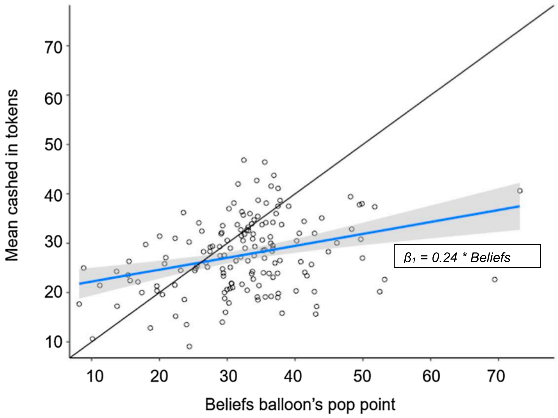

Fig. 7. The blue line (with 95% confidence region indicated by the grey area) illustrates the linear relationship between participants' mean belief in balloons' pop point and their pump behavior in the BERP (estimated via geom_smooth in r's ggplot package). Participants above (below) the 45 degree line in black are indicative of those participants who pump higher (lower) than their belief regarding the mean pop point.

cashed in lower than their beliefs in balloon pop points, risk neutral when this difference was less than 0.5 – as it will be the same rounded number of balloon pumps when indicated as an integer – and risk loving when they cashed in higher than their belief. Based on this variable we could classify 63% as risk averse (n = 99), 5% as risk neutral (n = 8) and 32% as risk loving (n = 51).

4.3. Economic risk and ambiguity

Besides participants' risk behavior in the BERP, we also elicited economic risk and ambiguity based on a multiple-choice list procedure (Sutter et al., 2013). Based on each participant's number of gambles in the risky prospect (noted as *nRG*) and ambiguous prospect (noted as *nAG*), we defined the parameters for each individual's risk and ambiguity preferences. The parameter *r* captures an individual's risk taking and is calculated by:

$$r = 1 - nRG/20$$

A value of *r* larger (smaller) than 0.5 indicates risk averse (risk seeking) behavior. A score of 0.5 indicates risk neutral behavior and corresponds to a participant who decided to gamble 9 or 10 times and opted for the safe option 9 times.[6]

The mean score for the parameter *r* was 0.473. This indicates risk neutral behavior as this score is not significantly different from 0.5 ($p = 0.46$, Wilcoxon signed-ranks test). Women (mean $r = 0.49$) were more risk averse than men (mean $r = 0.45$), but this difference is not statistically significant ($p = 0.418$). Age ($p = 0.799$) and educational attainment ($p = 0.697$) were not correlated with risk aversion in this task.

Individual ambiguity preferences are captured by the parameter *a*, which is calculated by:

$$a = (nRG - nAG)/(nRG + nAG)$$

The difference in the number of chosen gambles between the risky and ambiguous prospects is divided by the total number of chosen gambles in both prospects so that similar differences in chosen gambles between the two prospects will weigh more heavily for a risk averse participant than a risk neutral or risk seeking participant. The parameter *a* ranges from -1 (extreme ambiguity seeking) to 1 (extreme ambiguity aversion). A score of 0 indicates ambiguity neutrality (Wakker, 2010).

On average, participants are ambiguity averse with a mean score of *a* of 0.034 ($p = 0.03$, Wilcoxon signed-ranks test). There is a moderate trend ($p = 0.08$, Wilcoxon signed-ranks test) that women (mean $a = 0.092$) are more ambiguity averse than men (mean $a = -0.008$). Age ($p = 0.870$) and education ($p = 0.153$) have no effect on ambiguity preferences. Risk aversion was negatively correlated with mean cash-in behavior in the BERP and the variable risk-belief. The more risk averse participants were, as elicited from the choice list procedure (indicated by a larger *r*), the less participants pumped the balloon on cashed balloons in the BERP (Pearson's $r = -0.171$, p = 0.033 for mean cash-in behavior and Pearson's $r = -0.146$, p = 0.069 for risk-belief). Ambiguity was not related to these behaviors in the BERP, but did moderately correlate with the variable first cash (Pearson's $r = -0.134$, p = 0.096). This suggests that game play in the BERP could be seen as initially ambiguous (in spite of exposure to the pop-probabilities), but became more aligned with risk as type of uncertainty as the BERP unfolded.

Finally, in this study, 43.6% (46.1%) of participants switched back and forth between the gamble and the sure bet in the risky (ambiguity) context.[7] Although this behavior is not rational, it is not caused by single errors as the mean number of switches is 6.50

[6] Risk-neutral behavior would correspond to 9 gamble choices (r = 0.475) and 1 case of random behavior, which, if we assume would contribute on average ½ to *nRG*, corresponds to a mean *r* of 0.5.

[7] This is a much higher percentage compared to the percentage of switchers in Sutter et al. (2013). This is likely due to the stringent instructions in Sutter et al. (2013), which specifically mentioned a naturally occurring 'switching point' to participants and trained participants that switching

(6.58) in the risky (ambiguity) set of prospects. It is beyond the scope of our study to examine why participants had multiple switches in these economic tasks, but we suspect that they played the task dynamically instead of carefully reviewing each row as part of a monotonically increasing risk task. Participants who switched multiple times in either or both economic contexts also pump the balloon significantly ($p = 0.03$, Wilcoxon signed-ranks test) higher (29.35) than participants who switched once (26.53).[8]

4.4. Risky health behaviors

We first summarize the risky health behaviors of our experimental sample in Table 1. The mean score on the AUDIT was 8.205, implicating hazardous drinking. On average, participants used a variety of 1.75 drugs. The most commonly used drugs are marijuana (71.79% of the sample), followed by hallucinogens (31.41% of the sample) and cocaine (27.56% of the sample). Participants had on average two sexual partners in the previous year. Furthermore, about half of our sample smoked a cigarette last year, and substantially more males than females smoked. Over a quarter of this sample stole an item in the last year. Finally, driving without a seatbelt occurred frequently in the previous year with a percentage of 68.59%.

