

Mueller, Steffen Q.; Ring, Patrick; Schmidt, Maria

Working Paper

Forecasting economic decisions under risk: The predictive importance of choice-process data

Hamburg Contemporary Economic Discussions, No. 66

Provided in Cooperation with:

University of Hamburg, Chair for Economic Policy

Suggested Citation: Mueller, Steffen Q.; Ring, Patrick; Schmidt, Maria (2019) : Forecasting economic decisions under risk: The predictive importance of choice-process data, Hamburg Contemporary Economic Discussions, No. 66, ISBN 978-3-942820-47-9, University of Hamburg, Faculty of Business, Economics and Social Sciences, Chair for Economic Policy, Hamburg

This Version is available at:

<https://hdl.handle.net/10419/200699>

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.



Universität Hamburg

DER FORSCHUNG | DER LEHRE | DER BILDUNG

Faculty of Business,
Economics and Social Sciences
Chair for Economic Policy

STEFFEN Q. MUELLER / PATRICK RING / MARIA SCHMIDT

FORECASTING ECONOMIC DECISIONS UNDER RISK: THE PREDICTIVE IMPORTANCE OF CHOICE-PROCESS DATA

Urban
Transport
Media
Sports
Socio-
Regional
Real Estate
Architectural

HAMBURG CONTEMPORARY

ECONOMIC DISCUSSIONS

NO. 66

Hamburg Contemporary Economic Discussions

University of Hamburg
Faculty of Business, Economics and Social Sciences
Chair for Economic Policy
Von-Melle-Park 5
20146 Hamburg | Germany
Tel +49 40 42838 - 4622
Fax +49 40 42838 - 6251
<https://www.wiso.uni-hamburg.de/en/fachbereich-vwl/professuren/maennig/home.html>
Editor: Wolfgang Maennig

Steffen Q. Mueller
University of Hamburg
Faculty of Business, Economics and Social Sciences
Chair for Economic Policy
Von-Melle-Park 5
20146 Hamburg | Germany
Tel +49 40 42838 - 5297
Fax +49 40 42838 - 6251
Steffen.Mueller@uni-hamburg.de

Patrick Ring
Kiel Institute for the World Economy
Social and Behavioral Approaches to Global Problems
Kiellinie 66
24105 Kiel | Germany
Tel +49 431 8814 - 626
patrick.ring@ifw-kiel.de

Maria Schmidt
Kiel University
Department of Psychology
Olshausenstraße 62
24118 Kiel | Germany

Photo Cover: HQuality/Shutterstock.com

ISSN 1865 - 2441 (Print)
ISSN 1865 - 7133 (Online)

ISBN 978-3-942820-46-2 (Print)
ISBN 978-3-942820-47-9 (Online)

Steffen Q. Mueller, Patrick Ring, Maria Schmidt

Forecasting economic decisions under risk:

The predictive importance of choice-process data¹

Abstract: We investigate various statistical methods for forecasting risky choices and identify important decision predictors. Subjects ($n=44$) are presented a series of 50/50 gambles that each involves a potential gain and a potential loss, and subjects can choose to either accept or reject a displayed lottery. From this data, we use information on 8800 individual lottery gambles and specify four predictor-sets that include different combinations of input categories: lottery design, socioeconomic characteristics, past gambling behavior, eye-movements, and various psychophysiological measures that are recorded during the first three seconds of lottery-information processing. The results of our forecasting experiment show that choice-process data can effectively be used to forecast risky gambling decisions; however, we find large differences among models' forecasting capabilities with respect to subjects, predictor-sets, and lottery payoff structures.

Keywords: Forecasting, lottery, risk, choice-process tracing, experiments, machine learning, decision theory

JEL: C44, C45, C53, D87, D91

Version: January 2019

1 Introduction

We conduct a lottery gambling experiment to assess the predictive importance of choice-process data (CPD) by investigating various statistical methods for forecasting economic decisions under risk. Our data covers 44 subjects that each played 200 lotteries, thereby resulting in a total of 8800 observations. In each lottery, participants were offered a 50/50 gamble that involves a potential gain and a potential loss and subjects could decide to either play or not play the specified lottery.

To assess how much information on choice patterns can be inferred from CPD, eye movements and a variety of psychophysiological measures are recorded during the first

¹ We thank colleagues and seminar participants in Hamburg (University) and Kiel (IfW, University) for comments and suggestions. Particularly, we are grateful to Ingo Fiedler, René Glawion, Wolfgang Maennig, and Ulrich Schmidt. This project was partly funded by an Add-On Fellowship for Interdisciplinary Economics by the Joachim Herz Stiftung.

three seconds of lottery information processing. We specify four predictor-sets that correspond to various combinations of the following input categories: lottery design variables, socioeconomic characteristics, past gambling behavior, and information on eye-movements and psychophysiological reactions (pupil size, pulse, respiration and skin conductance).

Simple heuristics and decision rules can often explain a large share of the heterogeneity in individual decision making and they have been demonstrated to outperform advanced and knowledge-intensive methods in various forecasting domains (Goldstein & Gigerenzer, 2009). Similar to Stahl (2018), who evaluates lottery choices on the basis of judgmental heuristics, we discuss our empirical findings in comparison to two simple decision rules: a naive forecast that uses the most frequently observed class in the training data as a prediction for all test records and a statistical decision rule that is based on a lottery's expected value.

In addition to the naive decision rules that we use as forecasting benchmarks, we investigate methods of statistical learning, including linear (e.g., elastic net regression), non-linear (e.g., artificial neural networks) and tree-based ensemble algorithms (e.g., random forest). We apply a systematic grid-search to select values for sensitive model hyperparameters. We use subjects as strata and train algorithms on resampled partitions of 80% of the data via 10-fold stratified cross-validation (CV). The remaining 20% of the data are used as a hold-out test set. First, we assess the predictive importance of individual predictor-sets. Then, we develop a full model that is comprised of all input categories.

Data that are derived from choice-process tracing such as gaze patterns and pupil dilation provide valuable information on a wide range of economic behaviors. Numerous examples exist: Wang et al. (2016) demonstrate that pupil size can be used to predict truth telling and deception in sender-receiver games; Devetag et al. (2016) use eye-movements to identify various types of motivational attitudes in strategic behavior; Brocas et al. (2014) use mouse-tracking data to cluster participants according to look-up

patterns; and Reutskaja et al. (2011) employ an eye-tracking analysis to investigate search dynamics under time pressure and choice overload.

However, CPD that are derived from sources other than eye-tracking and attention-tracking also provide valuable information on predictable choice patterns; CPD are used to investigate facial reactions to emotional stimuli (Dimberg et al., 2002), monitor psychophysiological markers of stress and arousal to predict emotional states (Bendoly, 2013), and measure the extents to which neural activities reveal food preferences without choices (Smith, Douglas Bernheim, Camerer, & Rangel, 2014) and bargaining process-features predict disagreements (Camerer, Nave, & Smith, 2018). However, although there is an increasing interest in the analysis of CPD, the vast majority of studies focus on relatively few models, datasets, and CPD measures (e.g., Krol & Krol (2017)).

This study makes several important contributions: First, to the best of our knowledge, we provide the first systematic evaluation of statistical learning algorithms for forecasting gambling decisions. Second, we consider several types of CPD and assess their predictive importance by analyzing various predictor-sets and input categories. Third, this study demonstrates how an algorithmic modeling strategy can be used to make predictive inferences on correlations in complex systems using an example of the interplay between risky decisions and emotional and cognitive processes.

The out-of-sample accuracy results are summarized as follows: A naive decision rule that is produced by predicting all test records to be not-playing yields 54% accuracy; classifying subjects' choices on the basis of lotteries' expected values (EV) yields 67% accuracy; and the forecasting method that yields the highest mean CV accuracy for the model, which uses all input categories, results in 87% accuracy. The best CV-fit models that use CPD result in a mean accuracy of 62%, even though the corresponding predictor-sets do not include information on the lottery design variables. However, we observe large differences among models' forecasting capabilities for individual subjects, predictor-sets, and lottery payoff structures. In contrast to lotteries with positive expected values (PEV), many subjects strictly reject lotteries that have negative expected values

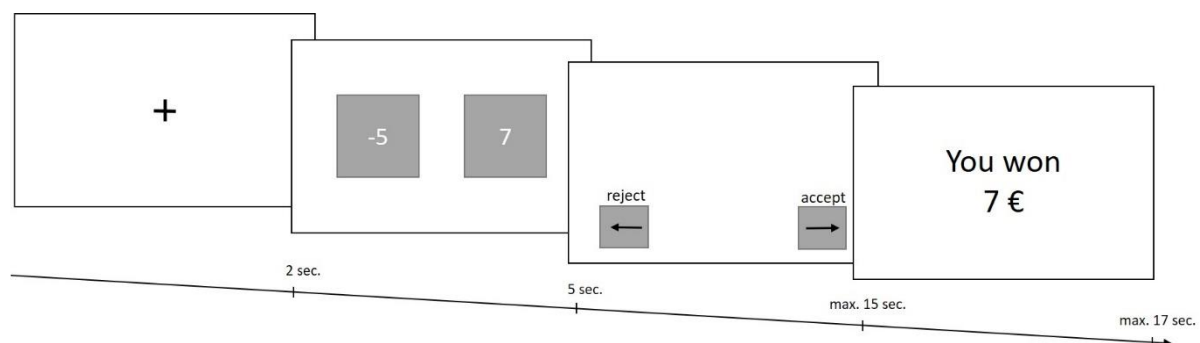
(NEV), though a small number of subjects play a relatively large share of lotteries with NEV. As a result, the model accuracy for the predictor-sets that include lottery design variables is substantially higher for NEV games compared to PEV games.

2 Experimental setup and descriptive analysis

We recruited 44 participants (mean age=28 years, SD=4) from the general population of Kiel, Germany via online advertisements. All participants gave written informed consent and could decide to discontinue participation at any time. The risk task consisted of 200 rounds. In each round, participants were offered a 50/50 gamble that involved a potential gain and a potential loss. Across lotteries, we manipulated the potential gain and loss (range of gains: +1 EUR to +20 EUR; range of losses -1 EUR to -10 EUR; both in 1 Euro steps). Participants could accept or reject the offered lottery by pressing a button (left or right arrow). It was not possible for subjects to execute their final gambling choices during the first three seconds that a lottery was displayed. Subjects received immediate feedback about the outcome of a lottery, if it was accepted. During the experiment, subjects were notified when they reached the 67th and 133th rounds. The order of the lotteries, the representation of gains and losses on the screen and the accept/reject buttons were randomized for each participant.

Figure 1 displays the sequence of events and the displayed screens for one round of lottery gambling vs. time (seconds).

Figure 1 Sequence of events and screens for one round of lottery gambling by time

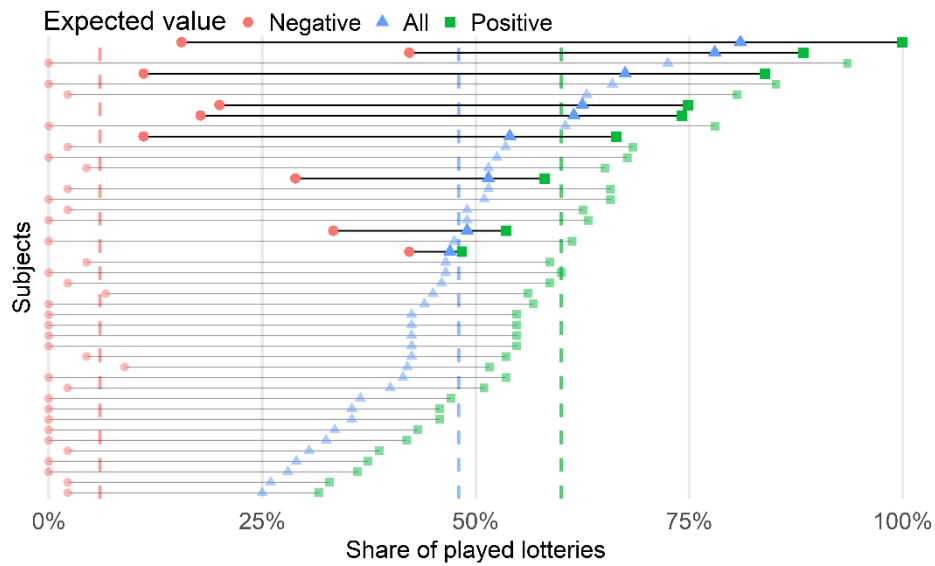


The first picture (left) shows a fixation cross and indicates that a lottery will be shown soon. The second picture shows the newly offered lottery for three seconds. The third picture shows the arrows that must be pressed to accept or reject the previously displayed lottery for a maximum of ten seconds. After a decision has been executed, the realized outcome is displayed. In this context, we define the *reaction time* as starting from the 5th second and ending when a final decision is executed. Hence, the reaction time is measured with respect to the time interval in which subjects could execute their gambling choices, not with respect to the first three seconds of lottery information processing. Furthermore, the fourth screen is omitted for rejected gambles.

All subjects started with an endowment of 10 EUR and at the end of the experiment, one round was randomly selected for the final payoff. If the subject had rejected the selected lottery, she kept the initial endowment of 10 EUR. If the subject had accepted the lottery, its outcome was realized and added to (subtracted from) the initial endowment in the case of a gain (loss). Detailed descriptions of the experimental design and descriptive statistics are included in the appendix.

Figure 2 shows the relative share of the number of played lotteries for each of the 44 individual subjects for all 200 lotteries: 155 PEVL and 45 NEVL. We note that there are ten lotteries that are associated with an EV of zero and we classify them as PEVL. The mean EV over all lotteries is 2.50 EUR (SD=3.30), the mean EV for NEV lotteries (NEVL) is -1.90 EUR (Min=-4.50, Max=-0.50, SD=1.20), and the mean EV for PEV lotteries (PEVL) is 3.80 EUR (Min=0.00, Max=9.50, SD=2.50).

Figure 2 Share of played lotteries by subject and lotteries' expected values



Notes: Each of the 44 subjects decided on 200 lotteries: 155 with positive expected values (PEV) and 45 with negative expected values (NEV). Subjects that play five or more NEV lotteries are highlighted (bold). Dashed lines correspond to mean shares of played lotteries by expected value. Subjects are ordered according to the highest share of the played lotteries (All).

The mean subject plays 48% of all lotteries, but there is high heterogeneity in the individual shares of played lotteries across subjects and within and between NEVL and PEVL. There are 21 subjects who do not play any NEVL and 14 subjects who play one to four NEVL. The remaining nine subjects play five or more NEVL. In contrast to an average share of played NEVL of 6%, the average share of played PEVL is 60% and no strict decision boundary that can be used to classify subjects as risk-averse for playing PEVL is identified. The distribution of the share of played PEVL is much smoother across individuals compared to NEVL.

3 Methods and data description

3.1 Motivation for an algorithmic modeling approach

Forecasting human decisions is difficult because there are many variables that impact choices in various and interdependent ways (Kleinberg, et al., 2018) and it is well known that there is considerable variation in individual risk-taking behaviors (Bruhin et al., 2010). As an example, the choice patterns that are presented in Figure 2 show that there is high variation in the individual shares of played lotteries for both NEVL and PEVL. Many studies that investigate issues that are related to CPD and risk preferences apply

data modeling strategies by employing linear and parametric models that require non-linear dependencies and higher-order interactions to be specified explicitly. However, it is unclear to what degree human behavior, cognitive and emotional processes, CPD, and experimental designs are related (Frederick, 2005; Loewenstein, 1999; Reutskaja et al., 2011; Ring, 2015).

Based on the high complexity and large number of relevant factors that determine the relationship between decisions and CPD we employ an algorithmic modeling strategy. Similar to Kleinberg et al. (2018, 2015) and Reutskaja et al. (2011) we posit that using a predictive inference approach to analyze systems that are complex by nature, such as the causal link between cognitive and emotional processes, can yield valuable insights on systematic patterns in human decision making.

3.2 Tested models and hyperparameters

Our study evaluates 1750 unique model specifications that correspond to four general modeling frameworks: naive, generalized linear, nonlinear and tree-based ensemble methods. In addition to a logistic regression model and linear penalized regressions methods (Elastic net), we evaluate methods that both automatically account for nonlinear interdependencies and higher-order interactions without the need for prespecification: Support Vector Machines (SVM), artificial neural networks (ANN), Random forests (RF), and tree-based gradient boosting machines (GBM). A general description for the machine learning methods that we employ in this study can be found in Hastie et al. (2009). Furthermore, we evaluate two simple forecasting benchmarks. The first predicts all test records as the most frequent class that we observe in the training data (not-played). We describe this naive forecast as a risk-averse decision rule (RDR) because the most risk-averse behavior is to reject all lotteries to receive the 10 EUR endowment as a final payoff. The second benchmark can be described as a simple statistical decision rule (SDR) that classifies gambling-decisions according to a lottery's expected value. According to this SDR, we classify NEVL ($EV < 0$) as not-played and PEVL ($EV \geq 0$) as played.

Except for the predefined naive forecasting rules and the logistic regression model, each of the machine learning algorithms that is evaluated in this study requires for selecting values for model-specific hyperparameters. We follow standard practice for determining sensitive parameter values via a systematic grid-search (e.g., Lessmann & Voß (2017)).

Table 1 presents an overview of the models and the corresponding hyperparameters that we evaluate in this study. Except for the two naive forecasting methods, we evaluate all models on the basis of four differently specified predictor-sets (see appendix for details).

Table 1 Summary of evaluated forecasting methods

Classification method	# Models	Hyperparameters	Tested values
<i>Naive Forecasts</i>			
Risk-averse Decision rule	1	-	-
Statistical Decision rule	1	-	-
<i>Generalized linear</i>			
Logistic regression	$4 * 1 = 4$	-	-
Elastic net*	$4 * 11 * 5$ $= 220$	Alpha* Lambda	0, 0.1, 0.2, ..., 1 (0.01, 0.025, 0.05, 0.1, 0.15)
<i>Nonlinear</i>			
Support vector machines (SVM)	$4 * 1 * 5 * 10$ $= 200$	Basis function kernel Cost Inverse kernel width (sigma)	<i>Radial</i> 1, 2, ..., 10 (0.01, 0.025, 0.05, 0.1, 0.15)
Artificial neural network (ANN)	$4 * 10 * 5$ $= 200$	Activation function No. of nodes in the hidden layer Weight decay	<i>Sigmoid</i> 1, 2, ..., 10 0.1, 0.2, ..., 0.5
<i>Tree-based ensemble</i>			
Random forest (RF)	$4 * 1 * 1 * 5 * 5$ $= 100$	Splitting rule No. of ensembled trees Min. no. of samples in each leaf node No. of predictors in each split**	<i>Gini</i> 1500 1, 5, 10, 20, 40 $ D_i (0.1, 0.2,$ 0.4, 0.8, 1)
Gradient boosting machines (GBM)	$4^5 = 1024$	Interaction depth No. of ensembled trees Shrinkage Min. no. of samples in each leaf node	5, 10, 20, 40 20, 40, 80, 160 0.1, 0.4, 0.7, 1 10, 40, 70, 100
No. of evaluated models	1750		

Notes: Except for the two naive forecasting methods, we evaluate all models on the basis of four differently specified predictor-sets. $|D_i|$ refers to the number of predictors for predictor-set i with $i = \{P, A, LSG, LSGPA\}$. * Depending on the hyperparameter alpha, the model takes the form of a lasso, ridge, or elastic net regression model. ** In addition to RF, we implicitly evaluate a bagged classification and regression tree (CART) ensemble by offsetting RF's random sub-space method via setting the number of randomly chosen predictors to the total number of predictor-set specific predictor variables, i.e., for $|D_i| * 1$.

3.3 Description of variables and data cleaning

Table 2 briefly describes all the predictors that we employ in this paper: lottery design variables (L), such as loss and win values; participants' socioeconomic characteristics (S), such as gender and age, education and income levels; past gambling behavior (G); psychophysiological CPD (P), such as pupil dilation and skin conductance measurements; and attention CPD (A).

