

# Supplemental Material: Representative Design in Psychological Assessment: A Case Study Using the Balloon Analogue Risk Task (BART)

Markus D. Steiner  
University of Basel

Renato Frey  
University of Basel

## Contents

<b>1</b>	<b>Details on the Results of the Reanalyses</b>	<b>2</b>
<b>2</b>	<b>Sequences of Explosion Points Used in the Experiment</b>	<b>2</b>
<b>3</b>	<b>Participants' General Task Representation in Study 1</b>	<b>3</b>
<b>4</b>	<b>Learning in the BART</b>	<b>3</b>
<b>5</b>	<b>Correlations of the other Risk-Taking Measures</b>	<b>4</b>
<b>6</b>	<b>Posterior Parameter Estimates of H2 and H3</b>	<b>5</b>
<b>7</b>	<b>Convergent Validity Between the BART and Other Risk-Taking Measures</b>	<b>5</b>
7.1	Study 1 . . . . .	5
7.2	Study 2 . . . . .	10
<b>8</b>	<b>Test–Retest Reliability of the Four BART Implementations</b>	<b>14</b>
<b>9</b>	<b>Summary of Indicators From Studies 1 and 2</b>	<b>14</b>
<b>10</b>	<b>Test–Retest Reliability of All Risk-Taking Measures</b>	<b>16</b>
<b>11</b>	<b>Simulating the Effect of the Distributional Form</b>	<b>16</b>
<b>12</b>	<b>References</b>	<b>17</b>

## 1 Details on the Results of the Reanalyses

In our reanalyses of the five datasets from Frey, Pedroni, Mata, Rieskamp, and Herwig (2017); Schürmann, Frey, and Pleskac (2018); Steiner and Frey (2020), we found that most participants judgments were best described by CDFs of normal distributions in the datasets from Schürmann et al. (2018); Steiner and Frey (2020). However, in the dataset from Frey et al. (2017), this proportion was somewhat lower, with large proportions of participants also being best described by a CPF of a normal distribution or a CDF of a uniform distribution.

A potential reason leading to this slightly less clear pattern may consist of the specific implementation of the probability rating task. That is, the spacing between the different stages at which participants provided their probability ratings, varied slightly between studies: In Frey et al. (2017), the first nine stages were equally spaced up to a participant’s adjusted BART score, and then went in steps of 10 from the adjusted BART score to the maximum number of inflations of the respective participant. Thus, the spacing could differ within and between participants and not necessarily the whole range of the uniform distribution was covered. As this procedure complicates the comparison of different participants, the two subsequent implementations used ten and eight equally-spaced intervals of the balloons, respectively, ranging from 1 to  $C - 1$  and then one more balloon with  $C$  inflations. Due to this difference in the implementation the CPF may have been favored in the dataset from Frey et al. (2017) due to less leveling off at the upper end of the probability ratings.

## 2 Sequences of Explosion Points Used in the Experiment

In the experiment, we used four different sequences of explosion points, one for each implementation of the BART. Within each implementation, we used the same order of the sequence, to minimize noise produced by random variation in the explosion points. The distribution and actual order of the explosion points are shown in Figure S1.

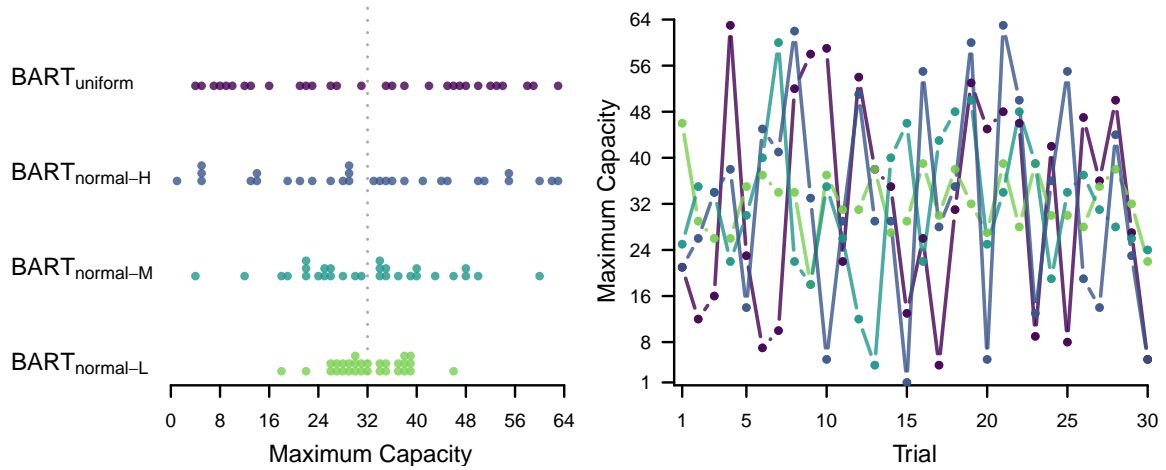
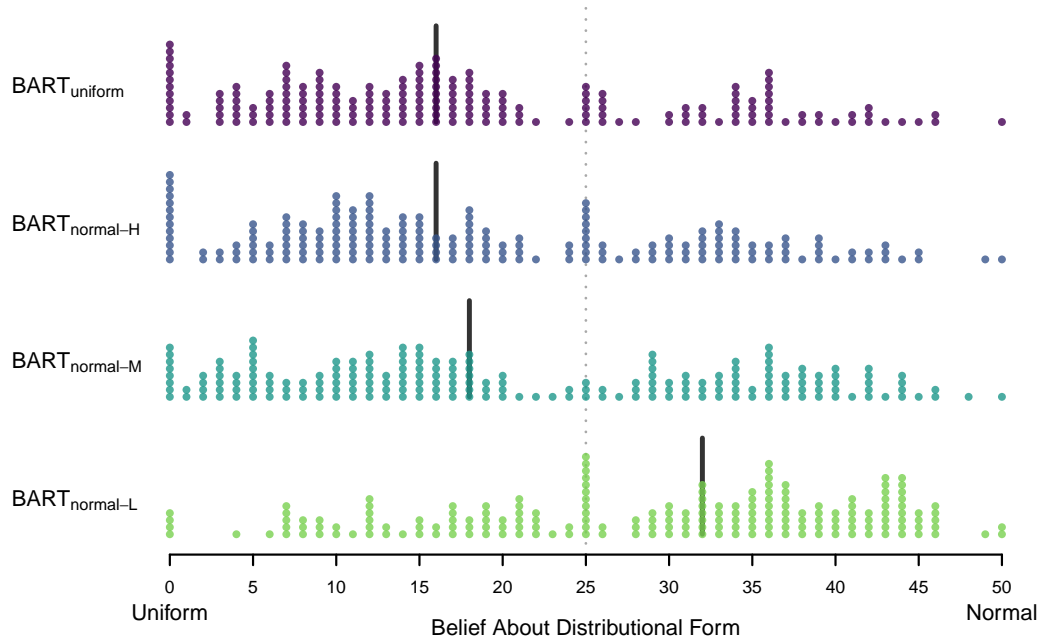


Figure S1. Distributions and the exact sequences of explosion points used in the four BART implementations in studies 1 and 2.