In accordance with previous research utilizing the BART, our results showed that participants' risk behavior in the BERP (mean cash-in value for cashed balloons) correlated with alcohol use, drug use and smoking behavior.[9] The more participants pumped on cashed-in balloons, the higher they scored on the AUDIT, which aims to characterize problematic drinking, the more drugs they used, and the more likely they were to smoke (Table 2). Risk and ambiguity preferences stemming from the economic tasks did not correlate with any of these health-related risk behaviors. Neither of our tasks correlated with number of sex partners, stealing or driving without a seatbelt.

These results[10] proved robust when controlling for demographic variables (Tables 3 and 4). Gender, which was dummy coded in three categories[11] with females serving as the reference category, and education were also related to alcohol use: males scored higher on the AUDIT than females and the more educated participants scored lower. In the logistical regression on smoking, gender was also a significant variable: males were more likely to smoke than females.

Except for smoking, the variable risk-belief, which compares pump behavior to participants' beliefs in balloon pops, also significantly affected risky health behaviors (Table A1 in Appendix). The variable first cash on the other hand does not relate to risky health behaviors (Table A1 in Appendix), as also indicated by its relationship to ambiguity instead of economic risk, suggests that it represents a different metric.

4.5. The BERP, impulsivity and sensation seeking

Finally, as part of this study, we wanted to shed some light on previous claims regarding the BART, impulsivity, sensation seeking and risky health behaviors. Namely, when Lejuez et al. (2002) introduced the BART, they aimed to show that the BART predicted a variety of healthy risk behaviors over and beyond the effects of impulsivity and sensation seeking. Their main analysis to demonstrate this, was conducted in two steps. They first performed a factor analysis on all self-report questionnaires, including impulsivity and sensation seeking, and all risky behaviors. These results led to one factor describing impulsivity and sensation seeking and two separate factors representing the risky behaviors delinquency, and substance use and sexual behavior. Subsequently they ran two separate hierarchical linear regression models with the factor delinquency as dependent variable in one model, and the factor substance use and sexual behavior as dependent variable in another model. After including demographic variables and the factor describing impulsivity and arousal-seeking as predictors as a first step, adding the variable BART as a second step significantly improved the R-squared in both models. Based on these results they concluded that the BART predicts risk behaviors over and beyond the factor impulsivity and sensation seeking.

Although our goal was to demonstrate the ecological validity of the BERP as a behavioral risk measure, we nevertheless wanted to investigate whether we could replicate the above findings; would the BERP predict health-related risky behaviors over and beyond the effects of impulsivity and sensation seeking?

Like previous work on the BART, we confirmed strong correlations between impulsivity, sensation seeking, and the BERP. As a

(*footnote continued*)
multiple times would not make sense. We purposely did not provide these clues, and made no mention of a switching point. The fact that we have a higher number of repeated switchers is worrisome in light of this method, but it pinpoints the exact problem of this method: if not instructed, participants may not have a naturally occurring switching point, or they do not understand the task well even though clear instructions (that did not include a single obligatory switch) were provided.

[8] As this categorization of multiple switchers versus those who only switch once lines up with their game play in the BERP, it is tricky to add this variable to our regression models. For completeness, we nevertheless verified all our main results by controlling for whether participants switched repeatedly versus only once by adding a variable (dummy coded in two categories where participants who switch more than once are the reference category) to our models (Tables 3 and 4). All results remained qualitatively valid.

[9] Using Holm's adjustment for multiple hypothesis testing (6 hypotheses comparing the BERP with 6 risky behaviors), the p value for drugs use becomes $p = 0.078$, alcohol use $p = 0.24$, and smoking $p = 0.33$.

[10] One can argue whether one more category of drug use bears the same intensity of unit increase along the scale, e.g. how to compare participants who both use two drug categories, but one uses cocaine and heroin, whereas the other uses marijuana and hallucinogens. Therefore, we also conducted an ordered ordinal regression on the same model as in Table 3. Qualitatively, the results remained the same.

[11] One participant preferred to not report a binary gender.

Table 1
Summary statistics for risky health behaviors.

Variable	Women (n = 69)	Men (n = 86)	Overall (n = 156)
Alcohol	7.159	9.140	8.205
Drugs	1.565	1.884	1.750
Sex partners	2.072	2.046	2.071
Smoking	39.13%	55.81%	48.72%
Stealing	31.88%	24.42%	28.21%
Driving without seatbelt	75.36%	62.79%	68.59%

Table 2
Correlation coefficients between risky healthy behaviors and risk parameters.

	Alcohol	Drugs	Sex partners	Smoking	Stealing	Driving without seatbelt
BERP	0.16**	0.20**	0.04	0.15*	0.05	0.04
Risk	0.02	−0.09	−0.01	0.09	−0.12	−0.06
Ambiguity	−0.05	0.06	0.09	−0.04	0.01	0.03

Note: the table shows Spearman correlations for drug use, smoking, stealing and driving without seatbelt; otherwise Pearson correlations are shown
*** Significant at the 1 percent level.
** Significant at the 5 percent level.
* Marginal at the 10 percent level.

Table 3
Least-square models of alcohol and drug use.