Table 2 Description of specified predictor variables

Lottery design (L)
5 variables that are related to win and loss values and expected payoff
4 variables that are related to the lottery trials
2 variables that are related to the area where a win or a loss value is displayed on the computer screen
Socioeconomic (S)
44 variables that account for subject-specific effects
5 variables that are related to qualifications and profession levels
4 variables that are related to education levels
4 variables that are related to age levels and gender
3 variables that are related to income levels
Gambling behavior^a (G)
14 variables that are related to subjects' past gambling behaviors and decision outcomes
Psychophysiological reactions^b (P)
10 variables that are related to a subject's skin conductance rates
4 variables that are related to a subject's blood volume pulse
4 variables that are related to a subject's blood volume pulse's amplitude
4 variables that are related to a subject's respiration rate
4 variables that are related to a subject's respiration rate's amplitude
4 variables that are related to a subject's heart rate
4 variables that are related to a subject's body temperature
4 variables that are related to a subject's pupils' sizes
Attention^b (A)
3 variables that are related to the lengths of time that a subject gazes at the displayed win and loss values

Notes: ^a With respect to the past five displayed lotteries and the corresponding choices and outcomes.

^b Except skin conductance data, all choice-process data are measured exclusively during the first three seconds of lottery information processing and relate to minimum, maximum and mean values and the difference between minimum and maximum values. Skin conductance data have a delay of approximately one second and we adjusted the time window accordingly (Boucsein, 2012).

We encode factor variables as dummy variables and specify 126 predictors. Then, we exclude the most frequently observed level for each category as the corresponding reference group. This encoding results in a total of 119 predictors: 42 numeric and 77 binary variables. Except for skin conductance data (SCD), all CPD are measured with respect to the first three seconds of lottery information processing. During this time period, it was not possible for subjects to execute their final gambling choices. Consequently, none of the predictor-sets include information on the reaction time. However, SCD has a delay of approximately one second and we adjusted the time window accordingly (Boucsein, 2012). SCD includes information on the number of significant (above-threshold) reactions and tonic and phasic activity based on the continuous decomposition analysis as

suggested by Benedek & Kaernbach (2010). All other CPD predictors relate to minimum, maximum and mean values and the difference between minimum and maximum values. In this context, with respect to gaze patterns (A), we follow Krol & Krol (2017) and argue that the complete scan-path sequence of information can be considered superfluous for games in which only two numbers are required for processing all information about a lottery's payoff structure.

Moreover, the P and A predictor-sets do not contain any information that relates to lottery design variables such as displayed win and loss values. However, for A , we implicitly include the side of the computer screen (left vs. right) on which a lottery's win and loss values are displayed in specifying subjects' look-up patterns.

In our data cleaning process, we exclude 220 observations because they include information on past gambling behavior in the *LSG* and *LSGPA* data and we exclude 212 observations due to missing eye-tracking data. The final dataset includes 8368 records. A detailed description of the empirical specifications is presented in the appendix.

3.4 Implementation

The classification task can be modeled as a choice, which is denoted as Y , that we specify as a binary outcome: $Y = 1$ for playing and $Y = 0$ for not-playing a lottery. Let us consider a function $f_i(\cdot)$ that relates Y to a predictor-set D_i with $i = \{P, A, LSG, LSGPA\}$. The predictor-sets include aggregated combinations of input categories: physiological CPD (P); attention CPD (A); lottery design, socioeconomic characteristics, and past gambling behavior (*LSG*); and a full model that is comprised of all input categories (*LSGPA*). We include dummy variables for subject-specific effects in all four datasets. The objective is to identify well-approximating functional relationships that relate the specified predictor-sets to the decision outcome by learning and identifying systematic choice patterns from the training data.

We use subjects as strata in both randomly selecting 80% of the cleaned data as a training sample and tuning models' hyperparameters via 10-fold stratified CV on the basis of the training sample. The remaining 20% of the data are used as a hold-out test set to

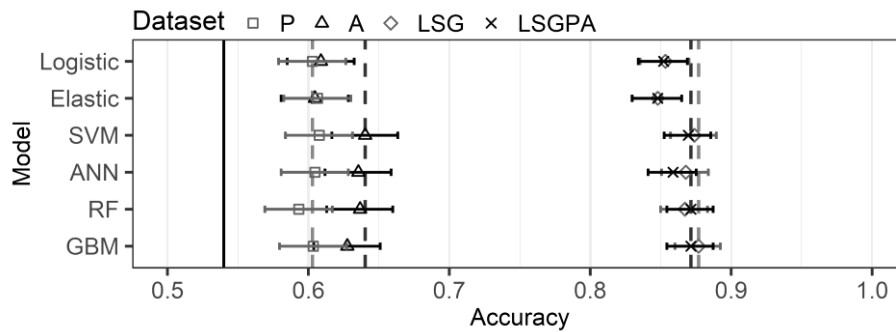
produce reasonable estimates of accuracy. This sampling procedure utilizes 6711 observations for model training and 1657 observations for model testing. We focus on assessing the predictive importance of CPD on binary gambling choices and we have no interest in maximizing the accuracy for specified outcomes, e.g., true-positive predictions. Since the average share of played lotteries is relatively balanced, we use the classification accuracy to assess models' predictive capabilities. Moreover, we set the cut-off value for classifying a record as played to a predicted probability of 50%. Last, we center and scale all numeric predictors separately with respect to the corresponding 10 training CV fold-sets and the test records.

4 Model performance evaluation

The out-of-sample accuracy results differ only marginally from the mean 10-fold CV performance on the training data. In the following sections, we focus on a visual inspection and a discussion of selected forecasting results. Detailed results for hyperparameter tuning and out-of-sample performance are included in the appendix. However, in terms of the Bonferroni-adjusted p-values for pairwise t-tests for detecting significant differences in models' CV accuracies, for the *LSG* and *LSGPA* predictor-sets, both generalized linear model frameworks (logistic and elastic) are consistently outperformed by data-driven algorithms (SVM, ANN, RF, and GBM).

For all results that we report in this study, we set the models' hyperparameters to the values that yield the highest mean CV accuracies. Figure 3 shows the out-of-sample classification accuracy for the 1657 records that are included in the test data.

Figure 3 Out-of-sample classification accuracy for playing a 50/50 gamble

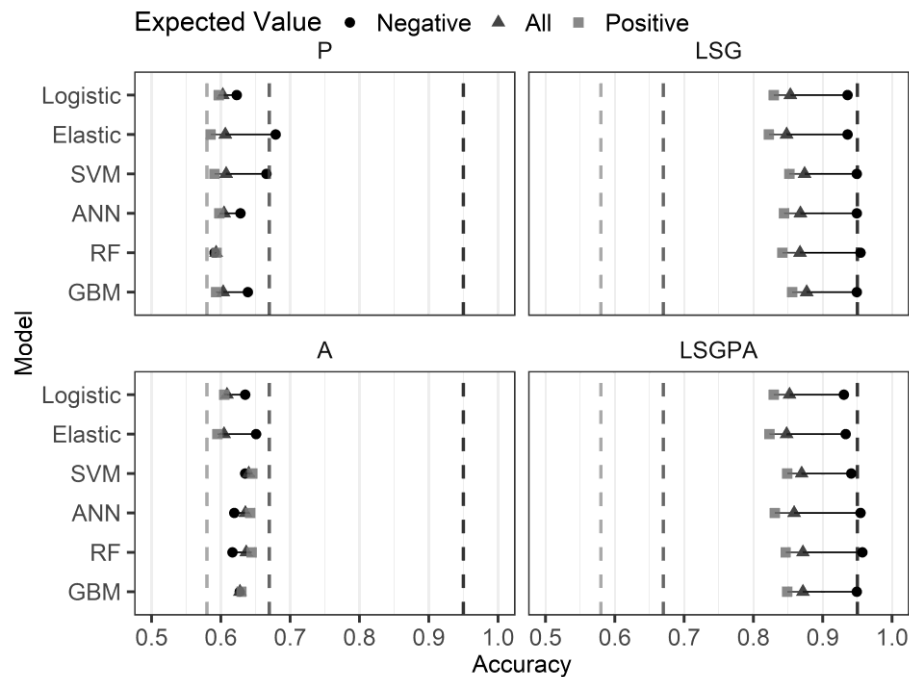


Notes: Test (training) data consist of 1657 (6711) records and the models' hyperparameters are chosen as the values that yield the highest mean 10-fold CV accuracy using subjects as strata. We evaluate logistic (Logistic) and penalized regression models (Elastic), support vector machines (SVM), artificial neural networks (ANN), random forests (RF), and tree-based gradient boosting machines (GBM) on the basis of psychophysiological (*P*) and attention (*A*) choice-process data; lottery design, socioeconomic characteristics, and information on past gambling behavior (*LSG*); and a full model that is comprised of all input categories (*LSGPA*). The error bars correspond to 95% confidence intervals. The dashed lines correspond to the accuracy that is realized on the basis of the models that yield the highest mean 10-fold CV training data accuracy. The solid line is a naive forecast that yields a test data accuracy of 54% by predicting all records as not-playing.

The naive forecast benchmark that is produced by the RDR that predicts all test records as not-playing (see the solid line in Figure 3) results in an out-of-sample accuracy of 54%. The models that we chose according to their best CV accuracy are indicated by dashed lines and we observe only marginal differences between the best CV-fit models and the best test-fit models. The best out-of-sample accuracy for *P* is observed with 61% (Elastic, SVM), for *A* with 64% (SVM, ANN, RF), for *LSG* with 88% (GBM), and for *LSGPA* with 87% (SVM, RF, GBM). In comparison to the generalized linear models (Logistic and Elastic), our results suggest that nonlinear models (SVM and ANN) and tree-based ensembles (RF and GBM) better utilize the information that is provided by the individual look-up patterns (*A*) and produce more accurate predictions on the basis of *LSG* and *LGSPA*. Figure 3 shows that *A* can be associated with a higher predictive importance than *P*, but there appear to exist only marginal differences between the *LSG* and *LSGPA* within-model accuracy. However, no dominant approach is identified among the nonlinear and tree-based algorithms when comparing mean test accuracy only.

In Figure 4, we present models' out-of-sample accuracy results in terms of lotteries' EV and using the SDR as a forecasting benchmark instead of the naive RDR.

Figure 4 Out-of-sample classification accuracy for playing a 50/50 gamble by lotteries' expected values

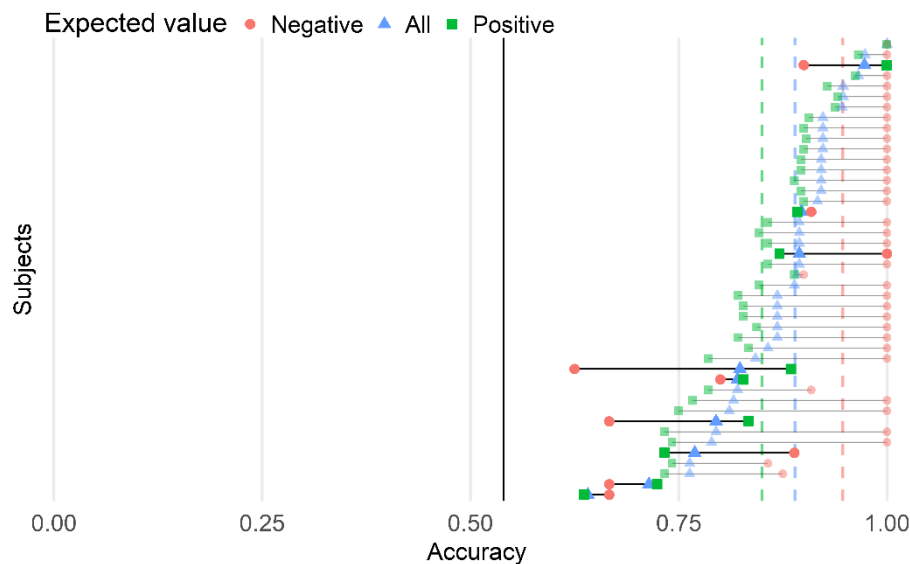


Notes: Out-of-sample classification accuracy for positive expected value (PEV), negative expected value (NEV), and all lotteries (All) for playing a 50/50 gamble. Test (training) data consist of 1657 (6711) records and models' hyperparameters are chosen as the values that yield the highest mean 10-fold CV accuracy using subjects as strata. We evaluate logistic and penalized linear regression models (Elastic), support vector machines (SVM), artificial neural networks (ANN), random forests (RF), and tree-based gradient boosting machines (GBM) on the basis of psychophysiological (*P*) and attention (*A*) choice-process predictors; lottery design, socioeconomic characteristics, and information on past gambling behavior (*LSG*); and a full-model that is comprised of all input categories (*LSGPA*). The dashed lines represent the predictive accuracy on the test data that results from predicting a lottery with negative expected values (NEV) as not-played (95%), positive expected values (PEV) as played (58%), and the corresponding accuracy based on all lotteries (67%).

SDR outperforms all tested models for *A* and *P* over all lotteries (All) with a test accuracy of 67%. In contrast to *A* and *P*, the *LSG* and *LSGPA* datasets include information on lottery design variables and both significantly outperform the SDR forecast. Moreover, the results for *A* and *P* in terms of EV do not vary substantially, except for the Elastic and SVM forecasts. For *A*, the Elastic model results in 65% for NEVL and 60% for PEVL. In contrast, for *P*, the SVM and Elastic models yield the best out-of-sample performance for NEVL, with accuracies of 67% and 68%, respectively. However, the SDR forecast yields an accuracy of 95%; hence, the vast majority of subjects do not play many NEVL. Similarly, the SDR results demonstrate that the mean subject played 58% out of all PEVL that are included in the test data.

Exemplarily, in Figure 5 we present the subject-specific test accuracy results by lotteries' EV for GBM for on the basis of the *LSGPA* predictor-set (see appendix section 3.5 for all individual model results).

Figure 5 GBM-LSGPA subject-specific out-of-sample classification accuracy for playing a 50/50 gamble by lotteries' expected values



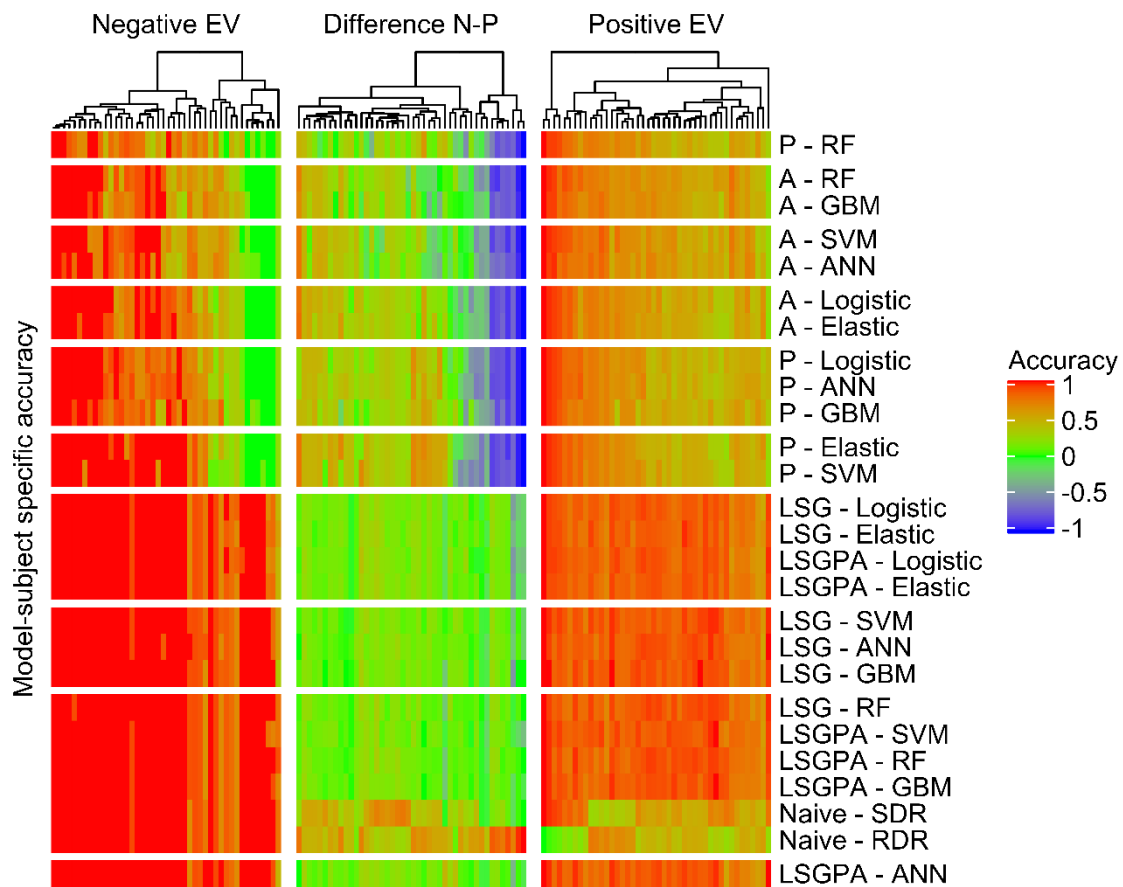
Notes: Out-of-sample classification accuracy for 44 subjects based on positive expected value (PEV), negative expected value (NEV), and all lotteries (All) for playing a 50/50 gamble. The test (training) data consist of 1657 (6711) records and models' hyperparameters are chosen as the values that yield the highest mean 10-fold CV accuracy using subjects as strata. The forecast is produced using a tree-based stochastic gradient boosting machine (GBM) on the basis of a predictor-set that is comprised of input categories (*LSGPA*) that relate to psychophysiological (*P*) and attention (*A*) choice-process data and lottery designs, socioeconomic characteristics, and information on past gambling behavior (*LSG*). Subject are ordered according to overall accuracy (All). The 9 subjects that play five or more NEV lotteries (NEVL) with respect to the 45 NEVL that are included in the full experimental data are highlighted (bold; see Figure 2, section 2). The solid line is a naive forecast that realizes an accuracy of 54% on the test data by predicting all records as not-playing.

On average, GBM on the basis of *LSGPA* produces accurate forecasts for gambling decisions across individuals. However, while the predictions for subjects that play a relatively large share of lotteries (highlighted) are less accurate for NEVL, they are more accurate for PEVL when compared to the large group of subjects that almost play none of the NEVL that are included in their corresponding test records. Moreover, GBM produces 100%-accurate forecasts for NEVL for 73% of the subjects (32 out of 44).

Figure 5 shows the results for GBM on the basis of *LSGPA* only. To assess the importance of CPD on forecasting risky choices for NEVL and PEVL across all models' test results, in

Figure 6 we present the predictor-set-, and subject-specific out-of-sample accuracy results for NEVL, PEVL, and the difference between NEVL and PEVL. To highlight differences across subject-specific results we cluster subjects (rows) with respect to their corresponding accuracy separately for NEVL, PEVL, and the difference between NEVL and PEVL via k-means hierarchical clustering.

Figure 6 Subject-specific out-of-sample classification accuracy for playing a 50/50 gamble by lotteries' expected values



Notes: Out-of-sample classification accuracy results for 44 subjects for negative expected value lotteries (NEVL) positive expected value lotteries (PEVL), and the difference between NEVL and PEVL for playing a 50/50 gamble. The test (training) data consist of 1657 (6711) records and models' hyperparameters are chosen as the values that yield the highest mean 10-fold CV accuracy using subjects as strata. To highlight differences across subjects' accuracy results, we first cluster subjects (rows) according to their corresponding accuracy separately for lotteries' EV via (hierarchical) k-means clustering, and then split models' subject-specific accuracy results across PEVL, NEVL, and the difference between NEVL and PEVL into 10 groups via k-means clustering; for both using the Euclidean distance measure. We evaluate logistic and penalized linear regression models (Elastic), support vector machines (SVM), artificial neural networks (ANN), random forests (RF), and gradient boosting machines (GBM) on the basis of psychophysiological (*P*) and attention (*A*) choice-process predictors; lottery-design, socioeconomic characteristics, and information on past gambling behavior (*LSG*); and a full-model that comprises all input categories (*LSGPA*). RDR is naive forecast that predicts all records as not-playing, and SDR results from predicting a lottery with NEV ($EV < 0$) as not-playing and PEV ($EV \geq 0$) as playing.

Figure 6 supports our previous results and reveals systematic differences in models' forecasting capabilities between subject-specific NEVL and PEVL choices. The corresponding subjects are characterized by the large differences within their NEVL results (red vs. green) and between their NEVL and PEVL results (red vs. blue). For example, Elastic and SVM for P result in more 100% accurate forecasts for individual subjects for NEVL than all other model-predictor-set combinations on the basis of A and P , and, except for the naive forecasts, they are the model-predictor-set combinations that correspond to the highest number of subject-specific forecast differences that yield a higher accuracy for NEVL than for PEVL. For LSG and $LSGPA$, that both provide information on lotteries' payoff structures, except for RDR and SDR all evaluated forecasting methods produce accurate forecasts for the vast majority of subjects, especially for NEVL. In this context, we highlight that the SDR and RDR on average produce highly accurate forecasts for NEVL choices that are on par with those of advanced machine learning algorithms. While there exist strategies for exploiting systematic patterns in between-model accuracy correlation differences across subjects and lottery designs, we do not evaluate heterogeneous model ensemble approaches in this study; we leave this for further research.