### 3 Participants' General Task Representation in Study 1

Figure S2 shows the distribution of participants' general task representations in study 1. Possibly due to some unclarity in the question, many participants provided ratings below the average of the scale (ranging from 0 to 50). However, as the retest of this assessment with a reformulated item in study 2 showed, also participants who indicated a confidence below the center of the scale, likely thought that explosion points clustered, but with a high variation.



*Figure S2.* Empirical distributions of participants' beliefs that the balloons' explosion points were randomly distributed (rating of 0) vs. that they clustered around a specific value (rating of 50). Beliefs were assessed at the end of the task. Dashed lines indicate the medians of the four experimental conditions. The dotted grey line indicates the center of the scale that was labelled "unsure."

### 4 Learning in the BART

To gauge the participants' learning curves in the task, we computed Spearman correlations between the number of inflations in trial in which the balloons did not explode, and the trial number, separately for each participant. Figure S3 shows the learning across trials separately for the four implementations. The correlations in the four panels are the average of the Spearman correlations calculated per individual within the respective implementation. As can be seen, the larger the variation in the explosion points, the more trials participants need to converge on a number of inflation. Moreover, the between-subjects variance is larger the higher the standard deviation of the explosion points (see also analysis of H3 in the main text).

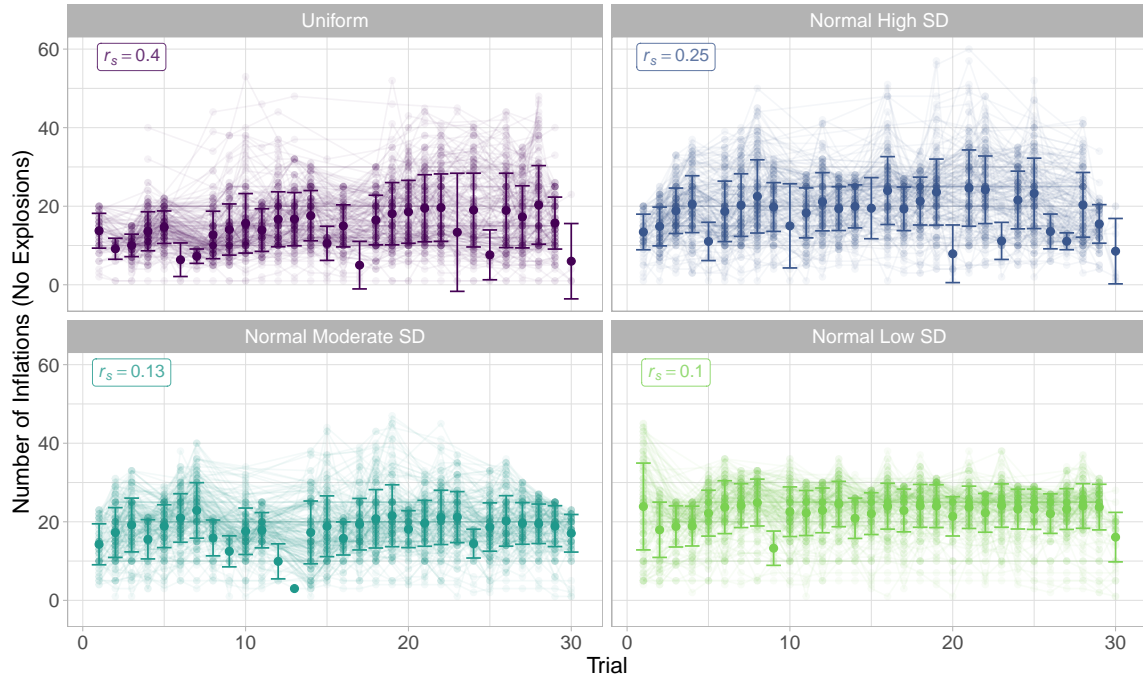


Figure S3. Number of inflations in the BART in trials where balloons did not explode.  $r_s$  = Spearman correlations between the trial and the number of inflations were first computed separately for each participant and then averaged.

## 5 Correlations of the other Risk-Taking Measures

Figure S4 provides an overview over the correlations between the other risk-taking measures used in studies 1 and 2. As can be seen, the propensity measures correlate relatively strongly, whereas the frequency measures only exhibit limited convergent validity.

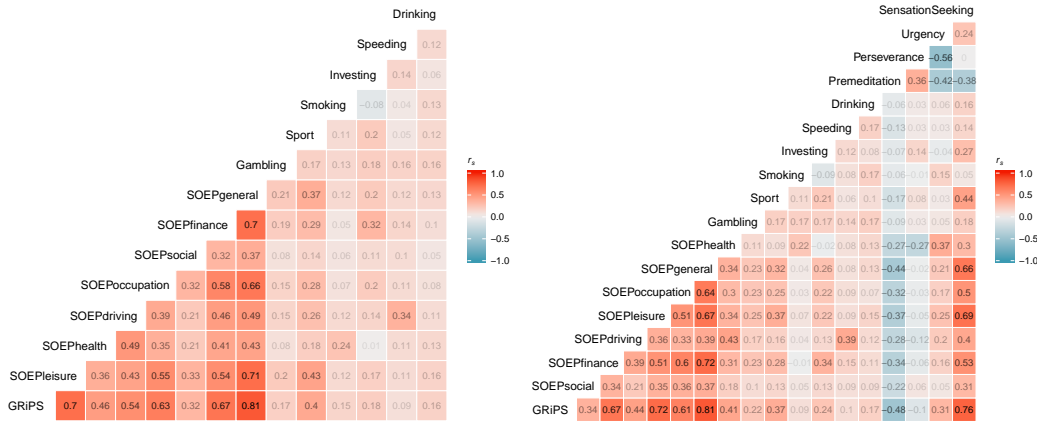


Figure S4. Correlations between the additional risk-taking measures in H4 of studies 1 and 2.

## 6 Posterior Parameter Estimates of H2 and H3

Table S1 presents information about the posteriors of the means, the standard deviations, and the deviance of the means from the objectively optimal behavior, of participants' beliefs about the optimal behavior, as well as their adjusted BART scores.