Independent variables	Dependent variables			
	Alcohol use		Drug use	
BERP		0.120*		0.037**
		(0.064)		(0.017)
Risk	1.541	2.161	0.220	0.414
	(2.198)	(2.205)	(0.592)	(0.591)
Ambiguity	−0.151	−0.116	−0.194	−0.183
	(1.427)	(1.415)	(0.384)	(0.379)
Male	1.821*	1.777*	0.238	0.224
	(0.957)	(0.950)	(0.258)	(0.255)
Unidentified gender	−7.868	−6.290	1.233	1.727
	(5.900)	(5.913)	(1.588)	(1.585)
Age	0.007	0.032	−0.065	−0.057
	(0.232)	(0.231)	0.063	0.062
Education	−0.682*	−0.700*	−0.155	−0.161
	(0.411)	(0.407)	0.111	0.109
Observations	156	156	156	156
Adjusted R-squared	0.03	0.05	0.03	0.06
F-statistic	$F(6,149) = 1.947$*	$F(7,148) = 2.191$**	$F(6,149) = 1.917$*	$F(7,148) = 2.358$**

Regression coefficients reported followed by standard errors in parentheses.
*** Significant at the 1 percent level.
** Significant at the 5 percent level.
* Marginal at the 10 percent level.

more straightforward analysis, we added impulsivity and sensation seeking as individual variables to the models described in Tables 3 and 4. We also ran a factor analysis and hierarchical linear regression as in Lejuez et al. (2002). Neither of our analyses support the idea that all three concepts – risk behavior in the BERP, impulsivity and sensation seeking – described as individual variables or as composed factors, remain significant predictors of risky health behaviors. Behavior in the BERP and both impulsivity and sensation seeking are non-orthogonal concepts and they clearly compete with one another when added to the same model.

5. Discussion

This study introduced the BERP, a modified protocol for the BART (Lejuez et al., 2002). The BART consistently relates to real-world risk taking (Hopko et al., 2006; Hunt et al., 2005; Lejuez et al., 2003b, 2003a, 2002), but suffers from methodological flaws and elicits extreme risk averse behavior. Unlike the BART, the BERP adheres to best practices in experimental economics and adds the capability to compare participant choice to optimal behavior and individual beliefs.

Table 4
Logistic regression model of smoking.

Independent variables	Dependent variable	
	Smoking	Smoking
BERP		0.047**
		(0.024)
Risk	0.890	1.143
	(0.805)	(0.825)
Ambiguity	−0.420	−0.445
	(0.537)	(0.552)
Male	0.624*	0.633*
	(0.344)	(0.349)
Unidentified gender	14.559	15.189
	(882.74)	(882.74)
Age	−0.130	−0.123
	(0.089)	(0.089)
Education	−0.151	−0.163
	(0.150)	(0.407)
Observations	156	156
Pseudo R-squared	0.14	0.17
LR χ^2	χ^2 (6) = 16.85***	χ^2 (7) = 20.75***

Regression coefficients reported followed by standard errors in parentheses.
*** Significant at the 1 percent level.
** Significant at the 5 percent level.
* Marginal at the 10 percent level.

We found that most (~69%) of BERP participants can be defined as risk averse, which is comparable to other behavioral measures of risk (e.g. Crosetto & Filippin, 2016). Behavioral risk measures and belief elicitations utilized as part of the BERP were also correlated, supporting the validity of the protocol. In addition to consistency with other behavioral measures of risk from economics, risk behavior in the BERP carries external validity, as it correlates with alcohol use, drug use, and smoking.

Our work extends previous efforts to improve the ecological validity of economic risk tasks. Although some researchers doubt the efficacy of economic risk tasks to predict risky behaviors (Eckel, 2016; Friedman et al., 2014; Trautmann, 2016), there have been numerous attempts to improve the poor ecological validity of current risk measures (Dohmen et al., 2011; Menkhoff and Sakha, 2017; Schildberg-Hörisch, 2018). For instance, utilizing a risk elicitation task in the same decision domain (Dohmen et al., 2011) and context (Zhou and Hey, 2018) as the risky behavior of interest. Although, economics has focused on revealed preferences that are directly incentivized (Charness et al., 2013), a general question about risky behavior performs surprisingly well (Dohmen et al., 2011; Menkhoff and Sakha, 2017). Another strategy that was proposed recently, suggests combining multiple risk elicitation tasks to calculate a normalized composite risk score. This method captures a common component of risk across multiple tasks in order to measure a generalizable aspect of risk that increases predictive performance and has the benefit of reducing measurement error (Menkhoff and Sakha, 2017; Pedroni et al., 2017; Schildberg-Hörisch, 2018).

In accordance with the strategies proposed above, the BERP (1) aligns well with many risky health behaviors and (2) captures different cognitive aspects of risk because of its composite nature. First, unlike many economic risk tasks, the BART is an intuitive, continuous and dynamic decision-making task with immediate outcome feedback. The risk dynamics in this game overlap with the risky health contexts under investigation. Just like substance abuse, the risk (of intoxication) is low for the first balloon pumps in the BERP, but increases over time. At the same time participants experience positive outcomes by accumulating tokens with every balloon pump. This adds to the difficulty of choosing to stop pumping the balloon, just like many indicate to find it difficult to stop alcohol intake beyond their self-imposed limit (Muraven et al., 2005a, 2005b). Second, the BART is an engaging task, which simultaneously captures two cognitive aspects of risk taking - diminishing returns and increasing hazard – and affective components such as hope, exhilaration and tension (Schonberg et al., 2011). Moreover, participants' risk behavior is obtained from multiple trials and is therefore not constructed on the basis of a single risk item (e.g. a switch point or an answer to a specific gamble). Altogether, this likely reduces measurement error and early support for this view is the fact that the first trial on which participants cashed their balloon did not correlate with risky health behaviors in this study, whereas the mean score of cashed balloons did predict substance use and smoking.

Researchers have also expressed concerns about the stability of measured preferences across economic risk tasks (Menkhoff and Sakha, 2017; Pedroni et al., 2017; Zhou and Hey, 2018). Our study allowed us to investigate risk preferences from the BERP in relation to other economic risk tasks often used in experimental economics. In accordance with previous findings (Deck et al., 2013; Crosetto & Filippin, 2016; Menkhoff and Sakha, 2017; Pedroni et al., 2017), we find a significant, but small correlation between risk elicited by the BERP and a multiple price list. This is not surprising as the risk tasks differ quite considerably. Like Pedroni et al. (2017) argued, we also believe participants adopt a variety of decision strategies in response to different choice architectures introduced across the available risk elicitation tasks.