5 Conclusions

Our study shows that CPD can be used to forecast risky decisions. However, we observe high variation in the differences in predictive measures, and within and across models and predictor-sets. We find that the forecasting accuracy highly depends on the lotteries' payoff structures and individual risk-taking behaviors. Such heterogeneity in differences across subjects and classification measures can have severe consequences for real life applications of artificial intelligence systems. As an example, let us consider the evaluation of methods for assessing job candidates' future risk preferences or of CPD that were recorded during a suspect's interrogation for detecting lies and truths. Here, the heterogeneity can be important for maximizing a model's predictive power with respect to specific measures such as the ratio of sensitivity to specificity.

References

- Benedek, M., & Kaernbach, C. (2010). A continuous measure of phasic electrodermal activity. *Journal of Neuroscience Methods*, 190(1), 80–91.
- Bendoly, E. (2013). Real-time feedback and booking behavior in the hospitality industry: Moderating the balance between imperfect judgment and imperfect prescription. *Journal of Operations Management*, 31(1–2), 62–71.
- Boucsein, W. (2012). Electrodermal activity. *New York: Springer*.
- Brocas, I., Carrillo, J. D., Wang, S., & Camerer, C. F. (2014). Imperfect choice or imperfect attention? Understanding strategic thinking in private information games. *Review of Economic Studies*, 81(3), 944–970.
- Bruhin, A., Fehr-Duda, H., & Epper, T. (2010). Risk and Rationality: Uncovering Heterogeneity in Probability Distortion. *Econometrica*, 78(4), 1375–1412.
- Camerer, C. F., Nave, G., & Smith, A. (2018). Dynamic Unstructured Bargaining with Private Information: Theory, Experiment, and Outcome Prediction via Machine Learning. *Management Science*, forthcoming, 1–25.
- Devetag, G., Di Guida, S., & Polonio, L. (2016). An eye-tracking study of feature-based choice in one-shot games. *Experimental Economics*, 19(1), 177–201.
- Dimberg, U., Thunberg, M., & Grunedal, S. (2002). Facial reactions to emotional stimuli: Automatically controlled emotional responses. *Cognition and Emotion*, 16(4), 449–471.
- Frederick, S. (2005). Cognitive Reflection and Decision Making. *Journal of Economic Perspectives*, 19(4), 25–42.
- Goldstein, D. G., & Gigerenzer, G. (2009). Fast and frugal forecasting. *International Journal of Forecasting*, 25(4), 760–772.
- Hastie, T., Tibshirani, R., & Friedman, J. (2009). *The Elements of Statistical Learning* (2nd ed.). *New York: Springer*.
- Kleinberg, J., Lakkaraju, H., Leskovec, J., Ludwig, J., & Mullainathan, S. (2018). Human decisions and machine predictions. *Quarterly Journal of Economics*, 133(1), 237–293.
- Kleinberg, J., Ludwig, J., Mullainathan, S., & Obemeyer, Z. (2015). Prediction Policy

- Problems. *American Economics Review*, 105(5), 491–495.
- Krol, M., & Krol, M. (2017). A novel approach to studying strategic decisions with eye-tracking and machine learning. *Judgment and Decision Making*, 12(6).
- Lessmann, S., & Voß, S. (2017). Car resale price forecasting: The impact of regression method, private information, and heterogeneity on forecast accuracy. *International Journal of Forecasting*, 33(4), 864–877.
- Loewenstein, B. G. (1999). Emotions in Economic Theory and Economic Behavior. *American Economic Review*, 90(2), 256–260.
- Reutskaja, E., Nagel, R., Camerer, C. F., & Rangel, A. (2011). Search Dynamics in Consumer Choice under Time Pressure: An Eye-Tracking Study. *American Economic Review*, 101(2), 900–926.
- Ring, P. (2015). The framing effect and skin conductance responses. *Frontiers in Behavioral Neuroscience*, 9(188), 1–7.
- Smith, A., Douglas Bernheim, B., Camerer, C. F., & Rangel, A. (2014). Neural activity reveals preferences without choices. *American Economic Journal: Microeconomics*, 6(2), 1–36.
- Wang, J. T., Spezio, M., & Camerer, C. F. (2016). Pinocchio 's Pupil: Using Eyetracking and Pupil Dilation to Understand Truth Telling and Deception in Sender-Receiver Games. *American Economic Review*, 100(3), 984–1007.

Appendix

1 Introduction

This appendix provides additional information on the experiment design, data cleaning and sampling process, empirical specifications, and descriptive statistics, and is complementing the main text by providing the detailed results for model training, hyperparameter tuning, and models' out-of-sample performances. Furthermore, this appendix replicates some text and results from the main paper for reasons of clarity.

2 Data and experimental design

In this section, we provide additional information on the experiment design, descriptive statistics, the data cleaning process, and variable specifications.

2.1 Experimental design

The experiment was conducted between 2018.07.23 and 2018.08.08 at the Psychology Department of the Kiel University, Germany. After subjects were instructed by the experimenter, they received an info sheet with general information, the informed consent, a worksheet to generate a personal code, and a survey that included questions on socio-economic characteristics such as age, gender, education and income level.

At the beginning of the experiment, subjects were seated in front of a computer screen and different sensor were attached to their bodies: First, subjects were asked to place their heads in the corresponding headrest to adjust the eye-tracking sensors (Tobii Pro X2-30) and Software (Tobii Pro SDK, Tobii Pro Eye Tracker Manager) for recording gaze-focus and pupils' sizes. Sampling rate was 50 Hz. After that, sensors were placed. We assessed physiological data with a 16-channel bioamplifier (Nexus-16; Mind Media B.V., the Netherlands). At first, a breast strap was attached to the thorax to assess breathing movement. Skin conductance was measured with adhesive non-reusable electrodes. They were attached to the distal phalange of the index and middle fingers of the non-dominant hand. Blood volume pulse was measured with a sensor on the annular finger.

A thermistor was taped to the auricular finger. Blood volume pulse was sampled with 128 Hz. All remaining data was recorded with 32 Hz. The recording software used to record these physiological reactions was Bio-trace (Mind Media B.V.).

The lottery gambling experiment started after we assured a successful calibration of all sensors. The first screen was a welcome page and, again, subjects were presented the concise instructions for the lottery gambling experiment. Then, subjects were asked to start the experiment by pressing the space-bar and deciding on three test trials. After subjects were asked if they had any remaining questions, they were shielded from acoustic disturbances with ear protectors and the actual experiment started. After deciding on 200 lottery trials, the experiment was finished, sensors were removed, and subjects received their final payoffs.

2.2 Descriptive statistics

Table A1 shows the absolute and relative number of played lotteries by subject and lotteries' expected values that correspond to the individual choice patterns that we present in Figure 2 (section 2) in the main text.

Table A1 Share of played lotteries by subject and lotteries' expected values

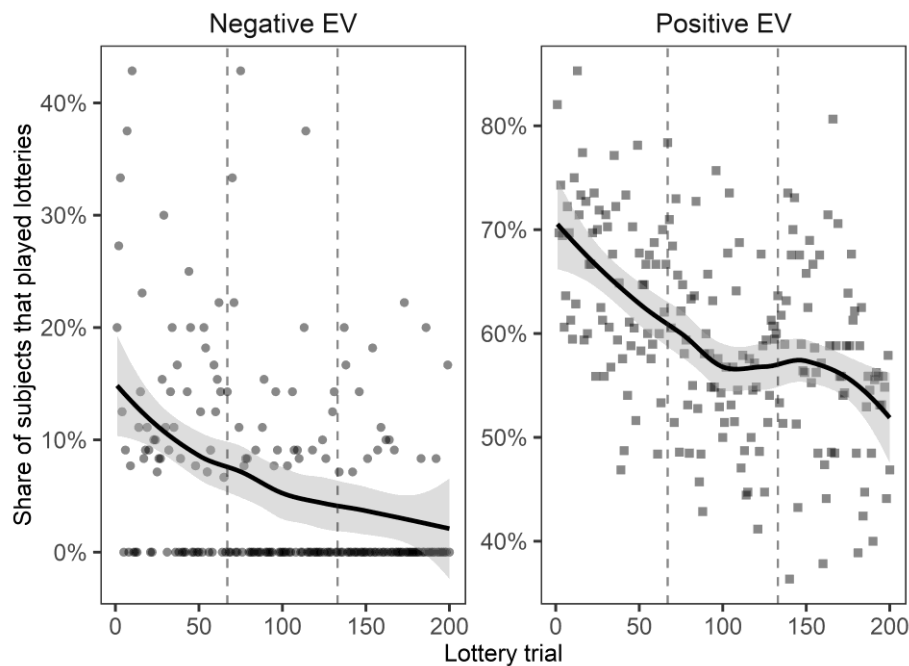
	All		Positive EV		Negative EV	
Subject ID	Played	Played (%)	Played	Played (%)	Played	Played (%)
nn07in21	162	81.00%	155	100.00%	7	15.56%
el05nd02	156	78.00%	137	88.39%	19	42.22%
er06er26	145	72.50%	145	93.55%	0	00.00%
er10in30	135	67.50%	130	83.87%	5	11.11%
hn06af11	132	66.00%	132	85.16%	0	0.00%
er05ed04	126	63.00%	125	80.65%	1	2.22%
er05ch04	125	62.50%	116	74.84%	9	20.00%
lz08rd25	123	61.50%	115	74.19%	8	17.78%
en05as01	121	60.50%	121	78.06%	0	0.00%
ea08nn28	108	54.00%	103	66.45%	5	11.11%
se06er29	107	53.50%	106	68.39%	1	2.22%
pp06ai18	105	52.50%	105	67.74%	0	0.00%
ergaas21	103	51.50%	102	65.81%	1	2.22%
nn06ld01	103	51.50%	90	58.06%	13	28.89%
ob04er02	103	51.50%	101	65.16%	2	4.44%
ch07nk26	102	51.00%	102	65.81%	1	2.22%
eechen11	98	49.00%	83	53.55%	15	33.33%
nn05er10	98	49.00%	98	63.23%	0	0.00%
nn08we28	98	49.00%	97	62.58%	1	2.22%
nn06rg19	95	47.50%	95	61.29%	0	0.00%
ce08le09	94	47.00%	75	48.39%	19	42.22%
mz04ed19	93	46.50%	93	60.00%	0	0.00%
os07ch18	93	46.50%	91	58.71%	2	4.44%
le06id19	92	46.00%	91	58.71%	1	2.22%
ix06er30	90	45.00%	87	56.13%	3	6.67%
usnaas21	88	44.00%	88	56.77%	0	0.00%
en05er24	85	42.50%	83	53.55%	2	4.44%
er09as11	85	42.50%	85	54.84%	0	0.00%
ermibe03	85	42.50%	85	54.84%	0	0.00%
le07ed18	85	42.50%	85	54.84%	0	0.00%
on04lf09	85	42.50%	85	54.84%	0	0.00%
er07nd25	84	42.00%	80	51.61%	4	8.89%
chsade20	83	41.50%	83	53.55%	0	0.00%
in07en17	80	40.00%	79	50.97%	1	2.22%
in07rk14	73	36.50%	73	47.10%	0	0.00%
er03ke09	71	35.50%	71	45.81%	0	0.00%
ersute14	71	35.50%	71	45.81%	0	0.00%
an07co25	67	33.50%	67	43.23%	0	0.00%
ke05rd12	65	32.50%	65	41.94%	0	0.00%
ld05us24	61	30.50%	60	38.71%	1	2.22%
er06er06	58	29.00%	58	37.42%	0	0.00%
ke08el19	56	28.00%	56	36.13%	0	0.00%
uz07im20	52	26.00%	51	32.90%	1	2.22%
ck07nz06	50	25.00%	49	31.61%	0	0.00%
Mean	95,25	47,62%	92,48	59.66%	2,77	6.16%
SD	26.41	-	24.78	-	4.98	-
N	4191	-	4069	-	122	-

Notes: Share of played lotteries by subject and lotteries' expected values (EV). Each of the 44 subjects decided on 200 lotteries: 155 with positive expected values (PEV) and 45 with negative expected values (NEV).

While Table A1 shows that there is high heterogeneity in the individual share of played lotteries across subjects, there are appear to exist further systematic differences in the relative share of subjects that play NEVL and PEVL with respect to lottery trials.

Figure A1 shows the relative share of subjects that played lotteries for each of the 155 PEVL and 45 NEVL by lottery trial. We observe PEVL and NEVL decisions at each trial. However, with respect to the absolute number of subjects that played lotteries, the minimum number of subjects that were offered PEVL (NEVL) per trial is 26 (4) and the maximum is 40 (18). To clarify, the number of subjects that were offered NEVL and PEVL varies with lottery trial. Precisely, the mean share of subjects that were offered PEVL is approximately three times as high as it is for NEVL, since the 200 lotteries include 45 NEVL and 155 PEVL.

Figure A1 Relative share of subjects that played lotteries by lottery trial



Notes: Relative share of subjects that played lotteries by lottery trial and lotteries' expected values (EV). Solid lines correspond to locally-weighted regression smoothing (loess) curves. 95% confidence intervals are indicated by grey-shaded areas. Each of the 44 subjects decided on playing 200 lotteries: 155 with positive EV and 45 with negative EV. Dashed lines indicate the 67th and 133th lottery-trial.

Figure A1 suggests that subjects' propensity for risking to play lotteries decreases as the experiment progresses; though, we observe differences for NEVL and PEVL. The minimum share of subjects that play NEVL is 0% (0 out of 11), and the maximum is 43% (3 out of 7). For subjects that play PEVL, we observe a minimum share of 36% (12 out of 31) and a maximum of 85% (29 out of 34).

2.3 Data cleaning and variables' specification

We exclude all 224 observations with missing eye-tracking data and we exclude 220 observations due to including information on past gambling behavior in the *LSG* and *LSGPA* data. There are 18 records that are associated with missing eye-tracking data and are related to one of the first 5 lottery gambling decisions. Hence, the cleaned dataset includes 8368 records.

In Table A2 we summarize the distribution of gambling choices by lotteries' expected values and data sample: the full data, the cleaned data, the training data, and the test data.

Table A2 Distribution of risky choices by data sample and lotteries' expected values

Data: Full sample	Expected value					
Choice	# All	All (%)	# PEV	PEV (%)	# NEV	NEV (%)
Played	4191	48	4069	60	122	6
Not played	4609	52	2751	40	1858	94
N	8800	100	6820	100	1980	100

Data: Cleaned	Expected value					
Choice	# All	All (%)	# PEV	PEV (%)	# NEV	NEV (%)
Played	4109	49	3235	40	629	23
Not played	4259	51	3850	60	2153	77
N	8368	100	7994	100	2782	100

Data: Training	Expected value					
Choice	# All	All (%)	# PEV	PEV (%)	# NEV	NEV (%)
Played	3217	48	3217	48	93	6
Not played	3494	52	3494	52	1406	94
N	6711	100	6711	100	1499	100

Data: Test	Expected value					
Choice	# All	All (%)	# PEV	PEV (%)	# NEV	NEV (%)
Played	892	54	536	42	18	5
Not played	765	46	747	58	356	95
N	1657	100	1283	100	374	100

Notes: The experiment data covers 44 subjects that each were offered 200 lotteries with 50/50 outcome probabilities, i.e., 8800 observations in total for the full data sample.

In Table A3 we present the variables' empirical specifications by predictor-set and groups together with distributional summary statistics. While most variables' description is self-explanatory, in the following we discuss further details with respect to CPD and their encodings and corresponding implications.

Table A3 Predictor variable specifications and summary statistics

Variable	Predictor-set	Variable group	Description	Min	Max	Mean	SD
Played	Outcome	Economic decision	Play vs. not-play	0	1	0.48	0.50
win_value	LSG, LSGPA	Win and loss values (L)	Potential win (EUR)	1.00	20.00	10.50	5.76
loss_value	LSG, LSGPA	Win and loss values (L)	Potential loss (EUR)	-10.00	-1.00	-5.50	2.87
exp_value	LSG, LSGPA	Win and loss values (L)	Expected value (EUR)	-4.50	9.50	2.50	3.21
neg_exp_value	LSG, LSGPA	Win and loss values (L)	Expected value < 0	0	1	0.22	0.42
trial	LSG, LSGPA	Trial (L)	Lottery trial	6	200	94.84	57.17
trial_D1_1_67	LSG, LSGPA	Trial (L)	Lottery trial 1 to 67	0	1	0.32	0.47
trial_D2_68_133*	LSG, LSGPA	Trial (L)	Lottery trial 68 to 133	0	1	0.34	0.47
trial_D2_134_200	LSG, LSGPA	Trial (L)	Lottery trial 134 to 200	0	1	0.34	0.47
win_right ^{LSG LSGPA}	LSG, LSGPA	Left vs. right (L)	Potential win is displayed on the right box	0	1	0.50	0.50
accept_right	LSG, LSGPA	Left vs. right (L)	Accept by pressing the right arrow	0	1	0.50	0.50
personD_1	P, A, LSG, LSGPA	Subject ID (S)	Dummy variable for individual subject 1	0	1	0.02	0.15
personD_2	P, A, LSG, LSGPA	Subject ID (S)	Dummy variable for individual subject 2	0	1	0.02	0.15
personD_3	P, A, LSG, LSGPA	Subject ID (S)	Dummy variable for individual subject 3	0	1	0.02	0.15
personD_4	P, A, LSG, LSGPA	Subject ID (S)	Dummy variable for individual subject 4	0	1	0.02	0.15
personD_5	P, A, LSG, LSGPA	Subject ID (S)	Dummy variable for individual subject 5	0	1	0.02	0.14
personD_6	P, A, LSG, LSGPA	Subject ID (S)	Dummy variable for individual subject 6	0	1	0.02	0.15
personD_7	P, A, LSG, LSGPA	Subject ID (S)	Dummy variable for individual subject 7	0	1	0.02	0.15
personD_8	P, A, LSG, LSGPA	Subject ID (S)	Dummy variable for individual subject 8	0	1	0.02	0.15
personD_9	P, A, LSG, LSGPA	Subject ID (S)	Dummy variable for individual subject 9	0	1	0.02	0.15
personD_10	P, A, LSG, LSGPA	Subject ID (S)	Dummy variable for individual subject 10	0	1	0.02	0.15
personD_11*	P, A, LSG, LSGPA	Subject ID (S)	Dummy variable for individual subject 11	0	1	0.02	0.15
personD_12	P, A, LSG, LSGPA	Subject ID (S)	Dummy variable for individual subject 12	0	1	0.02	0.14
personD_13	P, A, LSG, LSGPA	Subject ID (S)	Dummy variable for individual subject 13	0	1	0.02	0.15
personD_14	P, A, LSG, LSGPA	Subject ID (S)	Dummy variable for individual subject 14	0	1	0.02	0.15
personD_15	P, A, LSG, LSGPA	Subject ID (S)	Dummy variable for individual subject 15	0	1	0.02	0.15
personD_16	P, A, LSG, LSGPA	Subject ID (S)	Dummy variable for individual subject 16	0	1	0.02	0.15