Table S1

*Posteriors of Beliefs About Optimal Behavior*

Implementation	<i>M</i>	<i>SD</i>	$\Delta$ Optimal
<i>Belief about the optimal behavior</i>			
BART <sub>uniform</sub>	23.47 [21.65, 25.43]	10.31 [8.43, 12.36]	-8.53 [-10.35, -6.57]
BART <sub>normal-H</sub>	25.03 [23.51, 26.65]	8.50 [6.82, 10.37]	-2.97 [-4.49, -1.35]
BART <sub>normal-M</sub>	23.73 [22.69, 24.85]	5.62 [4.45, 6.88]	-1.27 [-2.31, -0.15]
BART <sub>normal-L</sub>	27.07 [26.27, 27.91]	4.27 [3.30, 5.35]	2.07 [1.27, 2.91]
<i>Adjusted BART score</i>			
BART <sub>uniform</sub>	16.70 [15.82, 17.57]	5.80 [5.06, 6.55]	-15.3 [-16.18, -14.43]
BART <sub>normal-H</sub>	21.17 [20.30, 22.04]	5.93 [5.23, 6.65]	-6.83 [-16.18, -5.96]
BART <sub>normal-M</sub>	19.63 [18.96, 20.31]	4.50 [3.74, 5.21]	-5.37 [-6.04, -4.69]
BART <sub>normal-L</sub>	24.44 [23.95, 24.91]	2.43 [1.92, 2.99]	-0.66 [-1.05, -0.09]

*Note:* Posterior estimates of the means, standard deviations, and deviation of the posterior mean from the optimal behavior of both participants' beliefs about the optimal behavior, and their adjusted BART scores, separate for the four implementations. Numbers indicate the median of the posterior distribution, and the 95% highest density interval in brackets. The optimal values for the BART<sub>uniform</sub>, BART<sub>normal-H</sub>, BART<sub>normal-M</sub>, and BART<sub>normal-L</sub> are 32, 28, 25, and 25, respectively.

## 7 Convergent Validity Between the BART and Other Risk-Taking Measures

### 7.1 Study 1

Tables S2 and S3 show the posterior estimates of the correlation between the adjusted BART scores and the total number of explosions per participants with the other risk-taking measures, respectively. The differences between the BART<sub>uniform</sub> and the three implementations of the BART<sub>normal</sub> are presented in Tables S4 and S5.

Table S2

*Convergent Validity of the Adjusted BART Scores and Other Risk-Taking Measures*

Measure	BART <sub>uniform</sub>	BART <sub>normal-H</sub>	BART <sub>normal-M</sub>	BART <sub>normal-L</sub>
GRiPS	.17 [.04, .31]	-.02 [-.15, .11]	<b>.23</b> [.11, .37]	.11 [-.04, .24]
SOEP <sub>general</sub>	.15 [.01, .28]	-.01 [-.15, .13]	<b>.27</b> [.14, .39]	.12 [-.01, .27]
SOEP <sub>driving</sub>	.06 [-.08, .20]	-.06 [-.21, .06]	.16 [.03, .30]	.02 [-.12, .17]
SOEP <sub>finance</sub>	.15 [.01, .29]	.06 [-.08, .19]	.18 [.05, .31]	.08 [-.05, .22]
SOEP <sub>health</sub>	.13 [-.01, .27]	-.06 [-.20, .08]	.14 [.00, .27]	.06 [-.08, .20]
SOEP <sub>leisure</sub>	.17 [.04, .31]	-.07 [-.20, .07]	<b>.24</b> [.12, .38]	.06 [-.09, .19]
SOEP <sub>occupation</sub>	.13 [-.01, .27]	.04 [-.10, .17]	.14 [.00, .27]	.07 [-.07, .21]
SOEP <sub>social</sub>	.12 [-.01, .25]	.06 [-.08, .19]	.09 [-.04, .23]	.07 [-.06, .21]
Drinking	.12 [-.02, .26]	-.16 [-.29, -.02]	.08 [-.06, .22]	.10 [-.04, .23]
Gambling	-.09 [-.23, .04]	-.13 [-.27, .01]	-.02 [-.16, .11]	-.02 [-.16, .12]
Investing	-.02 [-.16, .12]	-.05 [-.19, .08]	.03 [-.10, .17]	-.04 [-.18, .10]
Smoking	.02 [-.12, .16]	-.09 [-.23, .05]	-.03 [-.17, .11]	.03 [-.10, .17]
Speeding	.02 [-.12, .16]	-.11 [-.25, .03]	.04 [-.10, .17]	-.01 [-.14, .14]
Sport	.01 [-.14, .13]	-.10 [-.23, .04]	.08 [-.06, .21]	-.05 [-.20, .08]
Average	.08	-.05	.12	.04

*Note:* Posterior estimates of the correlations between participants' adjusted BART scores and the other risk-taking measures, separate for the four BART implementations. Numbers indicate the median of the posterior distribution, and the 95% highest density interval in brackets. Correlations in boldface indicate conclusive evidence for correlations greater or smaller than 0 according to the ROPE plus 95% HDI rule.

Table S3

*Convergent Validity of the Total Number of Explosions per Participant and the Other Risk-Taking Measures*

Measure	BART <sub>uniform</sub>	BART <sub>normal-H</sub>	BART <sub>normal-M</sub>	BART <sub>normal-L</sub>
GRiPS	.16 [.02, .29]	.01 [-.14, .14]	<b>.26</b> [.13, .39]	.12 [-.02, .26]
SOEP <sub>general</sub>	.11 [-.03, .24]	.02 [-.12, .15]	<b>.30</b> [.18, .43]	.10 [-.03, .24]
SOEP <sub>driving</sub>	.05 [-.09, .18]	-.02 [-.16, .12]	.14 [-.00, .27]	.02 [-.12, .16]
SOEP <sub>finance</sub>	.14 [.01, .28]	.09 [-.05, .22]	<b>.23</b> [.10, .36]	.08 [-.06, .22]
SOEP <sub>health</sub>	.13 [-.00, .27]	-.05 [-.18, .10]	<b>.20</b> [.07, .33]	-.04 [-.18, .10]
SOEP <sub>leisure</sub>	.15 [.01, .28]	-.06 [-.20, .08]	<b>.26</b> [.13, .39]	.08 [-.07, .21]
SOEP <sub>occupation</sub>	.10 [-.04, .23]	.05 [-.08, .19]	.18 [.04, .30]	.07 [-.07, .21]
SOEP <sub>social</sub>	.10 [-.03, .24]	.06 [-.08, .19]	.14 [-.00, .27]	.16 [.03, .29]
Drinking	.06 [-.08, .19]	-.09 [-.22, .05]	.07 [-.06, .21]	.04 [-.10, .18]
Gambling	-.10 [-.23, .04]	-.10 [-.24, .03]	.04 [-.09, .18]	.04 [-.10, .18]
Investing	-.06 [-.20, .08]	-.07 [-.21, .06]	.05 [-.08, .19]	.05 [-.07, .21]
Smoking	-.00 [-.13, .15]	-.10 [-.23, .04]	-.02 [-.15, .12]	-.04 [-.17, .11]
Speeding	-.01 [-.16, .12]	-.08 [-.21, .05]	-.02 [-.17, .11]	.01 [-.13, .15]
Sport	.01 [-.13, .15]	-.13 [-.26, .01]	.08 [-.05, .22]	-.03 [-.17, .11]
Average	.06	-.03	.14	.05

*Note:* Posterior estimates of the correlations between the total number of explosions per participant and the other risk-taking measures, separate for the four BART implementations. Numbers indicate the median of the posterior distribution, and the 95% highest density interval in brackets. Correlations in boldface indicate conclusive evidence for correlations greater or smaller than 0 according to the ROPE plus 95% HDI rule.