Lastly, this study also sought to investigate whether we could replicate the following findings as laid out by Lejuez et al. (2002); does balloon pumping behavior correlate with health-related risky behaviors over and beyond the effects of impulsivity and sensation seeking? Neither of our analyses support this claim. Like other studies (Lejuez et al., 2003b, 2003a, 2002), we also show that behavior in the BERP correlates with participants' self-reported impulsivity and sensation seeking. This is not surprising as these questionnaires contain elements that are very similar to the tools used to assess participants' real-world risk behavior (Eckel & Wilson, 2004; Lejuez et al., 2002) and may partially explain the large number of factors identified in risk questionnaires like the DOSPERT (Weber et al., 2002). Moreover, the BART was purposefully developed to clinically assess risk-taking in people (Lejuez et al., 2002). Its multifaceted nature means that aspects of both impulsivity and sensation seeking may be captured in the BART. This is supported by the fact that they clearly compete with one another when added to the same model. In accordance with the principle of revealed preferences, we believe the BERP captures choice preferences that may not be directly accessible to participants. A testable hypothesis extending from this supposition is that risk preferences expressed in the BERP would be more likely to correlate with real-world risk behaviors that were less similar to questions asked in standard sensation seeking surveys.

Like any other study, ours also comes with limitations. First, we use participants' self-reported answers regarding risky health behaviors, and although we guaranteed full anonymity, gave full disclosure regarding our data management plan, and provided credibility via a letter from the Institutional Review Board of the University where this study took place, we cannot know for sure if participants' descriptions of their risky activities actually occurred. Second, the diverse nature of the BERP is an asset, but has downsides as well. Namely, the BERP remains a mixture of both risk and ambiguity as types of uncertainty and learning is likely to contribute additional individual variance. In addition, affective responses towards the experience of pops and successful cash-ins are inevitable and could promote sensation seeking. Through additional modifications of the task it should be possible to isolate these components in future studies. Lastly, participants' belief distributions are non-uniform, although we made every effort to allow participants to form a true prior based on the balloon pops observed before the experiment started. In future iterations of the BERP we plan to elicit beliefs throughout the experiment to test whether participants recognize the uniform distribution as they progress in the game.

Our study highlights the value of using insights from economics and psychology to design a risk task that can support better mathematical models without sacrificing ecological validity. We look forward to isolating the cognitive processes that contribute to risk preferences in future iterations of the BERP and to more directly test additional domains of real-world risk.

Acknowledgements

Declarations of interest: None.

Appendix A

See Fig. A1 and Table A1.

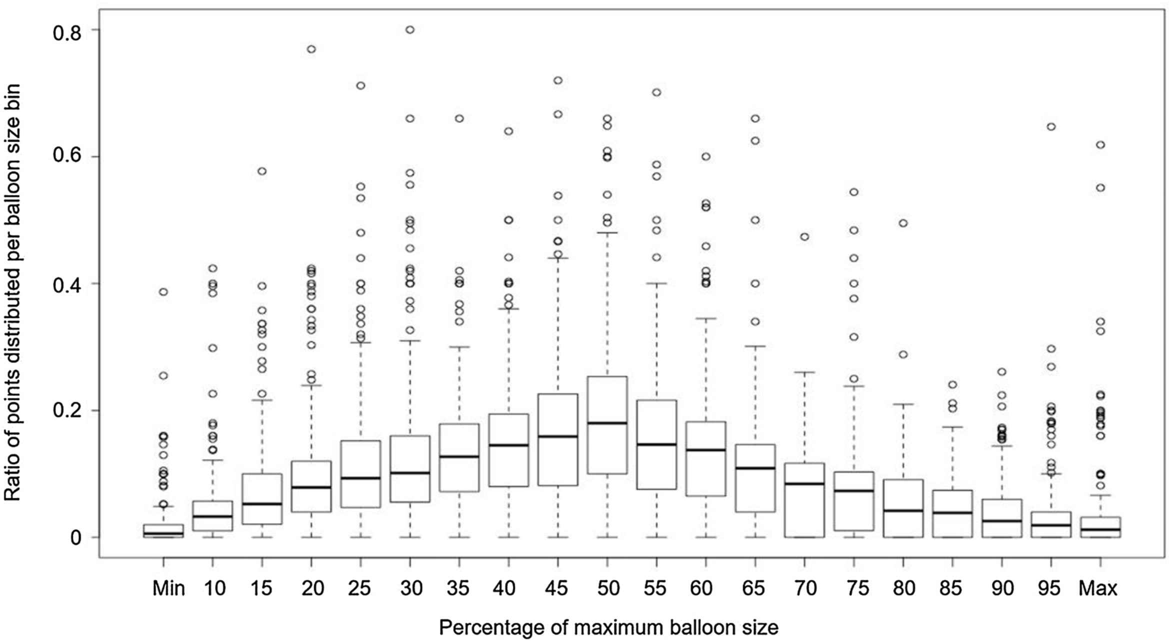

Fig. A1. Boxplots of participants' distributed points per balloon size bin. Participants with a ratio higher than 0.5 of points distributed to one balloon size bin are not depicted.

Table A1
Correlation coefficients between risky healthy behaviors and alternative risk metrics stemming from the BERP.

	Alcohol	Drugs	Sex partners	Smoking	Stealing	Driving without seatbelt
First cash	0.09	0.08	−0.02	0.11	0.00	−0.03
Risk-belief	0.16**	0.16**	0.15*	0.11	0.02	0.07

Note: the table shows Spearman correlations for smoking, stealing and driving without seatbelt; otherwise Pearson correlations are shown.
*** Significant at the 1 percent level.
** Significant at the 5 percent level.
* Marginal at the 10 percent level.