Variable	Predictor-set	Variable group	Description	Min	Max	Mean	SD
personD_17	P, A, LSG, LSGPA	Subject ID (S)	Dummy variable for individual subject 17	0	1	0.02	0.15
personD_18	P, A, LSG, LSGPA	Subject ID (S)	Dummy variable for individual subject 18	0	1	0.02	0.15
personD_19	P, A, LSG, LSGPA	Subject ID (S)	Dummy variable for individual subject 19	0	1	0.02	0.15
personD_20	P, A, LSG, LSGPA	Subject ID (S)	Dummy variable for individual subject 20	0	1	0.02	0.15
personD_21	P, A, LSG, LSGPA	Subject ID (S)	Dummy variable for individual subject 21	0	1	0.02	0.15
personD_22	P, A, LSG, LSGPA	Subject ID (S)	Dummy variable for individual subject 22	0	1	0.02	0.15
personD_23	P, A, LSG, LSGPA	Subject ID (S)	Dummy variable for individual subject 23	0	1	0.02	0.15
personD_24	P, A, LSG, LSGPA	Subject ID (S)	Dummy variable for individual subject 24	0	1	0.02	0.15
personD_25	P, A, LSG, LSGPA	Subject ID (S)	Dummy variable for individual subject 25	0	1	0.02	0.15
personD_26	P, A, LSG, LSGPA	Subject ID (S)	Dummy variable for individual subject 26	0	1	0.02	0.13
personD_27	P, A, LSG, LSGPA	Subject ID (S)	Dummy variable for individual subject 27	0	1	0.02	0.15
personD_28	P, A, LSG, LSGPA	Subject ID (S)	Dummy variable for individual subject 28	0	1	0.02	0.15
personD_29	P, A, LSG, LSGPA	Subject ID (S)	Dummy variable for individual subject 29	0	1	0.02	0.15
personD_30	P, A, LSG, LSGPA	Subject ID (S)	Dummy variable for individual subject 30	0	1	0.02	0.15
personD_31	P, A, LSG, LSGPA	Subject ID (S)	Dummy variable for individual subject 31	0	1	0.02	0.15
personD_32	P, A, LSG, LSGPA	Subject ID (S)	Dummy variable for individual subject 32	0	1	0.02	0.15
personD_33	P, A, LSG, LSGPA	Subject ID (S)	Dummy variable for individual subject 33	0	1	0.02	0.15
personD_34	P, A, LSG, LSGPA	Subject ID (S)	Dummy variable for individual subject 34	0	1	0.02	0.14
personD_35	P, A, LSG, LSGPA	Subject ID (S)	Dummy variable for individual subject 35	0	1	0.02	0.15
personD_36	P, A, LSG, LSGPA	Subject ID (S)	Dummy variable for individual subject 36	0	1	0.02	0.15
personD_37	P, A, LSG, LSGPA	Subject ID (S)	Dummy variable for individual subject 37	0	1	0.02	0.15
personD_38	P, A, LSG, LSGPA	Subject ID (S)	Dummy variable for individual subject 38	0	1	0.02	0.15
personD_39	P, A, LSG, LSGPA	Subject ID (S)	Dummy variable for individual subject 39	0	1	0.02	0.15
personD_40	P, A, LSG, LSGPA	Subject ID (S)	Dummy variable for individual subject 40	0	1	0.02	0.15
personD_41 ^{LSG LSGPA}	P, A, LSG, LSGPA	Subject ID (S)	Dummy variable for individual subject 41	0	1	0.02	0.15
personD_42	P, A, LSG, LSGPA	Subject ID (S)	Dummy variable for individual subject 42	0	1	0.02	0.15
personD_43 ^{LSG LSGPA}	P, A, LSG, LSGPA	Subject ID (S)	Dummy variable for individual subject 42	0	1	0.02	0.15
personD_44 ^{LSG LSGPA}	P, A, LSG, LSGPA	Subject ID (S)	Dummy variable for individual subject 44	0	1	0.02	0.15

Variable	Predictor-set	Variable group	Description	Min	Max	Mean	SD
quali_D1_abi*		Socioeconomic (S)	Highest education: A-levels	0	1	0.43	0.50
quali_D2_bach	LSG, LSGPA	Socioeconomic (S)	Highest education: Bachelor	0	1	0.25	0.43
quali_D3_real ^{LSG LSGPA}	LSG, LSGPA	Socioeconomic (S)	Highest education: GCSE	0	1	0.11	0.32
quali_D4_master ^{LSG LSGPA}	LSG, LSGPA	Socioeconomic (S)	Highest education: Master or similar degree	0	1	0.16	0.37
quali_D5_fachabi ^{LSG LSGPA}	LSG, LSGPA	Socioeconomic (S)	Highest education: "Fachabitur"	0	1	0.04	0.20
educ_D1_psy*	LSG, LSGPA	Socioeconomic (S)	Educational background: Psychology	0	1	0.16	0.36
educ_D2_eco ^{LSG LSGPA}	LSG, LSGPA	Socioeconomic (S)	Educational background: Economics	0	1	0.16	0.36
educ_D3_na ^{LSG LSGPA}	LSG, LSGPA	Socioeconomic (S)	Educational background: NA	0	1	0.09	0.29
educ_D4_other ^{LSG LSGPA}	LSG, LSGPA	Socioeconomic (S)	Educational background: Other	0	1	0.59	0.49
income_D1	LSG, LSGPA	Socioeconomic (S)	Income level ≤ 800 EUR	0	1	0.36	0.48
income_D2*	LSG, LSGPA	Socioeconomic (S)	Income level > 800 EUR & < 1200 EUR	0	1	0.41	0.49
income_D3	LSG, LSGPA	Socioeconomic (S)	Income level ≥ 1200 EUR	0	1	0.23	0.42
female_D1 ^{LSG}	LSG, LSGPA	Socioeconomic (S)	Gender: Male vs. female	0	1	0.45	0.50
age_D1_19_25 ^{LSG LSGPA}	LSG, LSGPA	Socioeconomic (S)	Age group: 19 to 25 years	0	1	0.32	0.46
age_D2_26_32*	LSG, LSGPA	Socioeconomic (S)	Age group: 26 to 32 years	0	1	0.57	0.49
age_D3_33_39 ^{LSG LSGPA}	LSG, LSGPA	Socioeconomic (S)	Age group: 19 to 39 years	0	1	0.11	0.32
played_lag_1_2	LSG, LSGPA	Gambling behavior (G)	lagged 1 x 2 played	0	1	0.25	0.43
played_lag_1_2_3	LSG, LSGPA	Gambling behavior (G)	lagged 1 x 2 x 3 played	0	1	0.14	0.35
played_lag_1_2_3_4	LSG, LSGPA	Gambling behavior (G)	lagged 1 x 2 x 3 x 4 played	0	1	0.09	0.28
played_lag_1_2_3_4_5	LSG, LSGPA	Gambling behavior (G)	lagged 1 x 2 x 3 x 4 x 5 played	0	1	0.06	0.23
lag1_pos_outcome	LSG, LSGPA	Gambling behavior (G)	Lag 1 positive outcome: won money	0	1	0.23	0.42
lag2_pos_outcome	LSG, LSGPA	Gambling behavior (G)	Lag 2 positive outcome: won money	0	1	0.23	0.42
lag3_pos_outcome	LSG, LSGPA	Gambling behavior (G)	Lag 3 positive outcome: won money	0	1	0.23	0.42
lag4_pos_outcome	LSG, LSGPA	Gambling behavior (G)	Lag 4 positive outcome: won money	0	1	0.23	0.42
lag5_pos_outcome	LSG, LSGPA	Gambling behavior (G)	Lag 5 positive outcome: won money	0	1	0.23	0.42
lag1_neg_outcome	LSG, LSGPA	Gambling behavior (G)	Lag 1 negative outcome: lost money	0	1	0.24	0.43
lag2_neg_outcome	LSG, LSGPA	Gambling behavior (G)	Lag 2 negative outcome: lost money	0	1	0.24	0.43
lag3_neg_outcome	LSG, LSGPA	Gambling behavior (G)	Lag 3 negative outcome: lost money	0	1	0.24	0.43

Variable	Predictor-set	Variable group	Description	Min	Max	Mean	SD
lag4_neg_outcome	LSG, LSGPA	Gambling behavior (G)	Lag 4 negative outcome: lost money	0	1	0.24	0.43
lag5_neg_outcome	LSG, LSGPA	Gambling behavior (G)	Lag 5 negative outcome: lost money	0	1	0.24	0.43
BVP_max	P, LSGPA	Blood volume pulse (P)	Blood volume pulse: maximum (relative (%) changes)	-13.90	221.43	76.60	38.36
BVP_min ^{P LSGPA}	P, LSGPA	Blood volume pulse (P)	Blood volume pulse: minimum (relative (%) changes)	-153.23	-14.61	-49.82	21.08
BVP_mean	P, LSGPA	Blood volume pulse (P)	Blood volume pulse: mean (relative (%) changes)	-46.89	1.14	-16.67	9.46
BVP_delta	P, LSGPA	Blood volume pulse (P)	Difference BVP min and max (relative (%) changes)	19.03	347.05	126.43	55.26
BVP_Amp_max	P, LSGPA	Blood volume pulse (P)	Blood volume pulse amplitude: maximum (millivolts)	14.33	247.10	104.00	43.81
BVP_Amp_min ^{P LSGPA}	P, LSGPA	Blood volume pulse (P)	Blood volume pulse amplitude: minimum (millivolts)	11.68	210.87	83.75	37.06
BVP_Amp_mean	P, LSGPA	Blood volume pulse (P)	Blood volume pulse amplitude: mean (millivolts)	13.36	229.88	93.98	40.53
BVP_Amp_delta	P, LSGPA	Blood volume pulse (P)	Difference BVP amplitude max and min (millivolts)	1.23	85.29	20.25	10.05
RSP_max	P, LSGPA	Respiration rate (P)	Respiration: mean (relative (%) changes)	910.30	1237.00	1066.00	69.88
RSP_min ^{P LSGPA}	P, LSGPA	Respiration rate (P)	Respiration: minimum (relative (%) changes)	900.70	1223.60	1051.60	68.73
RSP_mean	P, LSGPA	Respiration rate (P)	Respiration: maximum (relative (%) changes)	904.90	1230.70	1057.50	69.12
RSP_delta	P, LSGPA	Respiration rate (P)	Difference: RSP max and min (relative (%) changes)	0.80	124.09	14.42	11.09
RSP_rate_max	P, LSGPA	Respiration rate (P)	Respiration rate: maximum (breathes per min)	4.46	60.00	20.47	7.63
RSP_rate_min ^{P LSGPA}	P, LSGPA	Respiration rate (P)	Respiration rate: minimum (breathes per min)	4.28	58.18	16.11	4.45
RSP_rate_mean	P, LSGPA	Respiration rate (P)	Respiration rate: mean (breathes per min)	4.46	58.18	18.30	5.59
RSP_rate_delta	P, LSGPA	Respiration rate (P)	Difference: RSP rate max and min (breathes per min)	0.00	48.47	4.36	6.64
HR_max	P, LSGPA	Heart rate (P)	Heart rate: maximum (beats per min)	48.01	128.00	83.53	12.43
HR_min ^{P LSGPA}	P, LSGPA	Heart rate (P)	Heart rate: minimum (beats per min)	42.67	123.87	76.71	11.80
HR_mean	P, LSGPA	Heart rate (P)	Heart rate: mean (beats per min)	46.66	125.91	80.15	12.08
HR_delta	P, LSGPA	Heart rate (P)	Difference: Heart rate max and min (beats per min)	0.01	39.10	6.82	4.05
CDA_nSCR_D0	P, LSGPA	Skin conductance (P) ^{SCR}	0 significant skin conductance rates (SCRs)	0	1	0.23	0.42
CDA_nSCR_D1 [†]	P, LSGPA	Skin conductance (P) ^{SCR}	1 significant SCR	0	1	0.47	0.50
CDA_nSCR_D2	P, LSGPA	Skin conductance (P) ^{SCR}	2 significant SCR	0	1	0.20	0.40
CDA_nSCR_D3	P, LSGPA	Skin conductance (P) ^{SCR}	3 significant SCR	0	1	0.06	0.24
CDA_nSCR_D4	P, LSGPA	Skin conductance (P) ^{SCR}	4+ significant SCR	0	1	0.03	0.17
CDA.AmpSum	P, LSGPA	Skin conductance (P)	Sum of SCR-amplitudes of significant SCR (μS)	0.00	5.49	0.17	0.30

Variable	Predictor-set	Variable group	Description	Min	Max	Mean	SD
CDA.SCR	P, LSGPA	Skin conductance (P)	Average phasic driver (μS)	0.00	0.10	0.01	0.01
CDA.ISCR	P, LSGPA	Skin conductance (P)	Area of phasic driver (time integral) ($\mu S \times S$)	0	10.00	0.59	0.88
CDA.PhasicMax	P, LSGPA	Skin conductance (P)	Maximum value of phasic activity*	0	9.52	0.58	0.82
CDA.Tonic	P, LSGPA	Skin conductance (P)	Mean tonic activity*	1.34	33.78	8.11	5.57
Temp_max	P, LSGPA	Body temperature (P)	Temperature: maximum ($^{\circ}C$)	32.02	36.78	35.81	0.65
Temp_min ^{P LSGPA}	P, LSGPA	Body temperature (P)	Temperature: minimum ($^{\circ}C$)	32.02	36.77	35.80	0.65
Temp_mean	P, LSGPA	Body temperature (P)	Temperature: mean ($^{\circ}C$)	32.02	36.77	35.81	0.65
Temp_delta	P, LSGPA	Body temperature (P)	Difference: Temperature max and min ($^{\circ}C$)	0.00	0.14	0.01	0.01
t1_pupil_avg_lr_min	P, LSGPA	Pupil size (P)	Average pupil size: minimum (mm)	1.32	3.97	2.47	0.35
t1_pupil_avg_lr_max ^{P LSGPA}	P, LSGPA	Pupil size (P)	Average pupil size: maximum (mm)	2.11	6.43	3.05	0.35
t1_pupil_avg_lr_mean	P, LSGPA	Pupil size (P)	Average pupil size: mean (mm)	1.94	4.11	2.77	0.33
t1_delta_pupil_avg_lr	P, LSGPA	Pupil size (P)	Average pupil size: difference max min (mm)	0.04	4.40	0.58	0.26
t1_time_none	A, LSGPA	Gaze (A)	Time not spent on fixating boxes (sec)	0.03	1.00	3.00	0.64
t1_time_win	A, LSGPA	Gaze (A)	Time spent on fixating win (sec)	0.00	2.82	1.07	0.58
t1_time_loss	A, LSGPA	Gaze (A)	Time spent on fixating loss (sec)	0.00	2.73	0.94	0.55

Notes: We specified 126 predictor variables. Factor variables are encoded as dummy variables and we exclude the most frequently observed level for each category as corresponding reference groups. This encoding results in a total number of 119 predictors (42 numeric and 77 binary variables). We identify issues concerning multi-collinearity by QR decomposition for each predictor-set individually on the basis of the cleaned data. The *A* data does not include multi-collinear predictors. We indicate the removed variables for the predictor-sets *P*, *LSGP*, and *LSGPA* with ^P, ^{LSG}, and ^{LSGPA}, respectively. ^{SCR} Refers to the number of significant (above threshold) skin conductance rate measures. * Tonic and phasic activity based on the continuous decomposition analysis as suggested by (Benedek & Kaernbach, 2010).

3 Detailed model performance evaluation

3.1 Implementation

We use R (R Core Team, 2018) in combination with RStudio (RStudio Team) for the main computations and graphics in this paper. For the generalized linear methods employed in this study we use the R-package *glmnet* (Friedman et al., 2010), for support vector machines we use *kernlab* (Karatzoglou et al., 2004), for neural network models we use *nnet* (Venables & Ripley, 2002), for random forests we use *ranger* (Wright & Ziegler, 2017), and for stochastic gradient boosting machines we use *gbm* (Greenwell et al., 2018). We use *caret* for model training and model evaluation (Kuhn, 2018). Moreover, we use *tidyverse* (Wickham, 2016, 2017; Wickham et al., 2018), *lattice* (Sarkar, 2008), *circlize* (Gu et al., 2014), *RColorBrewer* (Neuwirth, 2014), and *complexHeatmap* (Gu et al., 2016) for data manipulations and creating the figures that are included this study. Last, we use *doParallel* (Calaway et al., 2017) for multicore processing where applicable.

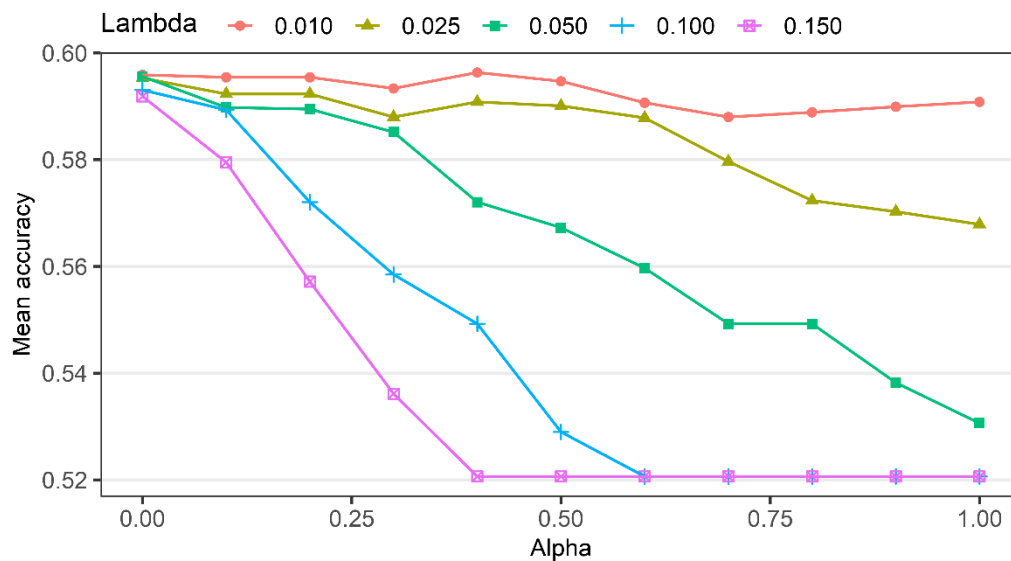
3.2 Model tuning

This section presents the results for the systematic-grid search that we employ to identify appropriate hyperparameter values for the machine learning algorithms that we use in our forecasting experiment.

3.2.1 Psychophysiological reactions (*P*)

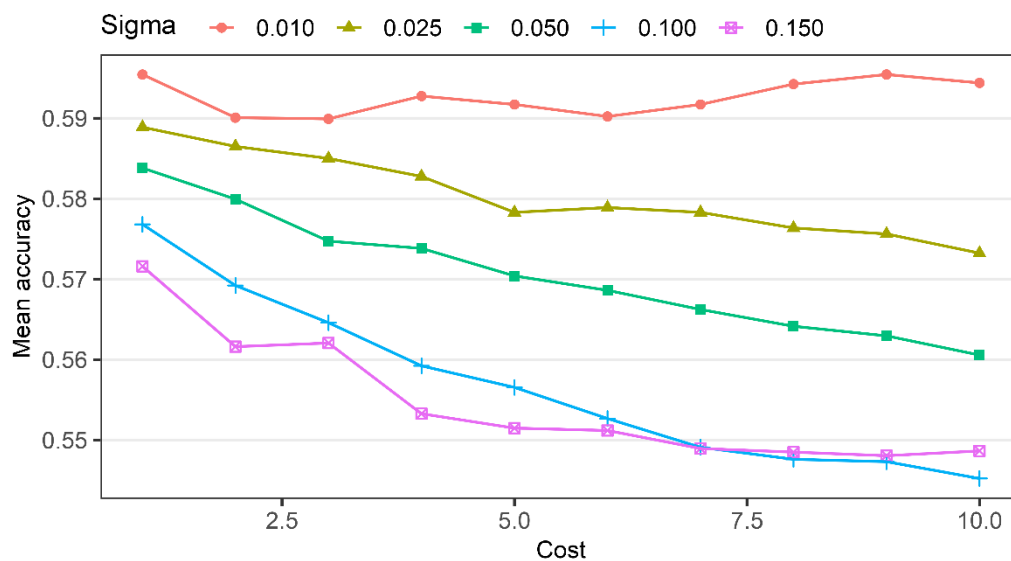
The following Figures (A2-A6) show the mean 10-fold CV results for determining sensitive parameter values via a systematic grid search on the basis of the predictor-set *P* that includes 73 predictor variables. Precisely, *P* includes information on individual psychophysiological reactions and dummy variables that account for subject specific effects.