Table S4

*Differences Between the BART Implementations in the Convergent Validity of the Adjusted BART Scores and Other Risk-Taking Measures*

Measure	BART <sub>uniform</sub> vs.		
	BART <sub>normal-H</sub>	BART <sub>normal-M</sub>	BART <sub>normal-L</sub>
GRiPS	-.19 [-.39, -.00]	.06 [-.13, .25]	-.07 [-.26, .12]
SOEP <sub>general</sub>	-.16 [-.36, .03]	.12 [-.07, .30]	-.03 [-.22, .16]
SOEP <sub>driving</sub>	-.12 [-.31, .08]	.10 [-.10, .29]	-.04 [-.23, .16]
SOEP <sub>finance</sub>	-.09 [-.29, .10]	.03 [-.16, .22]	-.06 [-.26, .13]
SOEP <sub>health</sub>	-.20 [-.38, .01]	.01 [-.17, .21]	-.07 [-.27, .12]
SOEP <sub>leisure</sub>	-.24 [-.43, -.05]	.07 [-.11, .27]	-.12 [-.31, .08]
SOEP <sub>occupation</sub>	-.09 [-.29, .11]	.01 [-.17, .21]	-.06 [-.25, .14]
SOEP <sub>social</sub>	-.05 [-.24, .15]	-.03 [-.22, .17]	-.05 [-.24, .15]
Drinking	<b>-.28</b> [-.47, -.09]	-.04 [-.23, .16]	-.02 [-.22, .17]
Gambling	-.04 [-.24, .15]	.07 [-.12, .27]	.08 [-.12, .28]
Investing	-.03 [-.23, .17]	.05 [-.14, .25]	-.02 [-.22, .17]
Smoking	-.11 [-.30, .08]	-.06 [-.25, .14]	.02 [-.18, .21]
Speeding	-.13 [-.32, .07]	.02 [-.17, .22]	-.03 [-.23, .17]
Sport	-.11 [-.30, .09]	.07 [-.12, .26]	-.05 [-.24, .16]
Average	-0.13	0.04	-0.04

*Note:* Differences in the posterior estimates of the correlations between participants' adjusted BART scores and the other risk-taking measures, between the BART<sub>uniform</sub> and the BART<sub>normal</sub> implementations. Numbers indicate the median of the posterior distribution, and the 95% highest density interval in brackets. Numbers in boldface indicate conclusive evidence for a correlation of a BART<sub>normal</sub> implementation to be greater or smaller than the one of the BART<sub>uniform</sub> according to the ROPE plus 95% HDI rule.



Table S5

*Differences Between the BART Implementations in the Convergent Validity of the Total Number of Explosions per Participant and the Other Risk-Taking Measures*

Measure	BART <sub>uniform</sub> vs.		
	BART <sub>normal-H</sub>	BART <sub>normal-M</sub>	BART <sub>normal-L</sub>
GRiPS	-.15 [-.34, .04]	.10 [-.08, .30]	-.04 [-.24, .15]
SOEP <sub>general</sub>	-.09 [-.28, .11]	.19 [.01, .38]	-.00 [-.20, .19]
SOEP <sub>driving</sub>	-.06 [-.27, .13]	.10 [-.10, .29]	-.03 [-.23, .17]
SOEP <sub>finance</sub>	-.05 [-.24, .14]	.10 [-.10, .28]	-.06 [-.25, .14]
SOEP <sub>health</sub>	-.18 [-.38, .00]	.06 [-.13, .25]	-.17 [-.36, .03]
SOEP <sub>leisure</sub>	-.21 [-.40, -.01]	.11 [-.07, .30]	-.08 [-.27, .12]
SOEP <sub>occupation</sub>	-.04 [-.23, .15]	.08 [-.11, .27]	-.03 [-.23, .17]
SOEP <sub>social</sub>	-.04 [-.23, .15]	.04 [-.14, .23]	.06 [-.13, .25]
Drinking	-.15 [-.35, .04]	.01 [-.18, .20]	-.02 [-.21, .19]
Gambling	-.00 [-.20, .20]	.13 [-.06, .33]	.13 [-.06, .33]
Investing	-.01 [-.20, .20]	.10 [-.09, .30]	.12 [-.09, .31]
Smoking	-.10 [-.29, .10]	-.02 [-.21, .18]	-.03 [-.23, .16]
Speeding	-.07 [-.26, .13]	-.01 [-.21, .19]	.03 [-.17, .23]
Sport	-.14 [-.32, .06]	.07 [-.13, .27]	-.03 [-.23, .16]
Average	-.09	.08	-.01

*Note:* Differences in the posterior estimates of the correlations between the total number of explosions per participants and the other risk-taking measures, between the BART<sub>uniform</sub> and the BART<sub>normal</sub> implementations. Numbers indicate the median of the posterior distribution, and the 95% highest density interval in brackets. Numbers in boldface indicate conclusive evidence for a correlation of a BART<sub>normal</sub> implementation to be greater or smaller than the one of the BART<sub>uniform</sub> according to the ROPE plus 95% HDI rule.

## 7.2 Study 2

Tables S6 and S7 show the posterior estimates of the correlation between the adjusted BART scores and the total number of explosions per participants with the other risk-taking measures, respectively. The differences between the BART<sub>uniform</sub> and the three implementations of the BART<sub>normal</sub> are presented in Tables S8 and S9.