Experimental instructions

NOTE: these instructions were visibly shown on each participant's computer screen as participants could simultaneously learn about the game details and play some practice balloons. We will categorize the experimental instructions by numbers referring to the individual screen slides.

Please email the corresponding author for a copy of the script if you wish to go through the instructions in an animated way.

1
Instructions Balloon pump game
To continue through these instructions, press the space bar
2
Your payment in this experiment
- Participation: $10/hour
- Bonus: throughout this experiment, you will use tokens to play the balloon pump game.
- Conversion rate: 1 token = $0.20
- After you have learned about today's experiment, we will explain in greater detail how your choices may earn you extra money.

3
Balloon pump game
- First, you will see a small balloon which is worth 0 tokens.
- At any time you may pump the balloon, which will cause the balloon to grow larger.
- With each pump, the balloon's value will also increase by one token.
- Holding the space bar will pump the balloon rapidly.
- You may press or hold the space bar as often as you'd like.

4
Pop point
- Each balloon has the same maximum possible balloon size.
- Each balloon can have a different 'pop point' though.
- This pop point may lie anywhere between the first pump and the maximum possible balloon size.
- If your balloon pops, your tokens will be set to zero tokens.

5
Cashing in
- At any time you can receive your accumulated tokens by cashing in.
- You will receive the number of tokens corresponding to the balloon's size when you cash in.
- If the balloon pops before you cash in, you will receive 0 tokens.

6
Balloon pump game: Alone
- You will play the balloon pump game by yourself.
- Each time before you see a new balloon, you will first see a picture of yourself.
- After a fixation cross a balloon will appear, which indicates that you may start pumping the balloon.

7
Do you have any questions?
Please raise your hand if something is not clear and we will come to you.
8
Observation
- Before you play the balloon pump game, you will simply observe several balloons pumping. More specifically, you will see 16 examples in which a balloon is pumped to its pop point.
- As explained before, the balloon can pop anywhere between the first pump and the maximum possible balloon size.
- From these balloons you will get a sense of how a balloon looks when pumped, and an idea of where any balloon might pop.
- You will only observe these balloons. You will not be able to pump them yourself.

9
Observation
If you are ready to observe these balloons, press the space bar.
10
- To familiarize you with the game and how to play it, you will now play a few balloons. The balloons you play with during the practice are strictly for your familiarization with the task, and will not contribute to your possible earnings.
- Press or hold the SPACE BAR to pump, and ENTER to cash in.

11
Payment in balloon pump game

- In addition to your compensation for participating today, you may also receive a bonus.
- We will randomly draw one balloon from one game and its corresponding tokens as a bonus for each person.
- You will receive 0 tokens if we draw a balloon that you popped.
- You will privately receive your payment in cash at the end of today's experiment.

12
Beliefs
Before you begin playing, we would like you to tell us a few of your expectations regarding the game.
13
Beliefs: where do you expect pop points to be?

- We would now like to know where you expect pop points to occur, relative to the maximum possible balloon's size.
- The scale (image below) is evenly divided into 20 units, each unit representing a balloon size stated as a certain percentage of the maximum possible balloon size.
- Please indicate your expectations by distributing points (at least 100) to any of the units on a scale like the one shown below.
- By distributing more points to a specific unit, you will be indicating that you expect the balloon's pop point to occur there more often than at other balloon's sizes.
- Example: you distribute 300 points, and 60 have been allocated to one specific unit. This would indicate that out of 300 balloons, you believe 60 would pop at that size.

14
Balloon pop point belief rating payment

- You will be paid for the accuracy of your ratings.
- At the end of the experiment, the computer will randomly pick one balloon's pop point and will compare your expectations to this value.
- The more points (relative to all points you used) that you distribute to the correct unit, the more accurate your expectation will be, and the more tokens you will receive.
- At most you can earn 20 tokens, for allocating all your points to the correct unit.
- You will receive 0 tokens if you allocate none of your points to the correct unit.

15
Beliefs: where do you expect pop points to lie?

- We have made it very easy for you to draw your expected distribution of pop points! Just click on any point above a unit and it will drag the bar to that point. When holding the cursor and moving it over all units, a distribution is created. Adjust as per your preferences.

16
A few more things before we begin

- One of the participants in today's experiment will play these games in the MRI scanner. We will now take that participant to the MRI scanner room.
- (S)he will receive some extra training in the scanner on which buttons to push.
- After the experiment we will ask you to fill in some remaining questionnaires.

17
Overview

- Do you have any questions?
- Please raise your hand now.
- If not, the balloon pump game will begin!

Risk/ambiguity task after the game play:
1

- The first part in this questionnaire consists of two choice tasks.
- In both tasks, you will see a table with 20 rows. In each row, you are asked to choose between Option A and B.
- Option A is a lottery and Option B is a sure amount of tokens.
- After the experiment the computer will randomly draw one row from both tasks.
- If you chose Opion A you will play the lottery. If you chose Option B you will receive the sure amount of tokens.

2
Choice task 1

- In this choice task the lottery consists of a bowl filled with 10 ping-pong balls. 5 ping-pong balls are orange and 5 ping-pong balls are white.
- Please indicate in each row if you want to play this lottery or receive the sure amount of tokens stated in that row.
- If the computer randomly selects a row in which you selected to play the lottery, you may draw a ping-pong ball from the bowl. If this drawn ping-pong ball matches the color you selected at the beginning of the experiment, you win 5 tokens. Else you win 0 tokens.
- If you select Option B, you will receive the sure amount of tokens.