Figure A2 Mean CV accuracy results for linear penalized regression models for P



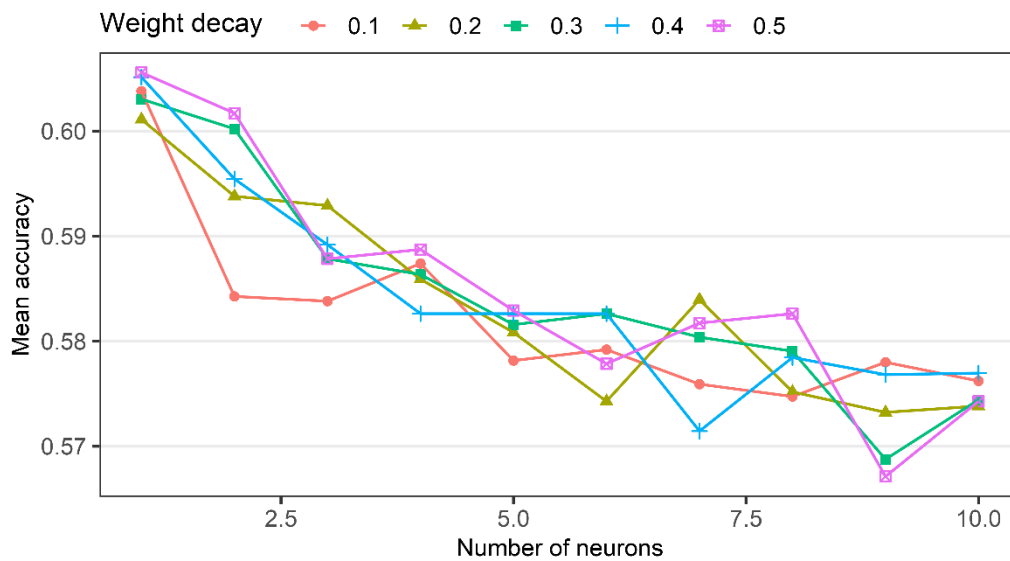
Notes: Mean 10-fold cross validated classification mean accuracy results for linear penalized regression models (lasso, ridge, and elastic net) as a function of the mixing percentage parameter $\alpha = \{0, 0.1, 0.2, \dots, 1\}$ and regularization parameter $\lambda = \{0.01, 0.025, 0.05, 0.1, 0.15\}$ on the basis of psychophysiological choice-process data (P).

Figure A3 Mean CV accuracy results for SVM for P



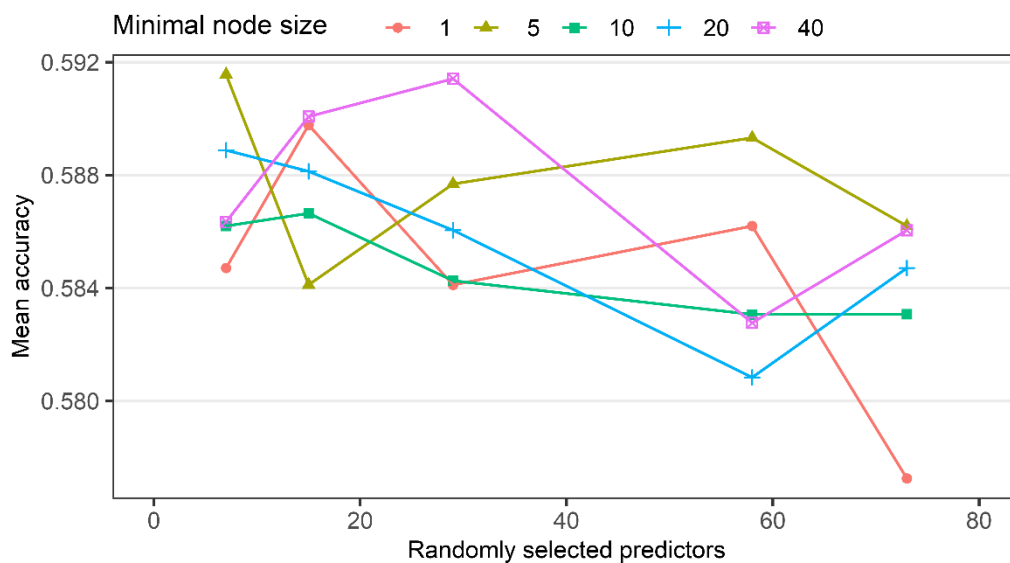
Notes: Mean 10-fold cross validated classification accuracy results for support vector machines (SVM) with a radial basis function kernel as a function of a cost parameter $C = \{1, 2, \dots, 10\}$ and inverse kernel width $\sigma = \{0.01, 0.025, 0.05, 0.1, 0.15\}$ on the basis of psychophysiological choice-process data (P).

Figure A4 Mean CV accuracy results for ANN for P



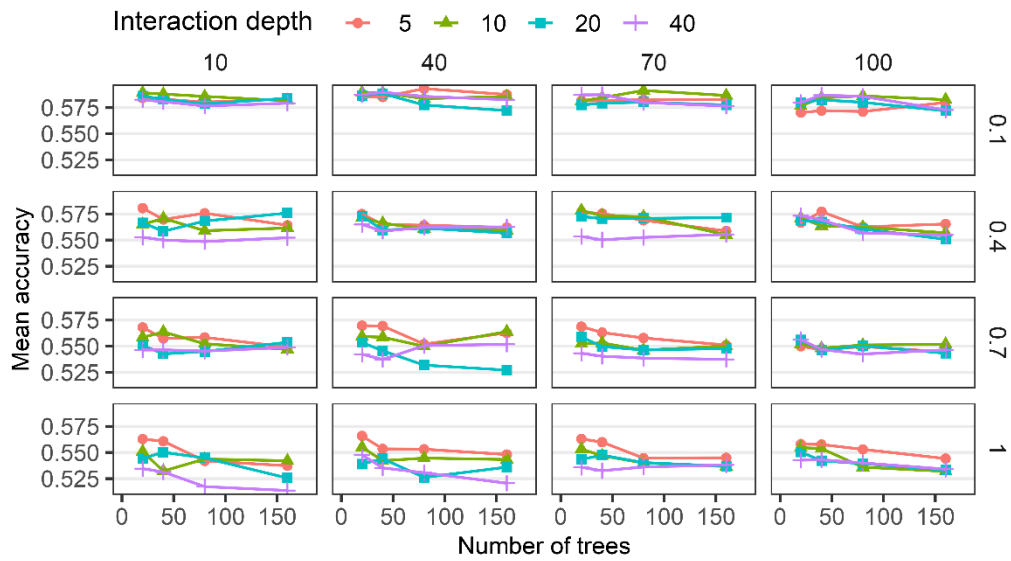
Notes: Mean 10-fold cross validated classification accuracy results for a single hidden layer artificial neural networks (ANN) using a sigmoid activation function, the number of neurons in the hidden layer $size = \{1, 2, \dots, 10\}$ and a regularization parameter $weight\ decay = \{0.1, 0.2, \dots, 0.5\}$ on the basis of psychophysiological choice-process data (P).

Figure A5 Mean CV accuracy results for RF for P



Notes: Mean 10-fold cross validated classification accuracy results for random forests (RF) using the Gini splitting rule, a number of 1500 individual bagged CART trees per forest, and as a function of the minimum number of samples in each leaf $min.\ node.\ size = \{1, 5, 10, 20, 40\}$ and the number of randomly selected predictors at each split $mtry = D_i(0.1, 0.2, 0.4, 0.8, 1)$ with D_i as the number of predictor variables (73) that are included in the psychophysiological choice-process data (P).

Figure A6 Mean CV accuracy results for GBM for P

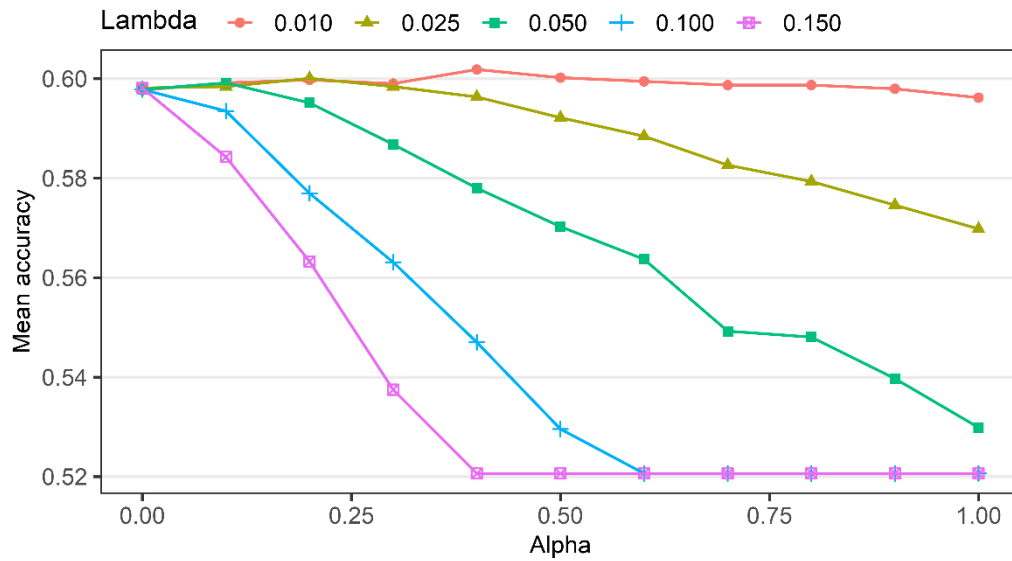


Notes: Mean 10-fold cross validated classification accuracy results for tree-based stochastic gradient boosting machines (GBM) as a function of the maximum tree depth $interaction.depth = \{5, 10, 20, 40\}$ and number of boosting iterations $n.trees = \{20, 40, 80, 160\}$, a shrinkage factor $shrinkage = \{0.1, 0.4, 0.7, 1\}$, and a minimum number of samples in each leaf $n.minobsinnode = \{10, 40, 70, 100\}$ on the basis of psychophysiological choice-process data (P).

3.2.2 Attention data (A)

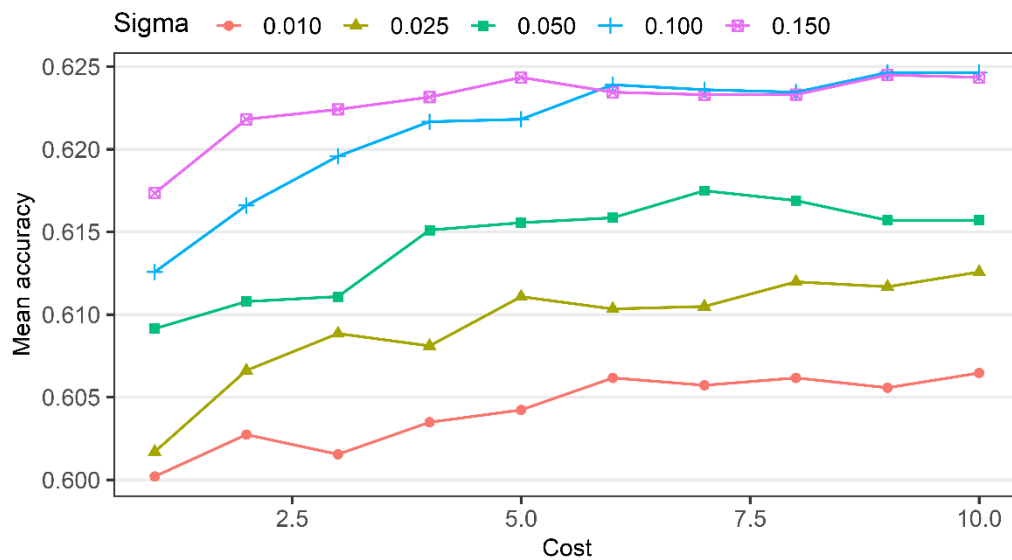
The following Figures (A7-A11) show the mean 10-fold cross-validation results for determining models' hyperparameters by a systematic grid search on the basis of a predictor-set A that includes 48 predictor variables. Precisely, A includes information on individual gaze-patterns and dummy variables that account for subject specific effects.

Figure A7 Mean CV accuracy results for linear penalized regression models for A



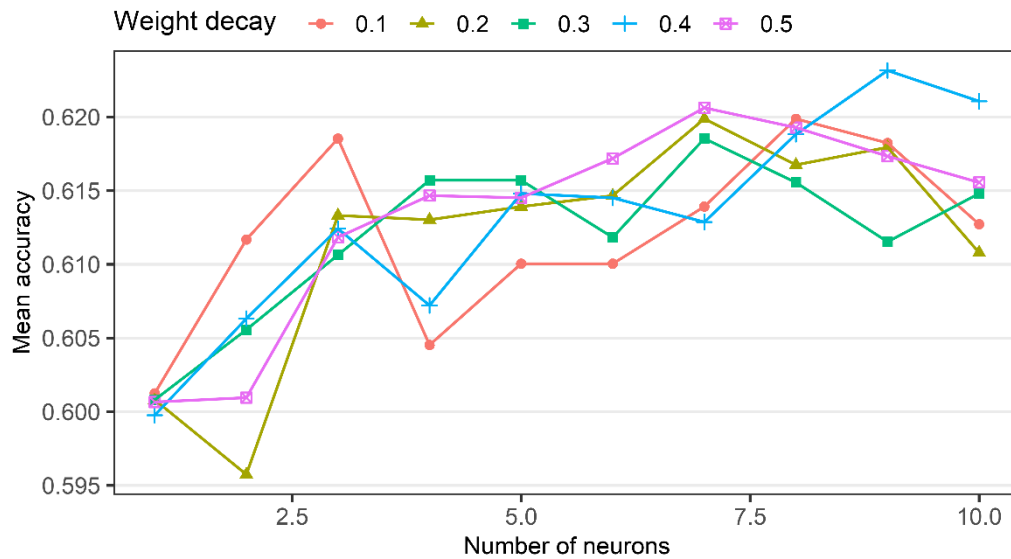
Notes: Mean 10-fold cross validated classification mean accuracy results for linear penalized regression models (lasso, ridge, and elastic net) as a function of the mixing percentage parameter $\alpha = \{0, 0.1, 0.2, \dots, 1\}$ and regularization parameter $\lambda = \{0.01, 0.025, 0.05, 0.1, 0.15\}$ on the basis of attention choice-process data (A).

Figure A8 Mean CV accuracy results for SVM for A



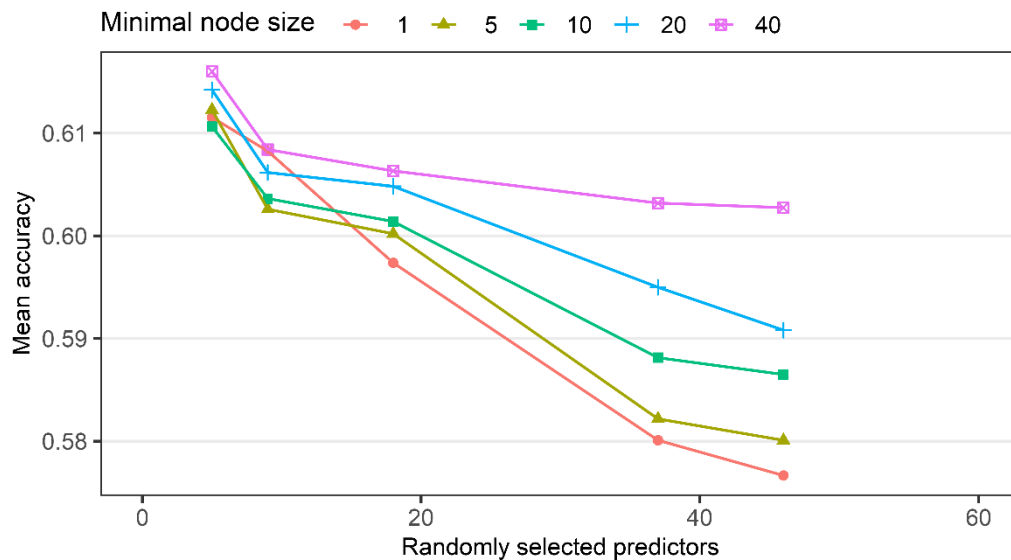
Notes: Mean 10-fold cross validated classification accuracy results for support vector machines (SVM) with a radial basis function kernel as a function of a cost parameter $C = \{1, 2, \dots, 10\}$ and inverse kernel width $\sigma = \{0.01, 0.025, 0.05, 0.1, 0.15\}$ on the basis of attention choice-process data (A).

Figure A9 Mean CV accuracy results for ANN for A



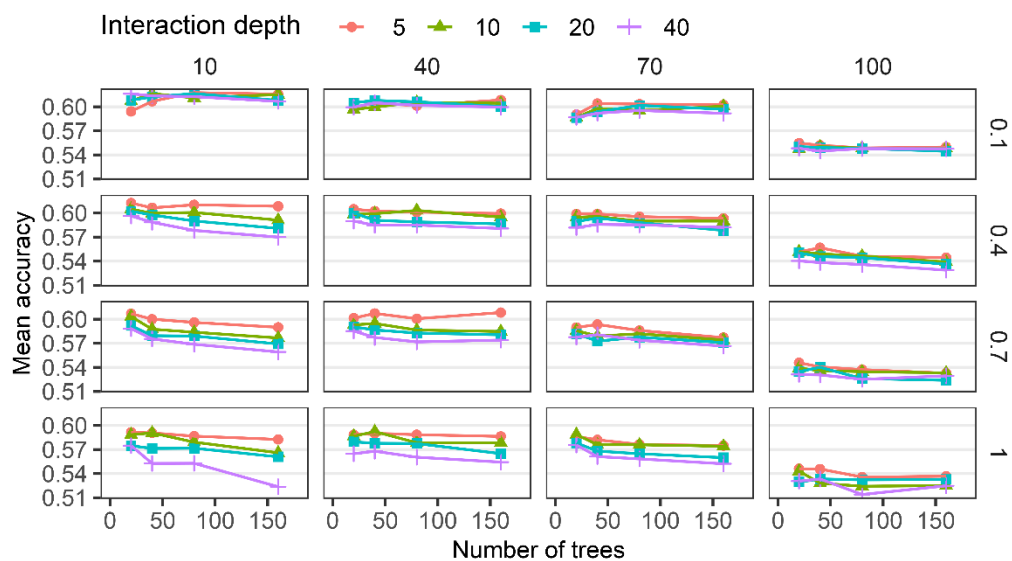
Notes: Mean 10-fold cross validated classification accuracy results for a single hidden layer artificial neural networks (ANN) using a sigmoid activation function, the number of neurons in the hidden layer $size = \{1, 2, \dots, 10\}$ and a regularization parameter $weight\ decay = \{0.1, 0.2, \dots, 0.5\}$ on the basis of attention choice-process data (A).

Figure A10 Mean CV accuracy results for RF for A



Notes: Mean 10-fold cross validated classification accuracy results for random forests (RF) using the Gini splitting rule, a number of 1500 individual bagged CART trees per forest, and as a function of the minimum number of samples in each leaf $min.\ node.\ size = \{1, 5, 10, 20, 40\}$ and the number of randomly selected predictors at each split $mtry = D_i(0.1, 0.2, 0.4, 0.8, 1)$ with D_i as the number of predictor variables (48) that are included in the attention data (A).

Figure A11 Mean CV accuracy results for GBM for A

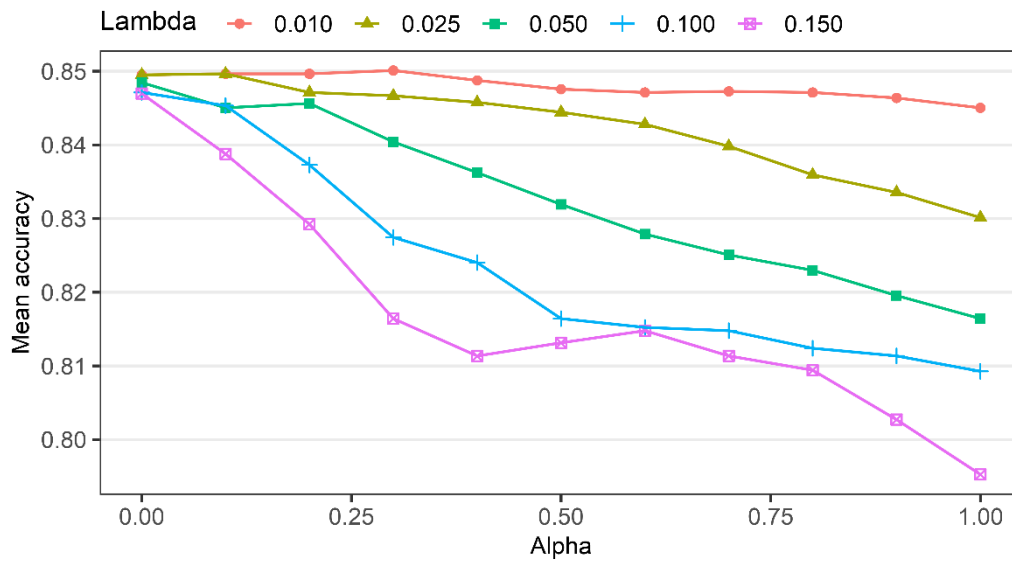


Notes: Mean 10-fold cross validated classification accuracy results for tree-based stochastic gradient boosting machines (GBM) as a function of the maximum tree depth $interaction.depth = \{5, 10, 20, 40\}$ and number of boosting iterations $n.trees = \{20, 40, 80, 160\}$, a shrinkage factor $shrinkage = \{0.1, 0.4, 0.7, 1\}$, and a minimum number of samples in each leaf $n.minobsinnode = \{10, 40, 70, 100\}$ on the basis of attention choice-process data (A).