Table S6

*Convergent Validity of the Adjusted BART Scores and Other Risk-Taking Measures in Study 2*

Measure	BART <sub>uniform</sub>	BART <sub>normal-H</sub>	BART <sub>normal-M</sub>	BART <sub>normal-L</sub>
GRiPS	<b>.24</b> [.08, .38]	.02 [-.12, .18]	<b>.21</b> [.08, .36]	-.01 [-.17, .14]
SOEP <sub>general</sub>	<b>.22</b> [.07, .36]	.05 [-.11, .20]	.19 [.04, .34]	.06 [-.09, .22]
SOEP <sub>driving</sub>	<b>.22</b> [.07, .36]	.03 [-.14, .17]	.14 [-.01, .30]	.04 [-.11, .19]
SOEP <sub>finance</sub>	.16 [.01, .31]	.19 [.04, .33]	.10 [-.05, .25]	.04 [-.11, .20]
SOEP <sub>health</sub>	.04 [-.10, .19]	.02 [-.13, .17]	.17 [.01, .31]	.09 [-.06, .23]
SOEP <sub>leisure</sub>	.15 [.00, .30]	.05 [-.11, .21]	<b>.21</b> [.07, .36]	.06 [-.08, .22]
SOEP <sub>occupation</sub>	.09 [-.08, .23]	.06 [-.09, .21]	.15 [-.00, .29]	-.01 [-.16, .14]
SOEP <sub>social</sub>	.12 [-.04, .27]	.04 [-.12, .19]	<b>.26</b> [.12, .40]	.13 [-.03, .28]
Drinking	.07 [-.07, .23]	-.12 [-.27, .03]	.06 [-.10, .21]	.03 [-.12, .17]
Gambling	.05 [-.10, .20]	.04 [-.11, .19]	.00 [-.14, .17]	.05 [-.11, .19]
Investing	-.14 [-.29, .00]	-.05 [-.20, .10]	.07 [-.09, .21]	-.02 [-.18, .13]
Smoking	.13 [-.03, .27]	-.09 [-.24, .07]	<b>-.20</b> [-.35, -.06]	.10 [-.05, .24]
Speeding	-.01 [-.16, .15]	.02 [-.14, .17]	-.02 [-.16, .14]	.06 [-.09, .22]
Sport	.08 [-.07, .24]	-.06 [-.22, .08]	.05 [-.10, .20]	.00 [-.14, .16]
Perseverance	-.01 [-.16, .15]	.01 [-.14, .16]	-.06 [-.21, .09]	.09 [-.06, .24]
Premeditation	-.10 [-.25, .05]	-.02 [-.17, .14]	-.07 [-.23, .07]	.07 [-.08, .23]
Urgency	-.01 [-.16, .14]	-.04 [-.19, .11]	.02 [-.12, .18]	.00 [-.14, .16]
Sensation Seeking	.17 [.02, .32]	.09 [-.08, .24]	.18 [.03, .32]	.08 [-.06, .24]
Average	.08	.01	.08	.05

*Note:* Posterior estimates of the correlations between participants' adjusted BART scores and the other risk-taking measures, separate for the four BART implementations. Numbers indicate the median of the posterior distribution, and the 95% highest density interval in brackets. Correlations in boldface indicate conclusive evidence for correlations greater or smaller than 0 according to the ROPE plus 95% HDI rule.

Table S7

*Convergent Validity of the Total Number of Explosions per Participant and the Other Risk-Taking Measures in Study 2*

Measure	BART <sub>uniform</sub>	BART <sub>normal-H</sub>	BART <sub>normal-M</sub>	BART <sub>normal-L</sub>
GRiPS	<b>.23</b> [.10, .37]	.07 [-.08, .23]	<b>.20</b> [.06, .35]	-.06 [-.21, .09]
SOEP <sub>general</sub>	<b>.24</b> [.09, .39]	.07 [-.08, .22]	.15 [-.01, .29]	.01 [-.15, .17]
SOEP <sub>driving</sub>	.16 [.02, .31]	.09 [-.07, .24]	.13 [-.01, .28]	.07 [-.08, .23]
SOEP <sub>finance</sub>	.14 [-.01, .29]	.19 [.03, .33]	.15 [.02, .31]	.05 [-.10, .20]
SOEP <sub>health</sub>	-.00 [-.15, .16]	.07 [-.08, .22]	.18 [.03, .33]	.10 [-.04, .25]
SOEP <sub>leisure</sub>	.19 [.04, .34]	.05 [-.10, .21]	.19 [.05, .34]	.05 [-.10, .20]
SOEP <sub>occupation</sub>	.09 [-.05, .25]	.05 [-.11, .19]	.16 [.02, .30]	-.04 [-.19, .10]
SOEP <sub>social</sub>	.12 [-.02, .28]	.05 [-.10, .20]	<b>.26</b> [.12, .40]	.14 [-.01, .28]
Drinking	.11 [-.04, .26]	-.09 [-.23, .06]	.01 [-.14, .16]	.09 [-.06, .23]
Gambling	.10 [-.06, .24]	.09 [-.06, .24]	.04 [-.11, .18]	.15 [-.00, .30]
Investing	-.13 [-.28, .02]	-.09 [-.24, .06]	.03 [-.12, .19]	.00 [-.15, .15]
Smoking	.02 [-.13, .17]	-.03 [-.18, .12]	-.19 [-.34, -.04]	.11 [-.05, .25]
Speeding	-.03 [-.17, .13]	.05 [-.11, .20]	-.03 [-.18, .12]	.15 [.01, .31]
Sport	.14 [-.02, .28]	-.08 [-.24, .07]	.02 [-.13, .17]	.07 [-.07, .22]
Perseverance	-.06 [-.22, .09]	-.03 [-.19, .11]	-.13 [-.28, .02]	.11 [-.04, .26]
Premeditation	-.12 [-.27, .03]	-.04 [-.20, .11]	-.07 [-.22, .08]	.09 [-.07, .24]
Urgency	.03 [-.12, .17]	.01 [-.15, .16]	.04 [-.10, .20]	-.00 [-.16, .15]
Sensation Seeking	<b>.22</b> [.06, .36]	.11 [-.04, .26]	.16 [.00, .30]	.06 [-.09, .20]
Average	.08	.03	.07	.06

*Note:* Posterior estimates of the correlations between the total number of explosions per participant and the other risk-taking measures, separate for the four BART implementations. Numbers indicate the median of the posterior distribution, and the 95% highest density interval in brackets. Correlations in boldface indicate conclusive evidence for correlations greater or smaller than 0 according to the ROPE plus 95% HDI rule.