3
In each row choose between Option A and B

Option A		Option B
1	O I choose the certain amount of 0.25 tokens	O I choose to draw a ball
2	O I choose the certain amount of 0.50 tokens	O I choose to draw a ball
3	O I choose the certain amount of 0.75 tokens	O I choose to draw a ball
4	O I choose the certain amount of 1 tokens	O I choose to draw a ball
5	O I choose the certain amount of 1.25 tokens	O I choose to draw a ball
6	O I choose the certain amount of 1.50 tokens	O I choose to draw a ball
7	O I choose the certain amount of 1.75 tokens	O I choose to draw a ball
8	O I choose the certain amount of 2 tokens	O I choose to draw a ball
9	O I choose the certain amount of 2.25 tokens	O I choose to draw a ball
10	O I choose the certain amount of 2.50 tokens	O I choose to draw a ball
11	O I choose the certain amount of 2.75 tokens	O I choose to draw a ball
12	O I choose the certain amount of 3 tokens	O I choose to draw a ball
13	O I choose the certain amount of 3.25 tokens	O I choose to draw a ball
14	O I choose the certain amount of 3.50 tokens	O I choose to draw a ball
15	O I choose the certain amount of 3.75 tokens	O I choose to draw a ball

16	O I choose the certain amount of 4 tokens	O I choose to draw a ball
17	O I choose the certain amount of 4.25 tokens	O I choose to draw a ball
18	O I choose the certain amount of 4.50 tokens	O I choose to draw a ball
19	O I choose the certain amount of 4.75 tokens	O I choose to draw a ball
20	O I choose the certain amount of 5 tokens	O I choose to draw a ball

4

Choice task 2

- In this choice task the lottery consists of a bowl filled with 10 ping-pong balls in an unknown composition of orange and white ping-pong balls.
- Please indicate in each row if you want to play this lottery or receive the sure amount of tokens stated in that row.
- If the computer randomly selects a row in which you selected to play the lottery, you may draw a ping-pong ball from the bowl. If this drawn ping-pong ball matches the color you selected at the beginning of the experiment, you win 5 tokens. Else you win 0 tokens.
- If you select Option B, you will receive the sure amount of tokens.

5

In each row choose between Option A and B

Option A		Option B
1	O I choose the certain amount of 0.25 tokens	O I choose to draw a ball
2	O I choose the certain amount of 0.50 tokens	O I choose to draw a ball
3	O I choose the certain amount of 0.75 tokens	O I choose to draw a ball
4	O I choose the certain amount of 1 tokens	O I choose to draw a ball
5	O I choose the certain amount of 1.25 tokens	O I choose to draw a ball
6	O I choose the certain amount of 1.50 tokens	O I choose to draw a ball
7	O I choose the certain amount of 1.75 tokens	O I choose to draw a ball
8	O I choose the certain amount of 2 tokens	O I choose to draw a ball
9	O I choose the certain amount of 2.25 tokens	O I choose to draw a ball
10	O I choose the certain amount of 2.50 tokens	O I choose to draw a ball
11	O I choose the certain amount of 2.75 tokens	O I choose to draw a ball
12	O I choose the certain amount of 3 tokens	O I choose to draw a ball
13	O I choose the certain amount of 3.25 tokens	O I choose to draw a ball
14	O I choose the certain amount of 3.50 tokens	O I choose to draw a ball
15	O I choose the certain amount of 3.75 tokens	O I choose to draw a ball
16	O I choose the certain amount of 4 tokens	O I choose to draw a ball
17	O I choose the certain amount of 4.25 tokens	O I choose to draw a ball
18	O I choose the certain amount of 4.50 tokens	O I choose to draw a ball
19	O I choose the certain amount of 4.75 tokens	O I choose to draw a ball
20	O I choose the certain amount of 5 tokens	O I choose to draw a ball

Appendix B. Supplementary material

Supplementary data to this article can be found online at https://doi.org/10.1016/j.joep.2019.04.005.

References

Aklin, W. M., Lejuez, C. W., Zvolensky, M. J., Kahler, C. W., & Gwadz, M. (2005). Evaluation of behavioral measures of risk taking propensity with inner city adolescents. *Behaviour Research and Therapy, 43*(2), 215–228.

Anderson, L. R., & Mellor, J. M. (2008). Predicting health behaviors with an experimental measure of risk preference. *Journal of Health Economics, 27*(5), 1260–1274.

Arrow, K. J. (1951). Alternative approaches to the theory of choice in risk-taking situations. *Econometrica: Journal of the Econometric Society, 19*(4), 404–437.

Artinger, F., Exadaktylos, F., Koppel, H., & Sääksvuori, L. (2010). Applying quadratic scoring rule transparently in multiple choice settings: A note. Jena economic research papers. https://www.econstor.eu/handle/10419/32581.

Barratt, E. S. (1985). Impulsiveness subtraits: Arousal and information processing. *Motivation, Emotion, and Personality, 5*, 137–146.

Bornovalova, M. A., Daughters, S. B., Hernandez, G. D., Richards, J. B., & Lejuez, C. W. (2005). Differences in impulsivity and risk-taking propensity between primary users of crack cocaine and primary users of heroin in a residential substance-use program. *Experimental and Clinical Psychopharmacology, 13*(4), 311–318.

Budner, S. (1962). Intolerance of ambiguity as a personality variable. *Journal of Personality, 30*, 29–50.

Charness, G., & Gneezy, U. (2012). Strong evidence for gender differences in risk taking. *Journal of Economic Behavior & Organization, 83*(1), 50–58.

Charness, G., Gneezy, U., & Imas, A. (2013). Experimental methods: Eliciting risk preferences. *Journal of Economic Behavior & Organization, 87*, 43–51.

Clasen, D. R., & Brown, B. B. (1985). The multidimensionality of peer pressure in adolescence. *Journal of Youth and Adolescence, 14*(6), 451–468.