3.2.3 Lottery design, socioeconomic and past gambling characteristics (LSG)

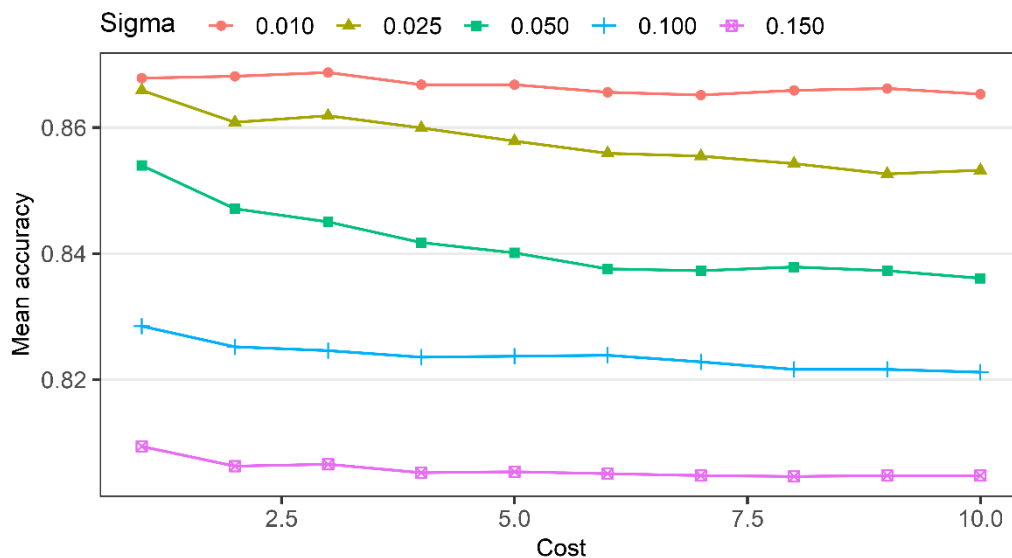
The following Figures (A12-A16) show the mean 10-fold cross-validation results for determining models' hyperparameters by a systematic grid search on the basis of a predictor-set *LSG* that includes 64 predictor variables. Precisely, *LSG* includes information on lottery design variables, socioeconomic characteristics, and past gambling behavior (*LSG*).

Figure A12 Mean CV accuracy results for linear penalized regression models for *LSG*



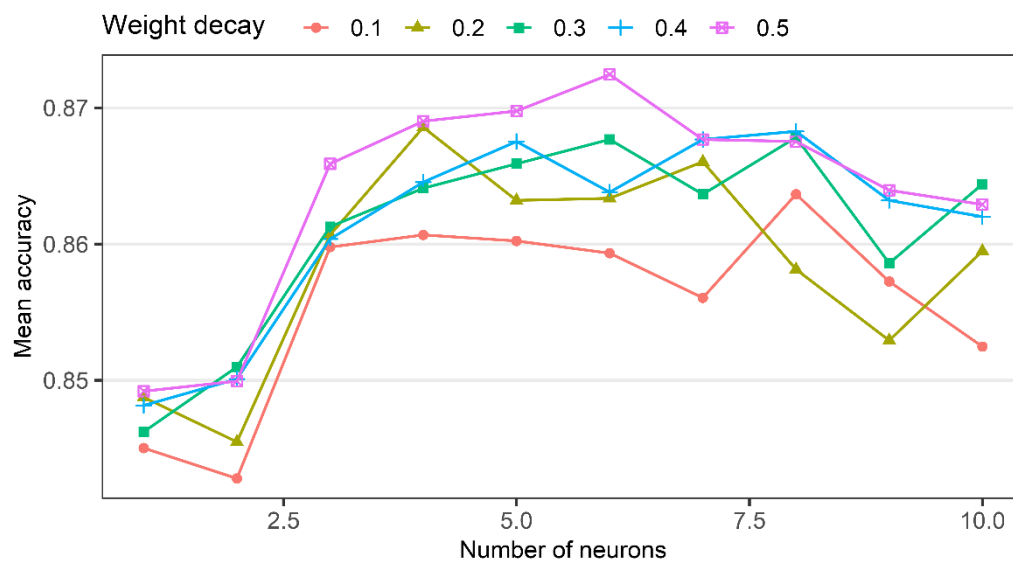
Notes: Mean 10-fold cross validated classification mean accuracy results for linear penalized regression models (lasso, ridge, and elastic net) as a function of the mixing percentage parameter $\alpha = \{0, 0.1, 0.2, \dots, 1\}$ and regularization parameter $\lambda = \{0.01, 0.025, 0.05, 0.1, 0.15\}$ on the basis of information that is related to lottery design variables, socioeconomic characteristics, and past gambling behavior (*LSG*).

Figure A13 Mean CV accuracy results for SVM for *LSG*



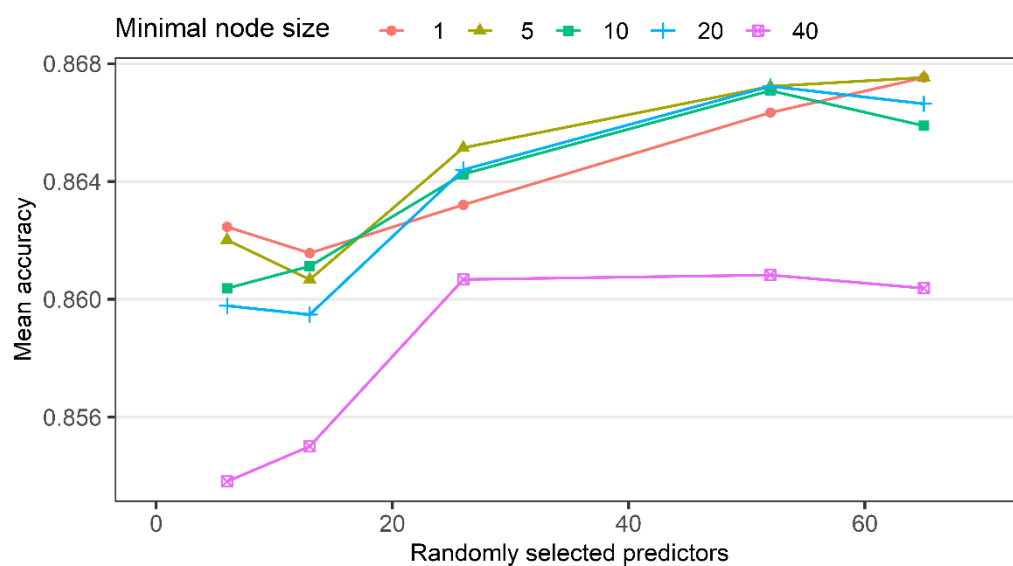
Notes: Mean 10-fold cross validated classification accuracy results for support vector machines (SVM) with a radial basis function kernel as a function of a cost parameter $C = \{1, 2, \dots, 10\}$ and inverse kernel width $\sigma = \{0.01, 0.025, 0.05, 0.1, 0.15\}$ on the basis of information that is related to lottery design variables, socioeconomic characteristics, and past gambling behavior (*LSG*).

Figure A14 Mean CV accuracy results for ANN for *LSG*



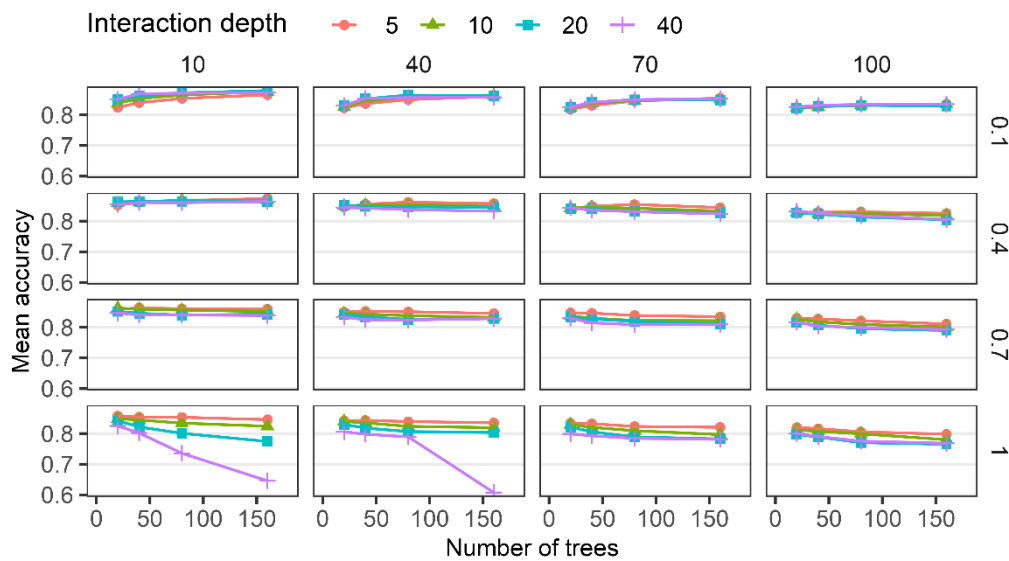
Notes: Mean 10-fold cross validated classification accuracy results for a single hidden layer artificial neural networks (ANN) using a sigmoid activation function, the number of neurons in the hidden layer $size = \{1, 2, \dots, 10\}$ and a regularization parameter $weight\ decay = \{0.1, 0.2, \dots, 0.5\}$ on the basis of information that is related to lottery design variables, socioeconomic characteristics, and past gambling behavior (*LSG*).

Figure A15 Mean CV accuracy results for RF for *LSG*



Notes: Mean 10-fold cross validated classification accuracy results for random forests (RF) using the Gini splitting rule, a number of 1500 individual bagged CART trees per forest, and as a function of the minimum number of samples in each leaf $min.\ node.\ size = \{1, 5, 10, 20, 40\}$ and the number of randomly selected predictors at each split $mtry = D_i(0.1, 0.2, 0.4, 0.8, 1)$ with D_i as the number of predictor variables that are related to lottery design variables, socioeconomic characteristics, and past gambling behavior (*LSG*).

Figure A16 Mean CV accuracy results for GBM for *LSG*

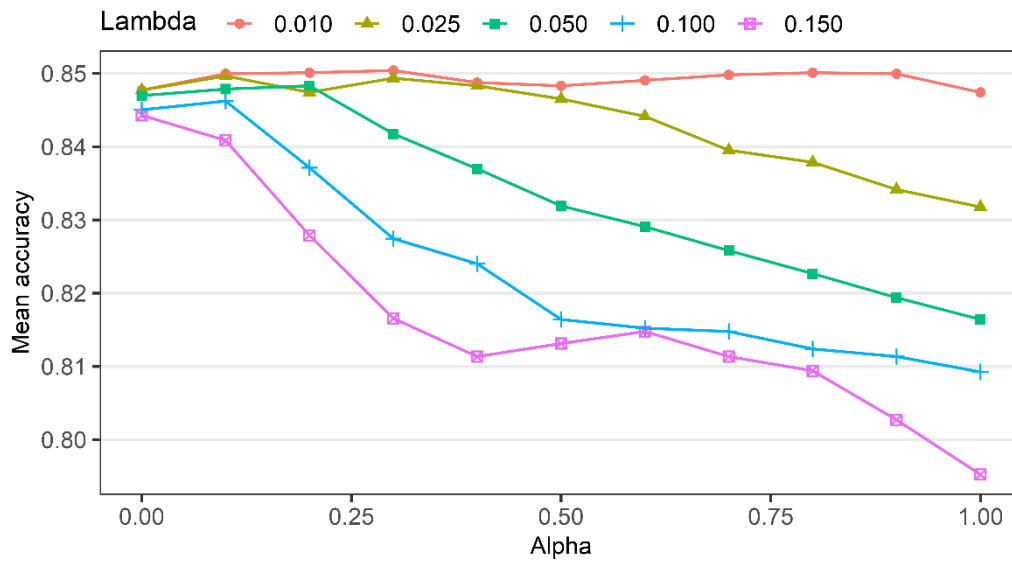


Notes: Mean 10-fold cross validated classification accuracy results for tree-based stochastic gradient boosting machines (GBM) as a function of the maximum tree depth *interaction.depth* = {5, 10, 20, 40} and number of boosting iterations *n.trees* = {20, 40, 80, 160}, a shrinkage factor *shrinkage* = {0.1, 0.4, 0.7, 1}, and a minimum number of samples in each leaf *n.minobsinnode* = {10, 40, 70, 100} on the basis of information that is related to lottery design variables, socioeconomic characteristics, and past gambling behavior (*LSG*).

3.2.4 Lottery design, socioeconomic and past gambling characteristics (*LSG*)

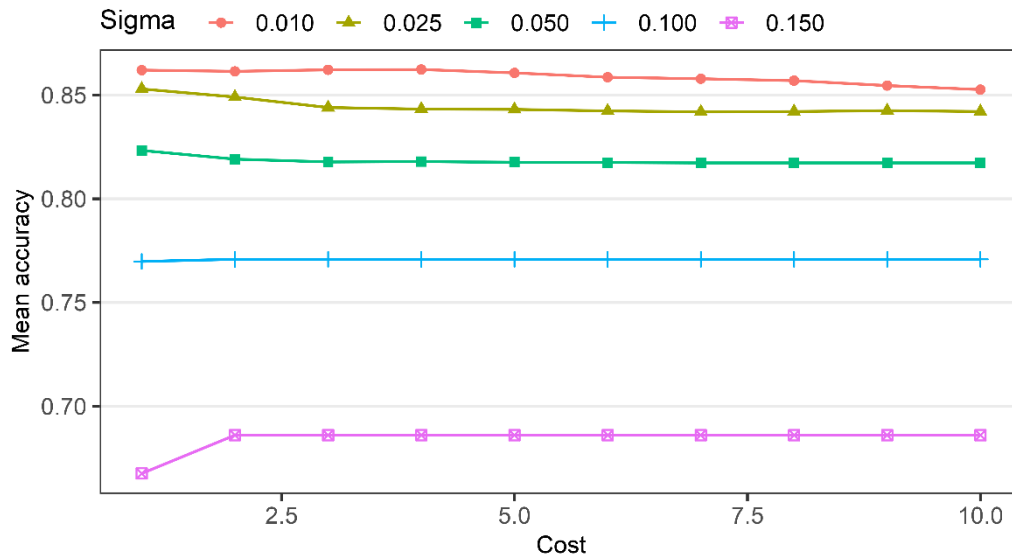
The following Figures (A17-A21) show the mean 10-fold cross-validation results for determining models' hyperparameters by a systematic grid search on the basis of the predictor-set *LSGPA* that includes 97 predictor variables. Precisely, *LSGPA* includes information on all predictor-sets and input categories: lottery design variables, socioeconomic characteristics, and past gambling behavior (*LSG*), individual psychophysiological reactions (*P*), and attention data (*A*).

Figure A17 Mean CV accuracy results for linear penalized regression models for *LSGPA*



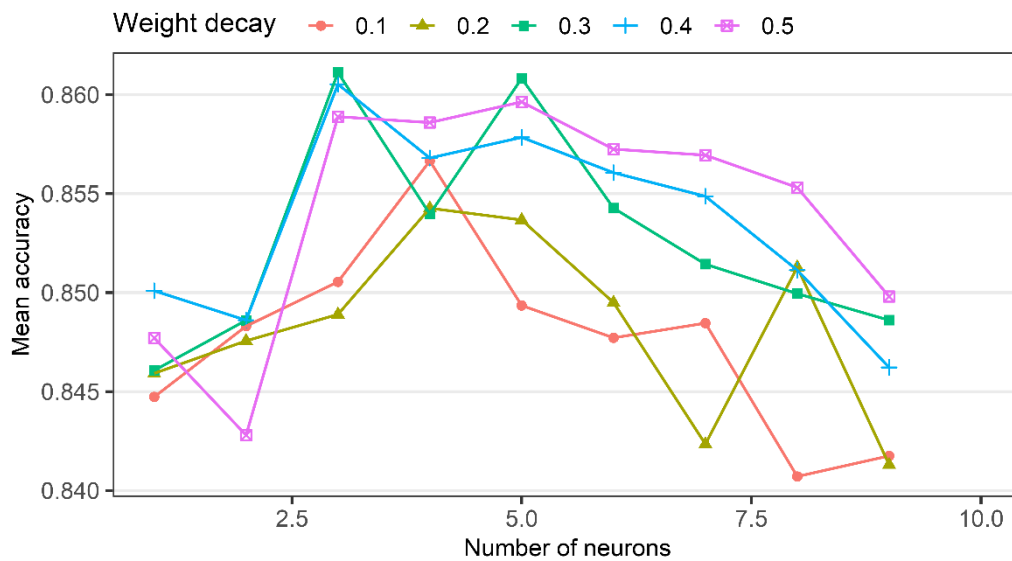
Notes: Mean 10-fold cross validated classification mean accuracy results for linear penalized regression models (lasso, ridge, and elastic net) as a function of the mixing percentage parameter $\alpha = \{0, 0.1, 0.2, \dots, 1\}$ and regularization parameter $\lambda = \{0.01, 0.025, 0.05, 0.1, 0.15\}$ on the basis of a predictor-set *LSG* that includes information that is related to lottery design variables, socioeconomic characteristics, and past gambling behavior (*LSG*), information on psychophysiological reactions (*P*) and attention data (*A*).

Figure A18 Mean CV accuracy results for SVM for *LSGPA*



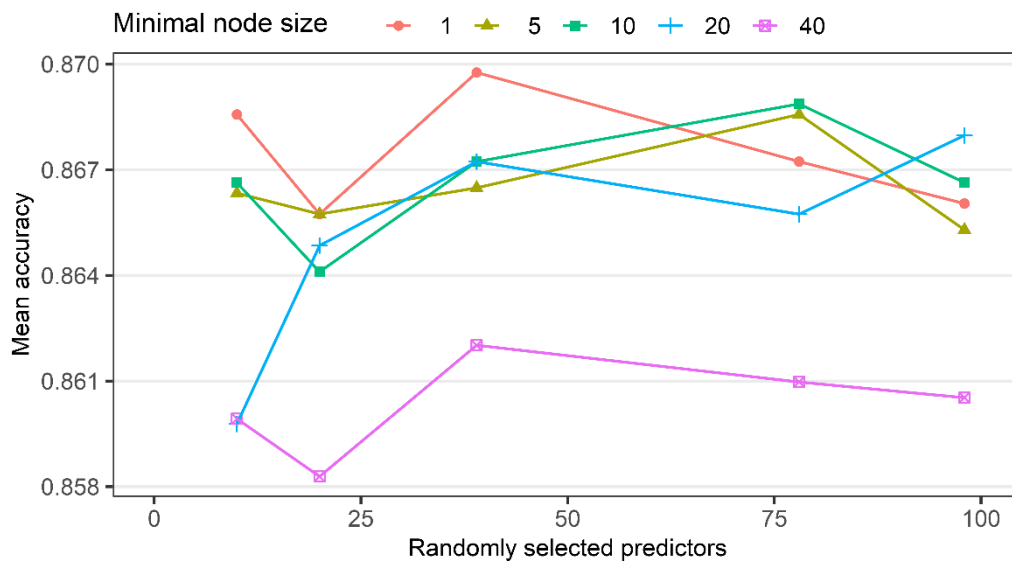
Notes: Mean 10-fold cross validated classification accuracy results for support vector machines (SVM) with a radial basis function kernel as a function of a cost parameter $C = \{1, 2, \dots, 10\}$ and inverse kernel width $\sigma = \{0.01, 0.025, 0.05, 0.1, 0.15\}$ on the basis of information that is related to lottery design variables, socioeconomic characteristics, and past gambling behavior (*LSG*), information on psychophysiological reactions (*P*) and attention data (*A*).