Table S8

*Differences Between the BART Implementations in the Convergent Validity of the Adjusted BART Scores and Other Risk-Taking Measures in Study 2*

Measure	BART <sub>uniform</sub> vs.		
	BART <sub>normal-H</sub>	BART <sub>normal-M</sub>	BART <sub>normal-L</sub>
GRiPS	-.22 [-.43, -.01]	-.02 [-.22, .20]	-.24 [-.45, -.03]
SOEP <sub>general</sub>	-.17 [-.38, .04]	-.03 [-.23, .19]	-.15 [-.36, .06]
SOEP <sub>driving</sub>	-.20 [-.41, .01]	-.08 [-.29, .13]	-.18 [-.39, .02]
SOEP <sub>finance</sub>	.03 [-.17, .25]	-.06 [-.27, .16]	-.12 [-.33, .10]
SOEP <sub>health</sub>	-.02 [-.24, .19]	.13 [-.08, .34]	.05 [-.16, .27]
SOEP <sub>leisure</sub>	-.10 [-.31, .12]	.06 [-.15, .27]	-.09 [-.29, .13]
SOEP <sub>occupation</sub>	-.03 [-.25, .18]	.06 [-.16, .27]	-.10 [-.31, .12]
SOEP <sub>social</sub>	-.08 [-.29, .14]	.15 [-.06, .36]	.02 [-.20, .23]
Drinking	-.19 [-.40, .03]	-.01 [-.23, .20]	-.04 [-.26, .17]
Gambling	-.01 [-.22, .21]	-.04 [-.26, .17]	-.00 [-.22, .20]
Investing	.09 [-.12, .31]	.21 [-.01, .42]	.12 [-.09, .33]
Smoking	-.21 [-.41, .01]	<b>-.33</b> [-.55, -.12]	-.03 [-.24, .18]
Speeding	.03 [-.19, .24]	-.00 [-.22, .22]	.08 [-.14, .29]
Sport	-.14 [-.35, .08]	-.03 [-.24, .19]	-.08 [-.30, .12]
Perseverance	.02 [-.19, .24]	-.05 [-.27, .17]	.10 [-.11, .31]
Premeditation	.08 [-.13, .30]	.02 [-.20, .24]	.17 [-.05, .37]
Urgency	-.03 [-.24, .19]	.04 [-.18, .25]	.01 [-.20, .23]
Sensation Seeking	-.08 [-.30, .12]	.01 [-.20, .22]	-.09 [-.30, .12]
Average	-0.07	0.00	-0.03

*Note:* Differences in the posterior estimates of the correlations between participants' adjusted BART scores and the other risk-taking measures, between the BART<sub>uniform</sub> and the BART<sub>normal</sub> implementations. Numbers indicate the median of the posterior distribution, and the 95% highest density interval in brackets. Numbers in boldface indicate conclusive evidence for a correlation of a BART<sub>normal</sub> implementation to be greater or smaller than the one of the BART<sub>uniform</sub> according to the ROPE plus 95% HDI rule.

Table S9

*Differences Between the BART Implementations in the Convergent Validity of the Total Number of Explosions per Participant and the Other Risk-Taking Measures in Study 2*

Measure	BART <sub>uniform</sub> vs.		
	BART <sub>normal-H</sub>	BART <sub>normal-M</sub>	BART <sub>normal-L</sub>
GRiPS	-.16 [-.36, .05]	-.03 [-.24, .17]	<b>-.29</b> [-.49, -.08]
SOEP <sub>general</sub>	-.18 [-.39, .04]	-.09 [-.31, .11]	-.23 [-.44, -.03]
SOEP <sub>driving</sub>	-.07 [-.28, .14]	-.03 [-.25, .18]	-.09 [-.30, .12]
SOEP <sub>finance</sub>	.05 [-.16, .26]	.02 [-.19, .23]	-.09 [-.31, .12]
SOEP <sub>health</sub>	.08 [-.14, .29]	.18 [-.03, .39]	.10 [-.11, .32]
SOEP <sub>leisure</sub>	-.15 [-.36, .05]	-.00 [-.22, .20]	-.15 [-.35, .07]
SOEP <sub>occupation</sub>	-.05 [-.26, .17]	.06 [-.14, .29]	-.13 [-.35, .07]
SOEP <sub>social</sub>	-.07 [-.28, .15]	.14 [-.07, .35]	.02 [-.19, .23]
Drinking	-.20 [-.42, .02]	-.10 [-.31, .12]	-.02 [-.24, .19]
Gambling	-.01 [-.23, .20]	-.06 [-.27, .16]	.05 [-.15, .28]
Investing	.04 [-.18, .25]	.17 [-.05, .38]	.13 [-.09, .34]
Smoking	-.05 [-.27, .16]	-.21 [-.43, .00]	.09 [-.12, .31]
Speeding	.08 [-.13, .29]	-.00 [-.23, .20]	.18 [-.03, .39]
Sport	-.21 [-.44, -.01]	-.11 [-.33, .11]	-.07 [-.28, .15]
Perseverance	.03 [-.19, .24]	-.07 [-.28, .15]	.17 [-.04, .38]
Premeditation	.08 [-.14, .29]	.06 [-.16, .27]	.21 [-.00, .42]
Urgency	-.01 [-.23, .21]	.02 [-.21, .22]	-.03 [-.24, .19]
Sensation Seeking	-.10 [-.31, .10]	-.06 [-.27, .15]	-.15 [-.36, .06]
Average	-.05	-.01	-.02

*Note:* Differences in the posterior estimates of the correlations between the total number of explosions per participants and the other risk-taking measures, between the BART<sub>uniform</sub> and the BART<sub>normal</sub> implementations. Numbers indicate the median of the posterior distribution, and the 95% highest density interval in brackets. Numbers in boldface indicate conclusive evidence for a correlation of a BART<sub>normal</sub> implementation to be greater or smaller than the one of the BART<sub>uniform</sub> according to the ROPE plus 95% HDI rule.

## 8 Test–Retest Reliability of the Four BART Implementations

Table S10 shows the posterior estimates of the test–retest reliabilities, as well as the coefficient of variation—a standardized measure of dispersion—of participants’ beliefs about the optimal behavior, the adjusted BART scores, and the total number of explosions per participant, separate for the four BART implementations. As can be seen in Table S10, the coefficient of variation differs somewhat between the extreme implementations (BART<sub>uniform</sub>, and BART<sub>normal-L</sub>) for the adjusted BART scores and the total number of explosions per participant.

Table S10

### *Test–Retest Reliabilities of BART Indicators*

Implementation	<i>r</i>	<i>cv</i>
<i>Belief about the optimal value</i>		
BART <sub>uniform</sub>	.40, [ .27, .52]	0.44
BART <sub>normal-H</sub>	.41, [ .28, .53]	0.4
BART <sub>normal-M</sub>	.26, [ .11, .40]	0.38
BART <sub>normal-L</sub>	.43, [ .31, .55]	0.33
<i>Adjusted BART score</i>		
BART <sub>uniform</sub>	.59, [ .48, .69]	0.33
BART <sub>normal-H</sub>	.73, [ .65, .80]	0.27
BART <sub>normal-M</sub>	.65, [ .55, .73]	0.24
BART <sub>normal-L</sub>	.42, [ .29, .54]	0.16
<i>Total number of explosions per participant</i>		
BART <sub>uniform</sub>	.47, [ .35, .59]	0.27
BART <sub>normal-H</sub>	.66, [ .57, .73]	0.31
BART <sub>normal-M</sub>	.63, [ .53, .72]	0.46
BART <sub>normal-L</sub>	.48, [ .37, .59]	0.55

*Note:* Coefficient of variation and posterior estimates of the test–retest correlations, separate for the four implementations. Reported are the medians of the posterior distributions, and the 95% highest density interval in brackets.