Crosetto, P., & Filippin, A. (2013). The “bomb” risk elicitation task. *Journal of Risk and Uncertainty, 47*(1), 31–65.

Crosetto, P., & Filippin, A. (2016). A theoretical and experimental appraisal of four risk elicitation methods. *Experimental Economics, 19*(3), 613–641.

Crowley, T. J., Raymond, K. M., Mikulich-Gilbertson, S. K., Thompson, L. L., & Lejuez, C. W. (2006). A risk-taking “set” in a novel task among adolescents with serious conduct and substance problems. *Journal of the American Academy of Child and Adolescent Psychiatry, 45*(2), 175–183.

Cubitt, R. P., Starmer, C., & Sugden, R. (1998). On the validity of the random lottery incentive system. *Experimental Economics, 1*(2), 115–131.

Deck, C., Lee, J., Reyes, J. A., & Rosen, C. C. (2013). A failed attempt to explain within subject variation in risk taking behavior using domain specific risk attitudes. *Journal of Economic Behavior & Organization, 87*, 1–24.

Dohmen, T., Falk, A., Huffman, D., Sunde, U., Schupp, J., & Wagner, G. G. (2011). Individual risk attitudes: Measurement, determinants and behavioral consequences.

Journal of the European Economic Association, 9(3), 522–550.
Eckel, C. C., & Grossman, P. J. (2002). Sex differences and statistical stereotyping in attitudes toward financial risk. *Evolution and Human Behavior: Official Journal of the Human Behavior and Evolution Society, 23*(4), 281–295.
Eckel, C. C., & Wilson, R. K. (2004). Is trust a risky decision? *Journal of Economic Behavior & Organization, 55*(4), 447–465.
Eckel, C. C. (2016). Risky Curves: On the empirical failure of expected utility theory, Daniel Friedman, R. Mark Isaac, Duncan James and Shyam Sunder. Routledge, 2014, xiii + 137 pages. Economics and Philosophy, 32(03), 540–548.
Ellsberg, D. (1961). Risk, ambiguity, and the savage axioms. *The Quarterly Journal of Economics, 75*(4), 643–669.
Figner, B., Mackinlay, R. J., Wilkening, F., & Weber, E. U. (2009). Affective and deliberative processes in risky choice: Age differences in risk taking in the Columbia Card Task. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 35*(3), 709–730.
Figner, B., & Murphy, R. O. (2011). Using skin conductance in judgment and decision making research. *A Handbook of Process Tracing Methods for Decision Research,* 163–184.
Figner, B., & Weber, E. U. (2011). Who takes risks when and why? Determinants of risk taking. *Current Directions in Psychological Science, 20*(4), 211–216.
Friedman, D., Mark Isaac, R., James, D., & Sunder, S. (2014). Risky Curves: On the Empirical Failure of Expected Utility. Routledge.
Fukunaga, R., Brown, J. W., & Bogg, T. (2012). Decision making in the Balloon Analogue Risk Task (BART): Anterior cingulate cortex signals loss aversion but not the infrequency of risky choices. *Cognitive, Affective & Behavioral Neuroscience, 12*(3), 479–490.
Gneezy, U., & Potters, J. J. M. (1997). An experiment on risk taking and evaluation periods. *The Quarterly Journal of Economics, 112*(2), 631–645.
Grant, S., Contoreggi, C., & London, E. D. (2000). Drug abusers show impaired performance in a laboratory test of decision making. *Neuropsychologia, 38*(8), 1180–1187.
Harrison, G. W., Martínez-Correa, J., Todd Swarthout, J., & Ulm, E. R. (2017). Scoring rules for subjective probability distributions. *Journal of Economic Behavior & Organization, 134*, 430–448.
Harrison, G. W. (2015). Risky Curves: On the empirical failure of expected utility, Daniel Friedman, R. Mark Isaac, Duncan James, Shyam Sunder (Eds.). Routledge, New York (2014). xiii+ 137 pp., $145.00 (hc), ISBN: 978-0415-636100. Journal of Economic Psychology, 48, 121–125.
Hertwig, R., Hogarth, R. M., & Lejarraga, T. (2018). Experience and description: Exploring two paths to knowledge. *Current Directions in Psychological Science, 27*(2), 123–128.
Holt, C. A., & Laury, S. K. (2002). Risk aversion and incentive effects. *The American Economic Review, 92*(5), 1644–1655.
Hopko, D. R., Lejuez, C. W., Daughters, S. B., Aklin, W. M., Osborne, A., Simmons, B. L., et al. (2006). Construct Validity of the Balloon Analogue Risk Task (BART): Relationship with MDMA Use by Inner-City Drug Users in Residential Treatment. *Journal of Psychopathology and Behavioral Assessment, 28*(2), 95–101.
Hoyle, R. H., Stephenson, M. T., & Palmgreen, P. (2002). Reliability and validity of a brief measure of sensation seeking. *Personality and Individual Differences, 32*(3), 401–414.
Hunt, M. K., Hopko, D. R., Bare, R., Lejuez, C. W., & Robinson, E. V. (2005). Construct validity of the Balloon Analog Risk Task (BART): Associations with psychopathy and impulsivity. *Assessment, 12*(4), 416–428.
Kagel, J. H., & Roth, A. E. (2016). The Handbook of Experimental Economics, Volume 2: The Handbook of Experimental Economics. Princeton University Press.
Kaivanto, K., & Kroll, E. B. (2012). Negative recency, randomization device choice, and reduction of compound lotteries. *Economics Letters, 115*(2), 263–267.