Figure A19 Mean CV accuracy results for SVM for *LSGPA*



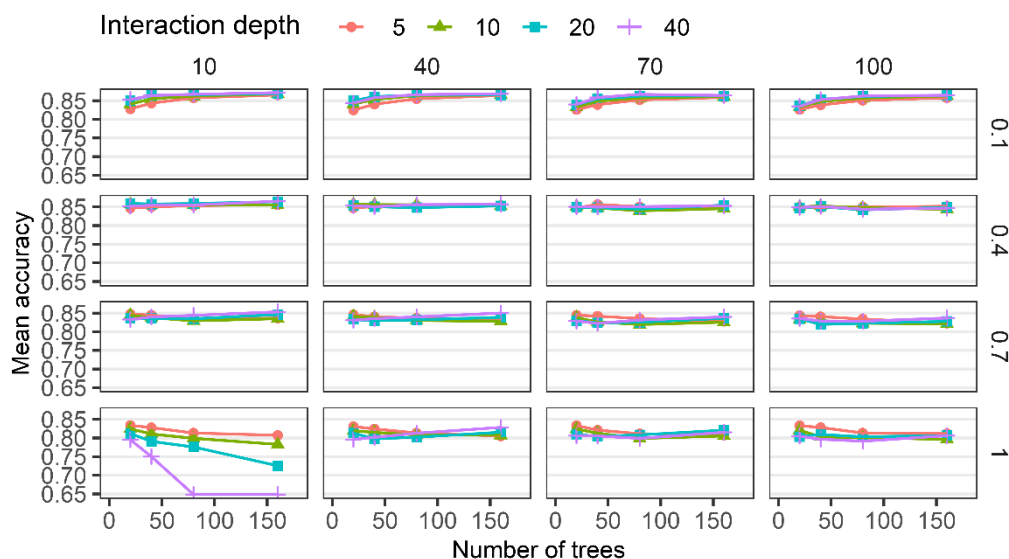
Notes: Mean 10-fold cross validated classification accuracy results for a single hidden layer artificial neural networks (ANN) using a sigmoid activation function, the number of neurons in the hidden layer $size = \{1, 2, \dots, 10\}$ and a regularization parameter $weight\ decay = 0.1, 0.2, \dots, 0.5$ on the basis of information that is related to lottery design variables, socioeconomic characteristics, and past gambling behavior (*LSG*), information on psychophysiological reactions (*P*) and attention data (*A*). We note that computations for $size = 10$ were not possible on the same CV partitions that we used for evaluating all specified models.

Figure A20 Mean CV accuracy results for RF for *LSGPA*



Notes: Mean 10-fold cross validated classification accuracy results for random forests (RF) using the Gini splitting rule, a number of 1500 individual bagged CART trees per forest, and as a function of the minimum number of samples in each leaf $min.\ node.\ size = \{1, 5, 10, 20, 40\}$ and the number of randomly selected predictors at each split $mtry = D_i(0.1, 0.2, 0.4, 0.8, 1)$ with D_i as the number of predictor variables (98) that are included in data that is related to information that is related to lottery design variables, socioeconomic characteristics, and past gambling behavior (*LSG*), information on psychophysiological reactions (*P*) and attention data (*A*).

Figure A21 Mean CV accuracy results for GBM for *LSGPA*



Notes: Mean 10-fold cross validated classification accuracy results for tree-based stochastic gradient boosting machines (GBM) as a function of the maximum tree depth $interaction.depth = \{5, 10, 20, 40\}$ and number of boosting iterations $n.trees = \{20, 40, 80, 160\}$, a shrinkage factor $shrinkage = \{0.1, 0.4, 0.7, 1\}$, and a minimum number of samples in each leaf $n.minobsinnode = \{10, 40, 70, 100\}$ on the basis of information that is related to lottery design variables, socioeconomic characteristics, and past gambling behavior (*LSG*), information on psychophysiological reactions (*P*) and attention data (*A*).

3.3 Cross validated accuracy results

Table A4 we show the selected models' hyperparameters according to highest CV mean accuracy. All other models' hyperparameter values are set to their default values (see the corresponding R packages, section 3.1). Moreover, we evaluate the model performance on the same left out CV observations for each round of CV.

Table A4 Selected model hyperparameters according to highest mean CV accuracy

Model: Elastic	<i>Dataset</i>			
<i>Hyperparameter</i>	P	A	LSG	LSGPA
Alpha	0.4	0.4	0.3	0.4
Lambda	0.01	0.01	0.01	0.01
CV Accuracy	0.60	0.60	0.85	0.85

Model: SVM	<i>Dataset</i>			
<i>Hyperparameter</i>	P	A	LSG	LSGPA
Kernel	Radial	Radial	Radial	Radial
Sigma	0.01	0.01	0.01	0.01
Cost	1	10	3	4
Accuracy	0.60	0.61	0.87	0.86

Model: ANN	<i>Dataset</i>			
<i>Hyperparameter</i>	P	A	LSG	LSGPA
Activation function	Sigmoid	Sigmoid	Sigmoid	Sigmoid
No. of neurons in the hidden layer	1	9	6	3
Weight decay	0.5	0.4	0.5	0.3
Accuracy	0.61	0.62	0.87	0.86

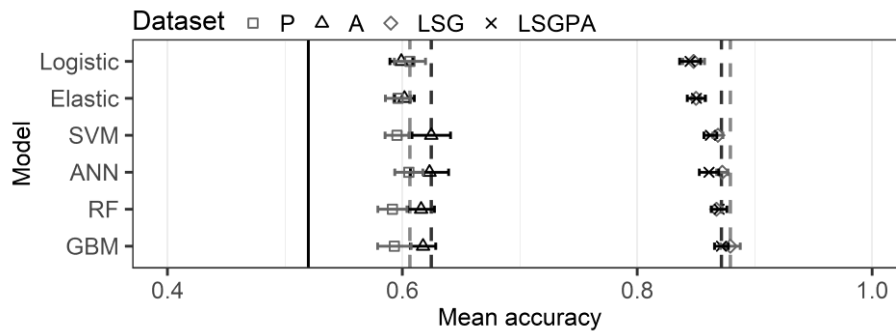
Model: RF	<i>Dataset</i>			
<i>Hyperparameter</i>	P	A	LSG	LSGPA
Splitting rule	Gini	Gini	Gini	Gini
No. of ensembled trees	1500	1500	1500	1500
Min. no. of samples in each leaf node	5	40	5	1
No. of predictors in each split	7	5	65	39
Accuracy	0.59	0.62	0.87	0.87

Model: GBM	<i>Dataset</i>			
<i>Hyperparameter</i>	P	A	LSG	LSGPA
Shrinkage	0.1	0.1	0.1	0.1
Interaction depth	5	5	20	40
Min. no. of samples in each leaf node	40	10	10	10
No. of ensembled trees	80	80	160	160
Accuracy	0.59	0.62	0.88	0.87

Notes: We determined models' hyperparameters via a systematic-grid search using 10-fold mean cross validation classification accuracy with subjects as strata. We evaluate logistic and penalized linear regression models (Elastic), support vector machines (SVM), artificial neural networks (ANN), random forests (RF), and tree-based gradient boosting machines (GBM) on the basis of psychophysiological (*P*) and attention (*A*) choice-process predictors; lottery-design, socioeconomic characteristics, and information on past gambling behavior (*LSG*); and a full-model that is comprised of all input categories (*LSGPA*).

For all results that we report in this study, we set the models' hyperparameters to the values that yield the highest mean CV accuracies. Figure 3 shows the out-of-sample classification accuracy for the 1657 records that are included in the test data.

Figure A22 Mean CV accuracy results for playing a 50/50 gamble



Notes: Mean 10-fold cross validation (CV) accuracy using 6711 observations for playing a 50/50 gamble with one potential loss- and one win-outcome using subjects as strata. We evaluate logistic (Logistic) and penalized linear regression models (Elastic), support vector machines (SVM), artificial neural networks (ANN), random forests (RF), and tree-based gradient boosting machines (GBM) on the basis of psychophysiological (*P*) and attention (*A*) choice-process predictors; lottery-design, socioeconomic characteristics, and information on past gambling behavior (*LSG*); and a full-model that is comprised of all input categories (*LSGPA*). Error bars correspond to 95% confidence intervals with respect to the 10 CV folds. The dashed lines correspond to the models that achieves the highest mean CV accuracy with respect to the different predictor-sets. The solid line is a naive forecast benchmark that yields a training data accuracy of 52% by predicting all records as not-playing.

Figure A22 shows that the more data-driven algorithms (SVM, ANN, RF, GBM) consistently outperform the generalized linear models that we evaluate in this study (Logistic and Elastic). While the nonlinear and tree-based ensemble methods result in more accurate forecasts based on *A* than on *P*, the corresponding differences for the linear methods are negligible. Moreover, the best CV-fit model's accuracy on *LSG* is marginally higher than on *LSGPA*, but as well negligible when considering the 10 CV-fold based 95% confidence intervals.

Table A5 shows the differences in models' CV accuracy estimates together with pairwise t-tests for detecting significant differences with Bonferroni-adjusted p-values.

Table A5 Models' differences in mean CV accuracies and pairwise t-test results

Data: P	Accuracy	Logistic	Elastic	SVM	ANN	RF	GBM
Logistic	0.61		0.01	0.01	0.00	0.01	0.01
Elastic	0.60	0.08		0.00	-0.01	0.00	0.00
SVM	0.60	0.29	1.00		-0.01	0.00	0.00
ANN	0.61	1.00	0.03**	0.11		0.01	0.01
RF	0.59	0.11	1.00	1.00	0.03**		0.00
GBM	0.59	0.97	1.00	1.00	0.72	1.00	

Data: A	Accuracy	Logistic	Elastic	SVM	ANN	RF	GBM
Logistic	0.60		0.00	-0.03	-0.02	-0.02	-0.02
Elastic	0.60	1.00		-0.02	-0.02	-0.01	-0.02
SVM	0.61	0.01***	0.22		0.00	0.01	0.01
ANN	0.62	0.10	0.26	1.00		0.01	0.01
RF	0.62	0.05**	0.44	1.00	1.00		0.00
GBM	0.62	0.00***	0.11	1.00	1.00	1.00	

Data: LSG	Accuracy	Logistic	Elastic	SVM	ANN	RF	GBM
Logistic	0.85		0.00	-0.02	-0.02	-0.02	-0.03
Elastic	0.85	1.00		-0.02	-0.02	-0.02	-0.03
SVM	0.87	0.00***	0.00***		-0.00	0.00	-0.01
ANN	0.87	0.01***	0.01***	1.00		0.00	-0.01
RF	0.87	0.03**	0.05**	1.00	1.00		-0.01
GBM	0.88	0.00***	0.00***	0.13	1.00	0.24	

Data: LSGPA	Accuracy	Logistic	Elastic	SVM	ANN	RF	GBM
Logistic	0.85		-0.01	-0.02	-0.02	-0.02	-0.03
Elastic	0.85	0.37		-0.01	-0.01	-0.02	-0.02
SVM	0.86	0.00***	0.01***		0.00	-0.01	-0.01
ANN	0.86	0.00***	0.21	1.00		-0.01	-0.01
RF	0.87	0.01***	0.01***	0.67	1.00		0.00
GBM	0.87	0.01***	0.01***	0.11	0.90	1.00	

Notes: Mean 10-fold CV accuracy differences for 6711 observations for playing a 50/50 gamble using subjects as strata. The upper triangle shows the pair-wise estimates of accuracy differences between models, and the lower triangle shows the Bonferroni adjusted p-values (with indicating significance as *** 0.01 ** 0.05 * 0.1) for pair-wise t-tests for detecting differences between models' accuracies with H_0 as a difference of zero. We evaluate logistic (Logistic) and penalized linear regression models (Elastic), support vector machines (SVM), artificial neural networks (ANN), random forests (RF), and tree-based gradient boosting machines (GBM) on the basis of psychophysiological (P) and attention (A) choice-process predictors; lottery-design, socioeconomic characteristics, and information on past gambling behavior (LSG); and a full-model that is comprised of all input categories (LSGPA). Model hyperparameters are selected via a systematic grid-search based on 10-fold CV.

Moreover, we report the linear correlation coefficients between models' CV accuracy estimates Table A6 to assess to which extent combining different modeling approaches may enhance our forecasting results.

Table A6 CV model accuracy correlations

Data: P	Mean	Logistic	Elastic	SVM	ANN	RF	GBM
Logistic	0.81		0.92	0.82	0.93	0.78	0.60
Elastic	0.90	0.92		0.94	0.94	0.93	0.78
SVM	0.86	0.82	0.94		0.88	0.86	0.78
ANN	0.86	0.93	0.94	0.88		0.87	0.68
RF	0.85	0.78	0.93	0.86	0.87		0.82
GBM	0.73	0.60	0.78	0.78	0.68	0.82	

Data: A	Mean	Logistic	Elastic	SVM	ANN	RF	GBM
Logistic	0.71		0.63	0.83	0.58	0.69	0.84
Elastic	0.51	0.63		0.43	0.47	0.45	0.59
SVM	0.73	0.83	0.43		0.82	0.86	0.69
ANN	0.67	0.58	0.47	0.82		0.87	0.62
RF	0.72	0.69	0.45	0.86	0.87		0.73
GBM	0.69	0.84	0.59	0.69	0.62	0.73	

Data: LSG	Mean	Logistic	Elastic	SVM	ANN	RF	GBM
Logistic	0.59		0.87	0.75	0.25	0.36	0.72
Elastic	0.41	0.87		0.60	<u>-0.05</u>	0.06	0.59
SVM	0.46	0.75	0.60		0.00	0.26	0.70
ANN	0.09	0.25	<u>-0.05</u>	0.00		0.01	0.24
RF	0.22	0.36	0.06	0.26	0.01		0.41
GBM	0.53	0.72	0.59	0.70	0.24	0.41	

Data: LSGPA	Mean	Logistic	Elastic	SVM	ANN	RF	GBM
Logistic	0.62		0.90	0.82	0.84	0.32	0.22
Elastic	0.65	0.90		0.85	0.63	0.46	0.40
SVM	0.67	0.82	0.85		0.63	0.47	0.57
ANN	0.46	0.84	0.63	0.63		0.09	0.10
RF	0.38	0.32	0.46	0.47	0.09		0.57
GBM	0.37	0.22	0.40	0.57	0.10	0.57	

Notes: Linear correlation coefficients and mean correlation for 10-fold mean cross validation classification accuracy on 6711 observations for playing a 50/50 gamble with one potential loss- and one win-outcome using subjects as strata. The diagonal correlations are left out for reasons of clarity. We evaluate logistic (Logistic) and penalized linear regression models (Elastic), support vector machines (SVM), artificial neural networks (ANN), random forests (RF), and tree-based gradient boosting machines (GBM) on the basis of psychophysiological (*P*) and attention (*A*) choice-process predictors; lottery-design, socioeconomic characteristics, and information on past gambling behavior (*LSG*); and a full-model that is comprised of all input categories (*LSGPA*). Model hyperparameters were selected via a systematic grid-search based on 10-fold CV. We highlight correlations that are smaller than 0.5 (bold) and negative (bold, underlined).

Table A6 shows that there is large between model accuracy correlation for the *P* data. With respect to the *A* data, Elastic, SVM, ANN, and RF show a correlation smaller than 0.5. However, we observe more variation for model-accuracy correlations for the *LSG*, and *LSGPA* predictor-sets. For *LSGPA*, there are 7 out of 15 relevant inter-model accuracy correlations that are smaller 0.5, and for *LSG* there are 9 correlations smaller than 0.5. Moreover, for *LSG* the correlation between ANN and Elastic predictions is negative (-

0.05); however, ANN is consistently outperforming the linear penalized regression methods.

3.4 Out-of-sample forecasting results by model, lotteries' expected values, and predictive measures

Table A7 reports various predictive measures for classification accuracy for the best training data CV fit methods together with the corresponding results for the RDR and SDR predictions.

Table A7 Out-of-sample classification performance measures for the best CV-fit models

Dataset: Model	RDR			P: Logistic			A: SVM		
Expected value	All	PEV	NEV	All	PEV	NEV	All	PEV	NEV
Accuracy	0.54	0.42	0.95	0.60	0.60	0.62	0.64	0.65	0.64
Sensitivity	0	0	0	0.50	0.50	0.72	0.56	0.56	0.53
Specificity	1	1	1	0.69	0.73	0.62	0.71	0.76	0.64
Pos pred value	0	0	0	0.58	0.72	0.09	0.62	0.76	0.07
Neg pred value	0.54	0.42	0.95	0.62	0.51	0.98	0.65	0.56	0.96
Dataset: Model	SDR			LSG: GBM			LSGPA: GBM		
Expected value	All	PEV	NEV	All	PEV	NEV	All	PEV	NEV
Accuracy	0.67	0.58	0.95	0.88	0.85	0.95	0.87	0.85	0.95
Sensitivity	0.98	1	0	0.88	0.89	0.39	0.87	0.89	0.28
Specificity	0.40	0	1	0.87	0.80	0.98	0.87	0.80	0.98
Pos pred value	0.58	0.58	0	0.85	0.86	0.47	0.85	0.86	0.45
Neg pred value	0.95	0	0.95	0.90	0.84	0.97	0.89	0.84	0.96

Notes: Out-of-sample classification performance metrics based on positive expected value (PEV), negative expected value (NEV), and all lotteries (All) for playing a 50/50 gamble. We evaluate two naive benchmark forecasts (RDR and SDR), logistic and penalized linear regression models (Elastic), support vector machines (SVM), artificial neural networks (ANN), random forests (RF), and tree-based gradient boosting machines (GBM) on the basis of psycho-physiological (*P*) and attention (*A*) choice-process data; lottery-design, socioeconomic characteristics, and information on past gambling behavior (*LSG*); and a full-model is comprised of all input categories (*LSGPA*). Models' hyperparameters were determined via a systematic grid-search using 10-fold cross validation and subjects as strata. The highest value for classification measure-specific performance by lotteries expected values is highlighted (bold).

Table A7 shows the differences across predictive measures for models' test data performance by lotteries' EV. Among the models that we would choose based on their CV performance with respect to the four predictor-sets, we find the logistic regression model using *P* to yield the best predictions for forecasting risky choices for NEVL in terms of negative predicted value accuracy with 98%. At the same time, the corresponding predictive accuracy in terms of positive predicted value (PPV) is only 9%. However, there are 347 NEVL included in the 1557 observations that we specify as test data, and we observe

18 played lotteries out of these 347 NEVL. As a consequence, the SDR shows a sensitivity of 98% over all test records.

Moreover, Table A7 reveals the large differences between the PPV results for NEVL that are produced for the Logistic model for *A* (7%) and SVM for *P* (9%) in comparison to GBM for *LSG* (46%) and GBM for *LSGPA* (63%). The best performing models that only include CPD input categories and dummy variables for subjects (SVM for *A* and Logistic for *P*) result in a mean test accuracy of 97% in terms of PPV for NEVL. With a corresponding PPV of 95% that is produced by the RDR and SDR forecasts, this increase in predictive power equates to a 60% classification error decrease. In contrast, with respect to GBM's test specificity for NEVL, the additional information that is included in *P* and *A* results in increasing the classification error from 1% (*LSG*) to 2% (*LSGPA*), i.e., by 100%.

The detailed out-of-sample results for all models, predictor-sets, and predictive classification measures by lotteries' expected value are reported in Table A8, A9, and A10.