## 9 Summary of Indicators From Studies 1 and 2

Table S11 presents means and SDs of the different indicators collected in the two empirical studies, separate for the four BART conditions.

Table S11

*Mean (SD) of the Indicators Collected in the two Empirical Studies*

Indicator	BART <sub>uniform</sub>	BART <sub>normal-H</sub>	BART <sub>normal-M</sub>	BART <sub>normal-L</sub>
<i>Study 1</i>				
Adj. BART score	16.83 (6.15)	21.20 (6.18)	19.68 (4.95)	23.31 (4.38)
Posterior belief	25.19 (13.02)	27.16 (12.16)	26.49 (10.91)	29.71 (10.51)
# explosions	8.98 (2.61)	9.91 (3.38)	7.46 (3.71)	6.11 (3.50)
Bel. dist.	18.06 (12.43)	19.15 (12.6)	20.85 (13.67)	29.01 (12.30)
GRiPS	2.22 (0.90)	2.26 (1.00)	2.34 (0.94)	2.29 (0.91)
SOEP general	3.82 (2.42)	3.97 (2.50)	4.14 (2.49)	3.73 (2.33)
SOEP driving	2.74 (2.40)	2.91 (2.71)	2.98 (2.51)	2.73 (2.58)
SOEP faith	4.78 (2.45)	4.33 (2.46)	4.57 (2.47)	4.38 (2.41)
SOEP finance	3.63 (2.43)	3.61 (2.52)	3.91 (2.61)	3.78 (2.47)
SOEP health	3.33 (2.41)	3.15 (2.36)	3.12 (2.41)	3.28 (2.45)
SOEP leisure	4.44 (2.63)	4.55 (2.80)	4.80 (2.59)	4.25 (2.60)
SOEP occupation	4.17 (2.64)	4.03 (2.47)	4.21 (2.56)	4.22 (2.62)
p(cigarettes)	0.23	0.31	0.29	0.27
p(drink)	0.76	0.74	0.73	0.78
p(gamble)	0.66	0.57	0.62	0.56
p(invest)	0.65	0.57	0.65	0.54
p(speed)	0.78	0.76	0.78	0.77
p(sport)	0.26	0.23	0.24	0.24
N	190	195	197	190
<i>Study 2</i>				
Adj. BART score	17.91 (5.66)	22.49 (5.64)	20.88 (4.62)	24.18 (3.16)
Posterior belief	28.38 (11.47)	31.08 (11.83)	29.69 (10.84)	29.81 (8.80)
# explosions	9.52 (2.66)	10.11 (2.96)	7.52 (3.41)	5.28 (2.87)
Bel. dist.	17.64 (27.44)	19.22 (28.10)	19.98 (27.25)	25.51 (25.56)
GRiPS	2.22 (0.94)	2.14 (0.92)	2.28 (0.84)	2.24 (0.86)
SOEP gen	3.93 (2.25)	3.75 (2.55)	3.94 (2.30)	3.81 (2.29)
SOEP driving	2.76 (2.32)	2.61 (2.57)	2.84 (2.31)	2.58 (2.19)
SOEP faith	4.56 (2.45)	4.51 (2.56)	4.93 (2.27)	4.54 (2.41)
SOEP finance	4.00 (2.48)	3.49 (2.47)	3.87 (2.64)	3.66 (2.45)
SOEP health	3.27 (2.30)	3.13 (2.39)	3.11 (2.38)	3.06 (2.14)
SOEP leisure	4.58 (2.59)	4.31 (2.81)	4.73 (2.50)	4.49 (2.66)
SOEP occupation	4.4 (2.63)	3.83 (2.55)	4.27 (2.42)	3.91 (2.44)
UPPS SS	2.28 (0.74)	2.19 (0.71)	2.24 (0.66)	2.28 (0.73)
UPPS perseverance	3.14 (0.57)	3.18 (0.56)	3.16 (0.53)	3.16 (0.51)
UPPS premeditation	3.18 (0.49)	3.25 (0.48)	3.24 (0.45)	3.26 (0.53)
UPPS urgency	2.07 (0.62)	2.08 (0.71)	2.11 (0.67)	2.08 (0.62)
p(cigarettes)	0.20	0.25	0.29	0.25
p(drink)	0.76	0.74	0.75	0.74
p(gamble)	0.62	0.5	0.62	0.53
p(invest)	0.66	0.59	0.67	0.65
p(speed)	0.78	0.73	0.78	0.79
p(sport)	0.23	0.25	0.25	0.24
N	157	158	157	160

*Note:* “Posterior belief” = Participants’ beliefs about the optimal number of inflations after having completed the BART (see hypothesis 2). “# explosions” = Number of balloon explosions per participant in the BART. “Bel. dist.” = Participants’ beliefs about the distributional form (see hypothesis 1). “p(…)” = Proportion of participants who indicated to engage in the respective behaviors > 0 times a day.