Lejuez, C. W., Aklin, W. M., Jones, H. A., Richards, J. B., Strong, D. R., Kahler, C. W., et al. (2003a). The Balloon Analogue Risk Task (BART) differentiates smokers and nonsmokers. *Experimental and Clinical Psychopharmacology, 11*(1), 26–33.
Lejuez, C. W., Aklin, W. M., Zvolensky, M. J., & Pedulla, C. M. (2003b). Evaluation of the Balloon Analogue Risk Task (BART) as a predictor of adolescent real-world risk-taking behaviours. *Journal of Adolescence, 26*(4), 475–479.
Lejuez, C. W., Read, J. P., Kahler, C. W., Richards, J. B., Ramsey, S. E., Stuart, G. L., et al. (2002). Evaluation of a behavioral measure of risk taking: The Balloon Analogue Risk Task (BART). *Journal of Experimental Psychology. Applied, 8*(2), 75–84.
List, J. A., Sadoff, S., & Wagner, M. (2011). So you want to run an experiment, now what? Some simple rules of thumb for optimal experimental design. *Experimental Economics, 14*(4), 439 Accessed 23 August 2017.
Menkhoff, L., & Sakha, S. (2017). Estimating risky behavior with multiple-item risk measures. *Journal of Economic Psychology, 59*, 59–86.
Muraven, M., Collins, R. L., Morsheimer, E. T., Shiffman, S., & Paty, J. A. (2005a). The morning after: Limit violations and the self-regulation of alcohol consumption. *Psychology of Addictive Behaviors: Journal of the Society of Psychologists in Addictive Behaviors, 19*(3), 253–262.
Muraven, M., Collins, R. L., Shiffman, S., & Paty, J. A. (2005b). Daily fluctuations in self-control demands and alcohol intake. *Psychology of Addictive Behaviors: Journal of the Society of Psychologists in Addictive Behaviors, 19*(2), 140–147.
Pedroni, A., Frey, R., Bruhin, A., Dutilh, G., Hertwig, R., & Rieskamp, J. (2017). The risk elicitation puzzle. *Nature Human Behaviour, 1*(11), 803.
Peirce, J. W. (2007). PsychoPy—Psychophysics software in Python. *Journal of Neuroscience Methods, 162*(1), 8–13.
Prinstein, M. J., Brechwald, W. A., & Cohen, G. L. (2011). Susceptibility to peer influence: Using a performance-based measure to identify adolescent males at heightened risk for deviant peer socialization. *Developmental Psychology, 47*(4), 1167–1172.
Qiu, J., & Weitzel, U. (2016). Experimental evidence on valuation with multiple priors. *Journal of Risk and Uncertainty, 53*(1), 55–74.
Rao, H., Korczykowski, M., Pluta, J., Hoang, A., & Detre, J. A. (2008). Neural correlates of voluntary and involuntary risk taking in the human brain: An fMRI Study of the Balloon Analog Risk Task (BART). *NeuroImage, 42*(2), 902–910.
Saunders, J. B., Aasland, O. G., Babor, T. F., de La Fuente, J. R., & Grant, M. (1993). Development of the Alcohol Use Disorders Identification Test (AUDIT): WHO Collaborative Project on Early Detection of Persons with Harmful Alcohol Consumption-II. *Addiction, 88*(6), 791–804.
Schildberg-Hörisch, H. (2018). Are risk preferences stable? *Journal of Economic Perspectives, 32*(2), 135–154.
Schonberg, T., Fox, C. R., Mumford, J. A., Congdon, E., Trepel, C., & Poldrack, R. A. (2012). Decreasing ventromedial prefrontal cortex activity during sequential risk-taking: an FMRI investigation of the balloon analog risk task. *Frontiers in Neuroscience, 6*, 80.
Schonberg, T., Fox, C. R., & Poldrack, R. A. (2011). Mind the gap: Bridging economic and naturalistic risk-taking with cognitive neuroscience. *Trends in Cognitive Sciences, 15*(1), 11–19.
Starmer, C., & Sugden, R. (1991). Does the random-lottery incentive system elicit true preferences? An experimental investigation. *The American Economic Review, 81*(4), 971–978.
Sutter, M., Kocher, M. G., Glätzle-Rützler, D., & Trautmann, S. T. (2013). Impatience and uncertainty: Experimental decisions predict adolescents' field behavior. *The American Economic Review, 103*(1), 510–531.
Trautmann, S. (2016). Risky Curves: On the empirical failure of expected utility, Daniel Friedman, R. Mark Isaac, Duncan James, Shyam Sunder (Eds.). Routledge, New York (2014). xiii+ 137 pp., $145.00 (hc), ISBN: 978-0415-636100. Journal Of Economic Psychology, 53, 178–179.
Tversky, A., & Kahneman, D. (1992). Advances in prospect theory: Cumulative representation of uncertainty. *Journal of Risk and Uncertainty, 5*(4), 297–323.
US Department of Health and Human Services; Centers for Disease Control (CDC). (n.d.). Assessing health risk behaviors among young people: Youth risk behavior surveillance system 2004. PsycEXTRA Dataset. doi: 10.1037/e421292005-001.
von Neumann, J., & Morgenstern, O. (1944). *Theory of games and economic behavior.* Princeton University Press Princeton.
Wakker, P. P. (2010). *Prospect Theory: For Risk and Ambiguity.* Cambridge University Press.
Ward, J. (1990). The HIV Risk-taking Behaviour Scale (HRBS) Manual. National Drug and Alcohol Research Centre, University of New South Wales.
Weber, E. U., Blais, A.-R., & Betz, N. E. (2002). A domain-specific risk-attitude scale: Measuring risk perceptions and risk behaviors. *Journal of Behavioral Decision Making, 15*(4), 263–290.
Zhou, W., & Hey, J. (2018). Context matters. *Experimental Economics, 21*(4), 723–756.
Zuckerman, M., Kolin, E. A., Price, L., & Zoob, I. (1964). Development of a sensation-seeking scale. *Journal of Consulting Psychology, 28*(6), 477.