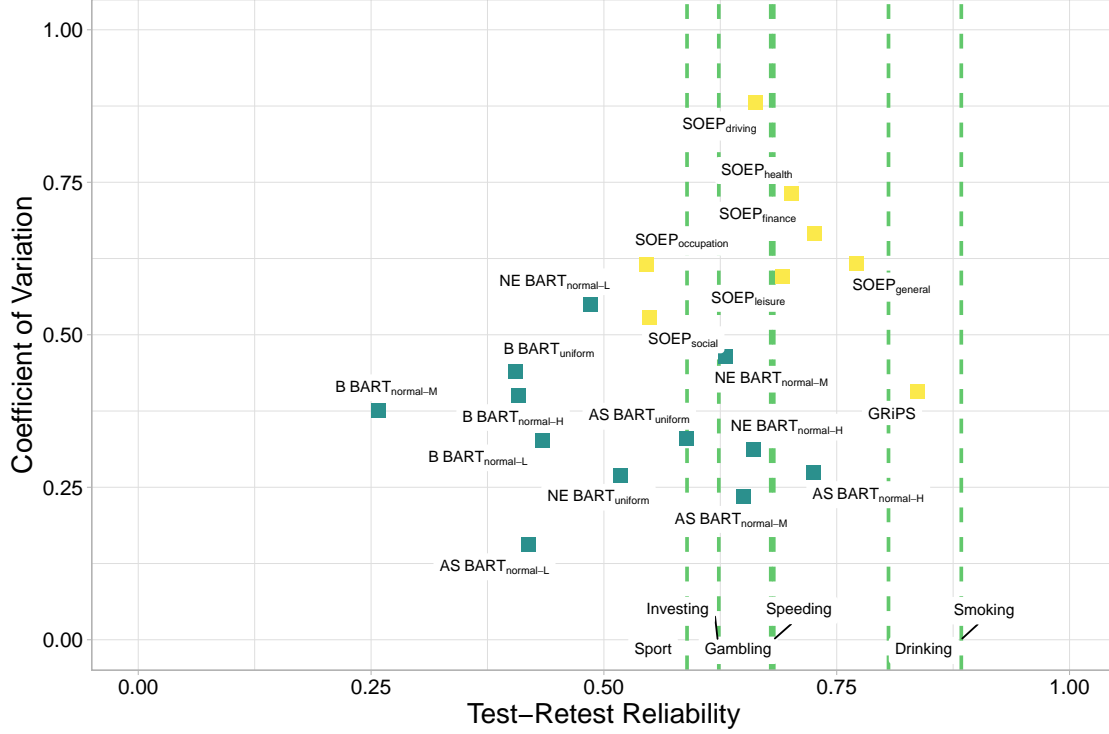


Figure S5. Test-retest reliabilities and coefficient of variation (a standardized measure of dispersion) of the risk-taking measures employed in studies 1 and 2. Indicators of the BART appear in blue. Propensity measures appear in yellow. Frequency measures appear in green. As we binarized the frequency measures and thus the coefficient of variation is not interpretable for them, only their reliabilities are shown as vertical lines.

## 10 Test-Retest Reliability of All Risk-Taking Measures

Figure S5 presents an overview over the test-retest reliabilities and the coefficients of variation of the different risk-taking measures assessed in studies 1 and 2. As we binarized the frequency measures and thus the coefficient of variation is not interpretable for them, we only plot their test-retest reliabilities as vertical lines.

## 11 Simulating the Effect of the Distributional Form

To test the effects the different distributions of explosion points have on the BART scores, adjusted BART scores, and the number of explosions, we ran several simulations. First, for each of the four distributions of explosion points, we simulated  $9 \times 1,000$  participants with a fixed target value per group, ranging from one inflation to the maximum capacity of the balloons in steps of 8, namely,  $target \in \{1, 8, 16, \dots, 64\}$ . Then, in each of 30 trials (i.e., as in our empirical studies) the simulated participants probabilistically determined a current target to pump to, by sampling a value from the distribution centered at their fixed target value (e.g., 1, 8, 16, ... 64) and a  $SD = 5$ . Finally, we let these simulated participants carry out “their” pumping behavior, separately for the four experimental conditions that we had implemented in our empirical studies.



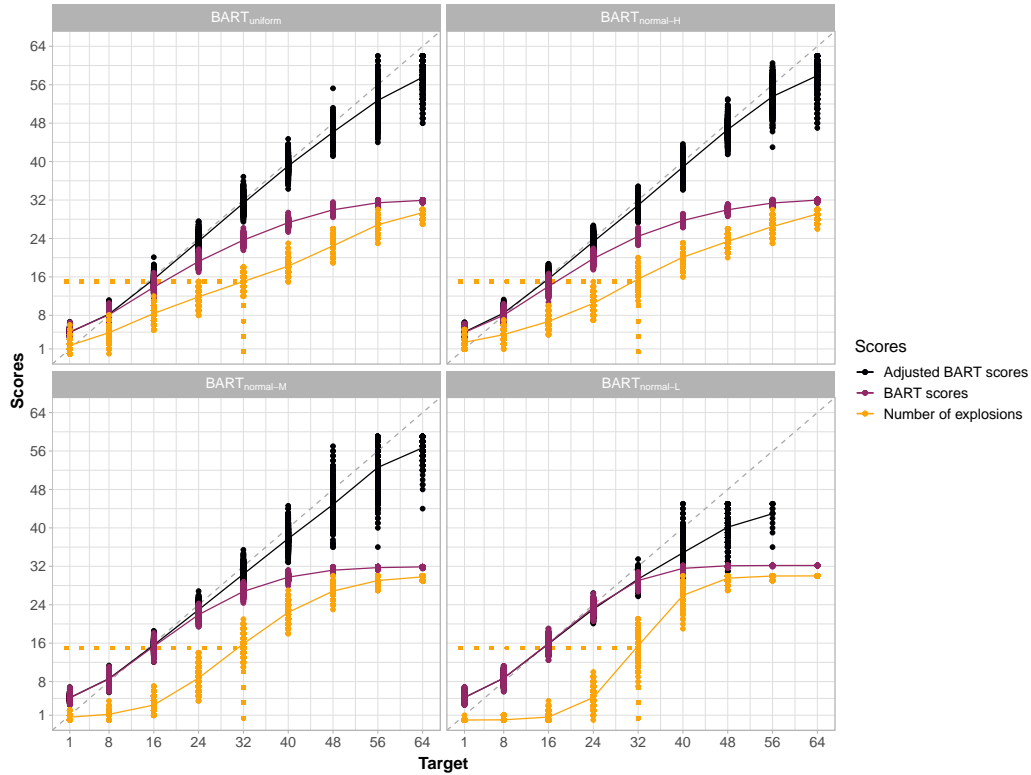


Figure S6. Simulation analysis of players of the BART that use different target values to pump to, with  $target \in \{1, 8, 16, \dots, 64\}$ . Each panel shows the results obtained using one of the four distribution of explosion points used in the Studies 1 and 2. The dashed grey lines indicate a perfect relationship between the target values and the scores. The dotted yellow line indicates the point where half the balloons exploded.

As can be seen in Figure S6, with an increasingly higher target a participant will naturally experience relatively more explosions (orange lines). For the uniform and normal-H distributions, these increases were virtually linear and almost identical (not surprisingly, as we intentionally implemented the normal-H distribution to match the statistical properties of the uniform distribution). Whereas the increase in the number of explosions follows a more sigmoid shape for the normal-M and normal-L distributions, around the mean breaking point the number of explosions are identical across the four distributions. This is the case because we explicitly implemented the three normal distributions to have the same means as the uniform distribution. Thus, based on a model with a fixed target value, we would expect similar average adjusted BART scores across the four distributions based on this simulation analysis.

## 12 References

- Frey, R., Pedroni, A., Mata, R., Rieskamp, J., & Hertwig, R. (2017). Risk preference shares the psychometric structure of major psychological traits. *Science Advances*, 3(10), e1701381. doi: 10.1126/sciadv.1701381

- Schürmann, O., Frey, R., & Pleskac, T. J. (2018). The role of risk perception in dynamic risk-taking behavior. *Journal of Behavioral Decision Making*, 1–12. doi: 10.1002/bdm.2098
- Steiner, M. D., & Frey, R. (2020). Beyond risk preference? Exploring alternative ways to modeling risk taking. *Manuscript in Preparation*.