Table A8 Model out-of-sample performance metrics over all lotteries

<i>Measure</i>	Logistic	Elastic	SVM	ANN	RF	GBM	Data
Accuracy	0.60	0.61	0.61	0.60	0.59	0.60	P
Lower accuracy	0.58	0.58	0.58	0.58	0.57	0.58	P
Upper accuracy	0.63	0.63	0.63	0.63	0.62	0.63	P
Sensitivity	0.50	0.40	0.44	0.49	0.48	0.49	P
Specificity	0.69	0.79	0.75	0.70	0.69	0.70	P
Positive predicted value	0.58	0.61	0.60	0.59	0.57	0.58	P
Negative predicted value	0.62	0.60	0.61	0.62	0.61	0.62	P
Balanced accuracy	0.60	0.59	0.60	0.60	0.59	0.60	P

<i>Measure</i>	Logistic	Elastic	SVM	ANN	RF	GBM	Data
Accuracy	0.61	0.60	0.64	0.64	0.64	0.63	A
Lower accuracy	0.58	0.58	0.62	0.61	0.61	0.60	A
Upper accuracy	0.63	0.63	0.66	0.66	0.66	0.65	A
Sensitivity	0.51	0.48	0.56	0.57	0.59	0.55	A
Specificity	0.69	0.71	0.71	0.69	0.68	0.69	A
Positive predicted value	0.59	0.59	0.62	0.61	0.61	0.61	A
Negative predicted value	0.62	0.61	0.65	0.65	0.66	0.64	A
Balanced accuracy	0.60	0.60	0.63	0.63	0.63	0.62	A

<i>Measure</i>	Logistic	Elastic	SVM	ANN	RF	GBM	Data
Accuracy	0.85	0.85	0.87	0.87	0.87	0.88	LSG
Lower accuracy	0.84	0.83	0.86	0.85	0.85	0.86	LSG
Upper accuracy	0.87	0.86	0.89	0.88	0.88	0.89	LSG
Sensitivity	0.86	0.85	0.88	0.88	0.85	0.88	LSG
Specificity	0.85	0.84	0.86	0.86	0.88	0.87	LSG
Positive predicted value	0.83	0.82	0.85	0.84	0.86	0.85	LSG
Negative predicted value	0.88	0.87	0.90	0.89	0.87	0.90	LSG
Balanced accuracy	0.85	0.85	0.87	0.87	0.87	0.88	LSG

<i>Measure</i>	Logistic	Elastic	SVM	ANN	RF	GBM	Data
Accuracy	0.85	0.85	0.87	0.86	0.87	0.87	LSGPA
Lower accuracy	0.83	0.83	0.85	0.84	0.85	0.85	LSGPA
Upper accuracy	0.87	0.86	0.89	0.88	0.89	0.89	LSGPA
Sensitivity	0.87	0.86	0.86	0.85	0.87	0.87	LSGPA
Specificity	0.84	0.84	0.87	0.87	0.88	0.87	LSGPA
Positive predicted value	0.82	0.82	0.86	0.85	0.86	0.85	LSGPA
Negative predicted value	0.88	0.87	0.88	0.87	0.88	0.89	LSGPA
Balanced accuracy	0.85	0.85	0.87	0.86	0.87	0.87	LSGPA

Notes: Out-of-sample classification performance metrics for all lotteries' EV for playing a 50/50 gamble. We evaluate two naive benchmark forecasts (RDR and SDR), logistic (Logistic) and penalized linear regression models (Elastic), support vector machines (SVM), artificial neural networks (ANN), random forests (RF), and tree-based gradient boosting machines (GBM) on the basis of psychophysiological (P) and attention (A) choice-process data; lottery-design, socioeconomic characteristics, and information on past gambling behavior (LSG); and a full-model is comprised of all input categories (LSGPA). Models' hyperparameters were determined via a systematic grid-search using 10-fold CV and subjects as strata.

Table A9 Out-of-sample performance metrics for positive expected value lotteries

<i>Measure</i>	Logistic	Elastic	SVM	ANN	RF	GBM	Data
Accuracy	0.60	0.59	0.59	0.60	0.59	0.59	P
Lower accuracy	0.57	0.56	0.56	0.57	0.57	0.57	P
Upper accuracy	0.62	0.61	0.62	0.62	0.62	0.62	P
Sensitivity	0.50	0.39	0.44	0.48	0.48	0.49	P
Specificity	0.73	0.85	0.80	0.76	0.75	0.74	P
Positive predicted value	0.72	0.79	0.76	0.73	0.73	0.72	P
Negative predicted value	0.51	0.50	0.51	0.51	0.51	0.51	P
Balanced accuracy	0.62	0.62	0.62	0.62	0.62	0.61	P

<i>Measure</i>	Logistic	Elastic	SVM	ANN	RF	GBM	Data
Accuracy	0.61	0.60	0.65	0.64	0.64	0.63	A
Lower accuracy	0.58	0.57	0.62	0.62	0.62	0.60	A
Upper accuracy	0.63	0.62	0.67	0.67	0.67	0.66	A
Sensitivity	0.51	0.47	0.56	0.57	0.59	0.55	A
Specificity	0.74	0.76	0.76	0.74	0.72	0.74	A
Positive predicted value	0.73	0.73	0.76	0.75	0.75	0.75	A
Negative predicted value	0.52	0.51	0.56	0.55	0.56	0.54	A
Balanced accuracy	0.62	0.62	0.66	0.66	0.66	0.65	A

<i>Measure</i>	Logistic	Elastic	SVM	ANN	RF	GBM	Data
Accuracy	0.83	0.82	0.85	0.84	0.84	0.86	LSG
Lower accuracy	0.81	0.80	0.83	0.82	0.82	0.84	LSG
Upper accuracy	0.85	0.84	0.87	0.86	0.86	0.87	LSG
Sensitivity	0.88	0.88	0.90	0.89	0.86	0.90	LSG
Specificity	0.76	0.75	0.79	0.78	0.82	0.80	LSG
Positive predicted value	0.84	0.83	0.85	0.85	0.87	0.86	LSG
Negative predicted value	0.82	0.81	0.85	0.84	0.81	0.85	LSG
Balanced accuracy	0.82	0.81	0.84	0.83	0.84	0.85	LSG

<i>Measure</i>	Logistic	Elastic	SVM	ANN	RF	GBM	Data
Accuracy	0.83	0.82	0.85	0.83	0.85	0.85	LSGPA
Lower accuracy	0.81	0.80	0.83	0.81	0.83	0.83	LSGPA
Upper accuracy	0.85	0.84	0.87	0.85	0.87	0.87	LSGPA
Sensitivity	0.89	0.88	0.88	0.86	0.88	0.89	LSGPA
Specificity	0.75	0.75	0.81	0.79	0.80	0.80	LSGPA
Positive predicted value	0.83	0.83	0.86	0.85	0.86	0.86	LSGPA
Negative predicted value	0.83	0.81	0.83	0.80	0.83	0.83	LSGPA
Balanced accuracy	0.82	0.81	0.84	0.83	0.84	0.84	LSGPA

Notes: Out-of-sample classification performance metrics for positive expected value (PEV) lotteries for playing a 50/50 gamble. We evaluate two naive benchmark forecasts (RDR and SDR), logistic and penalized linear regression models (Elastic), support vector machines (SVM), artificial neural networks (ANN), random forests (RF), and tree-based gradient boosting machines (GBM) on the basis of psychophysiological (*P*) and attention (*A*) choice-process data; lottery-design, socioeconomic characteristics, and information on past gambling behavior (*LSG*); and a full-model that a full-model is comprised of all input categories (*LSGPA*). Models' hyperparameters were determined via a systematic grid-search using 10-fold cross validation and subjects as strata.

Table A10 Out-of-sample performance metrics for negative expected value lotteries

<i>Measure</i>	Logistic	Elastic	SVM	ANN	RF	GBM	Data
Accuracy	0.62	0.68	0.67	0.63	0.59	0.64	P
Lower accuracy	0.57	0.63	0.62	0.58	0.54	0.59	P
Upper accuracy	0.67	0.73	0.71	0.68	0.64	0.69	P
Sensitivity	0.72	0.50	0.44	0.78	0.56	0.61	P
Specificity	0.62	0.69	0.68	0.62	0.59	0.64	P
Positive predicted value	0.09	0.08	0.07	0.09	0.06	0.08	P
Negative predicted value	0.98	0.96	0.96	0.98	0.96	0.97	P
Balanced accuracy	0.67	0.59	0.56	0.70	0.57	0.63	P

<i>Measure</i>	Logistic	Elastic	SVM	ANN	RF	GBM	Data
Accuracy	0.64	0.65	0.64	0.62	0.62	0.63	A
Lower accuracy	0.58	0.60	0.58	0.57	0.57	0.58	A
Upper accuracy	0.68	0.70	0.68	0.67	0.67	0.68	A
Sensitivity	0.74	0.68	0.53	0.58	0.68	0.68	A
Specificity	0.63	0.65	0.64	0.62	0.61	0.62	A
Positive predicted value	0.09	0.09	0.07	0.07	0.08	0.09	A
Negative predicted value	0.98	0.98	0.96	0.97	0.97	0.97	A
Balanced accuracy	0.68	0.67	0.58	0.60	0.65	0.65	A

<i>Measure</i>	Logistic	Elastic	SVM	ANN	RF	GBM	Data
Accuracy	0.94	0.94	0.95	0.95	0.95	0.95	LSG
Lower accuracy	0.91	0.91	0.92	0.92	0.93	0.92	LSG
Upper accuracy	0.96	0.96	0.97	0.97	0.97	0.97	LSG
Sensitivity	0.11	0.00	0.33	0.28	0.39	0.39	LSG
Specificity	0.98	0.98	0.98	0.98	0.98	0.98	LSG
Positive predicted value	0.20	0.00	0.46	0.45	0.54	0.47	LSG
Negative predicted value	0.96	0.95	0.97	0.96	0.97	0.97	LSG
Balanced accuracy	0.54	0.49	0.66	0.63	0.69	0.68	LSG

<i>Measure</i>	Logistic	Elastic	SVM	ANN	RF	GBM	Data
Accuracy	0.93	0.93	0.94	0.95	0.96	0.95	LSGPA
Lower accuracy	0.90	0.90	0.91	0.93	0.93	0.92	LSGPA
Upper accuracy	0.95	0.96	0.96	0.97	0.98	0.97	LSGPA
Sensitivity	0.11	0.00	0.28	0.44	0.28	0.28	LSGPA
Specificity	0.97	0.98	0.97	0.98	0.99	0.98	LSGPA
Positive predicted value	0.17	0.00	0.36	0.53	0.63	0.45	LSGPA
Negative predicted value	0.96	0.95	0.96	0.97	0.96	0.96	LSGPA
Balanced accuracy	0.54	0.49	0.63	0.71	0.63	0.63	LSGPA

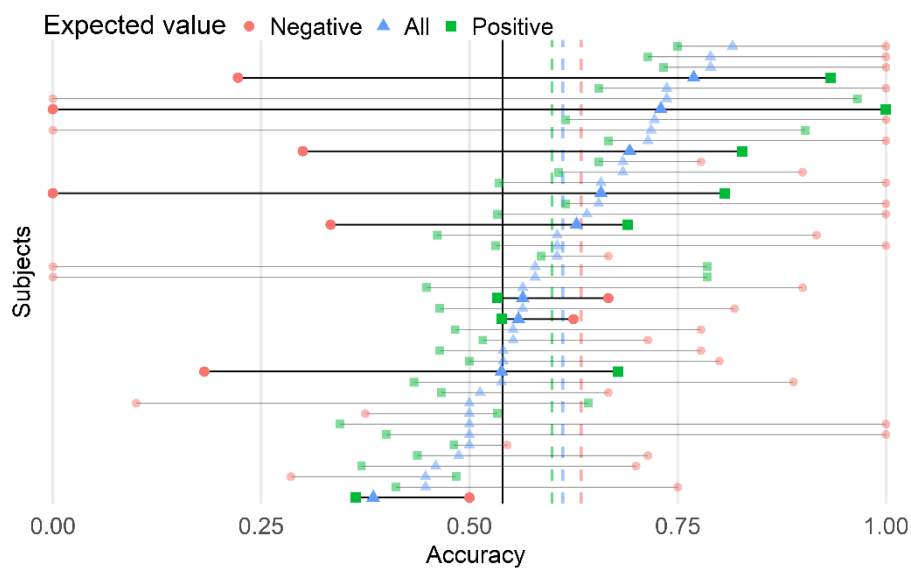
Notes: Out-of-sample classification performance metrics for negative expected value (NEV) lotteries (All) for playing a 50/50 gamble. We evaluate two naive benchmark forecasts (RDR and SDR), logistic and penalized linear regression models (Elastic), support vector machines (SVM), artificial neural networks (ANN), random forests (RF), and tree-based gradient boosting machines (GBM) on the basis of psychophysiological (P) and attention (A) choice-process data; lottery-design, socioeconomic characteristics, and information on past gambling behavior (LSG); and a full-model a full-model is comprised of all input categories (LSGPA). Models' hyperparameters were determined via a systematic grid-search using 10-fold cross validation and subjects as strata.

3.5 Out-of-sample forecasting results by model, predictor-set and subjects

In this section we present the omitted out-of-sample accuracy results for the 44 subjects by lotteries' EV that are produced by the models' hyperparameter specifications that yield the most accurate mean 10-fold subject specific stratified CV accuracy results on the training data.

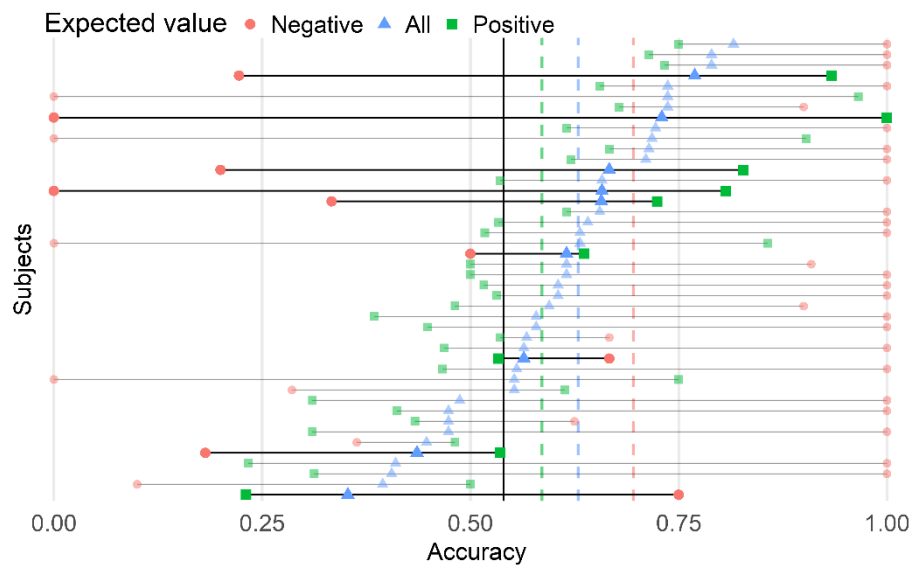
3.5.1 Psychophysiological reactions (*P*)

Figure A23 Out-of-sample accuracy results for the logistic regression model for *P*



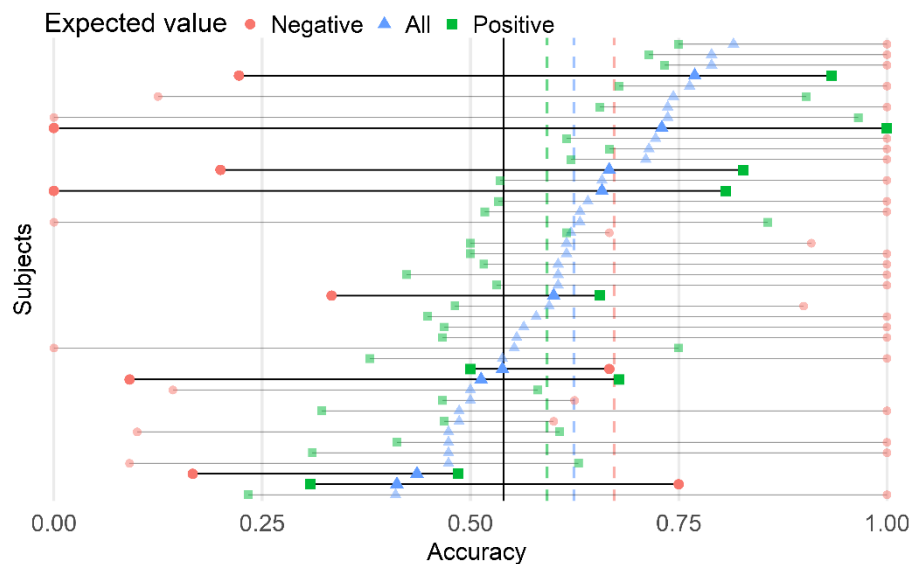
Notes: Out-of-sample accuracy for 44 subjects by positive expected value (PEV), negative expected value (NEV), and all lotteries (All) for playing a 50/50 gamble by a logistic regression model on the basis of predictors that relate to individual psychophysiological (*P*) choice-process data. The 9 subjects that play five or more NEV lotteries (NEVL) with respect to 45 NEVL that are included in the full experiment data are highlighted (bold; see Figure 2, section 2). Dashed lines correspond to average subject mean accuracy by lotteries' EV. The solid line is a naive forecast that achieves an accuracy of 54% on the test data by predicting all records as not-playing.

Figure A24 Out-of-sample accuracy results for the elastic regression model for P



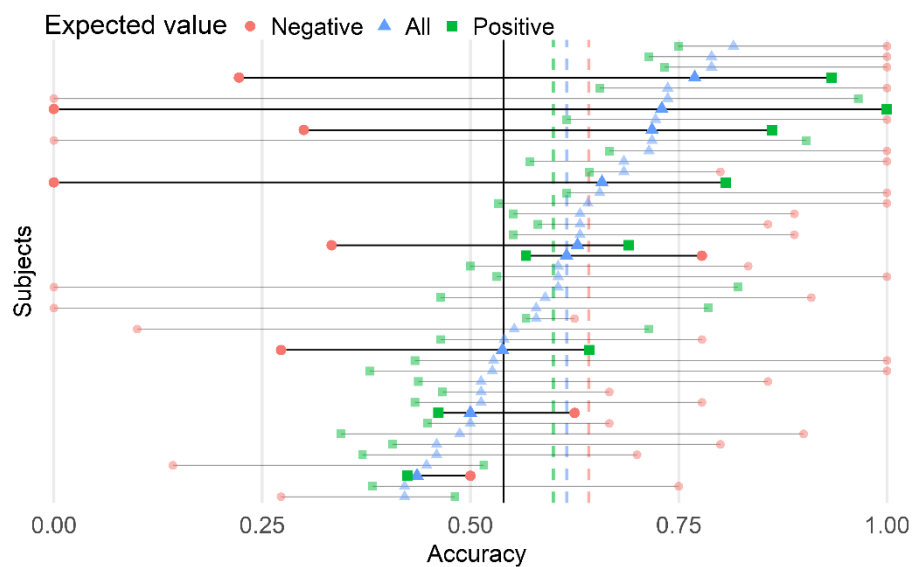
Notes: Out-of-sample accuracy for 44 subjects by positive expected value (PEV), negative expected value (NEV), and all lotteries (All) for playing a 50/50 gamble by an Elastic regression model on the basis of predictors that relate to individual psychophysiological (P) choice-process data. The 9 subjects that play five or more NEV lotteries (NEVL) with respect to 45 NEVL that are included in the full experiment data are highlighted (bold; see Figure 2, section 2). Dashed lines correspond to average subject mean accuracy by lotteries' EV. The solid line is a naive forecast that achieves an accuracy of 54% on the test data by predicting all records as not-playing.

Figure A25 Out-of-sample accuracy results for SVM for P



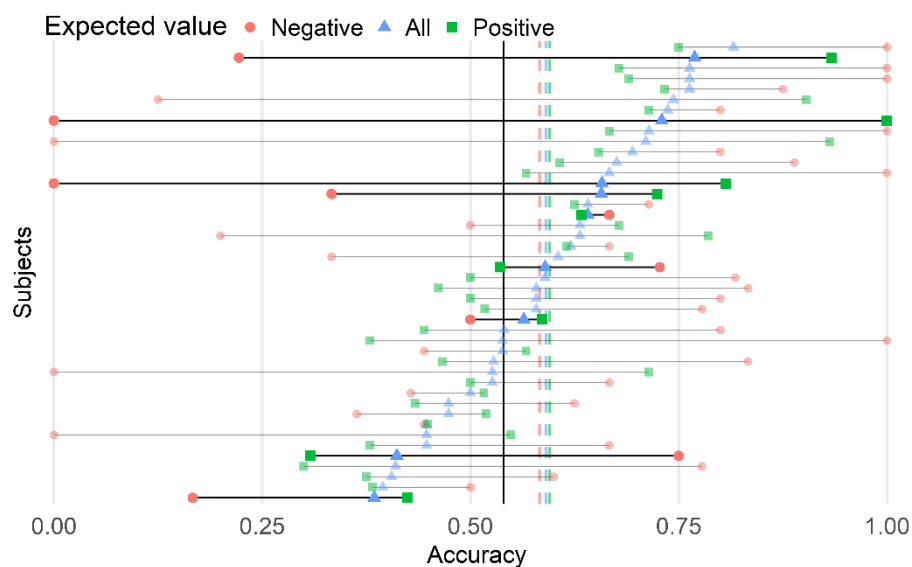
Notes: Out-of-sample accuracy for 44 subjects by positive expected value (PEV), negative expected value (NEV), and all lotteries (All) for playing a 50/50 gamble by an SVM model on the basis of predictors that relate to individual psychophysiological (P) choice-process data. The 9 subjects that play five or more NEV lotteries (NEVL) with respect to 45 NEVL that are included in the full experiment data are highlighted (bold; see Figure 2, section 2). Dashed lines correspond to average subject mean accuracy by lotteries' EV. The solid line is a naive forecast that achieves an accuracy of 54% on the test data by predicting all records as not-playing.

Figure A26 Out-of-sample accuracy results for ANN for P



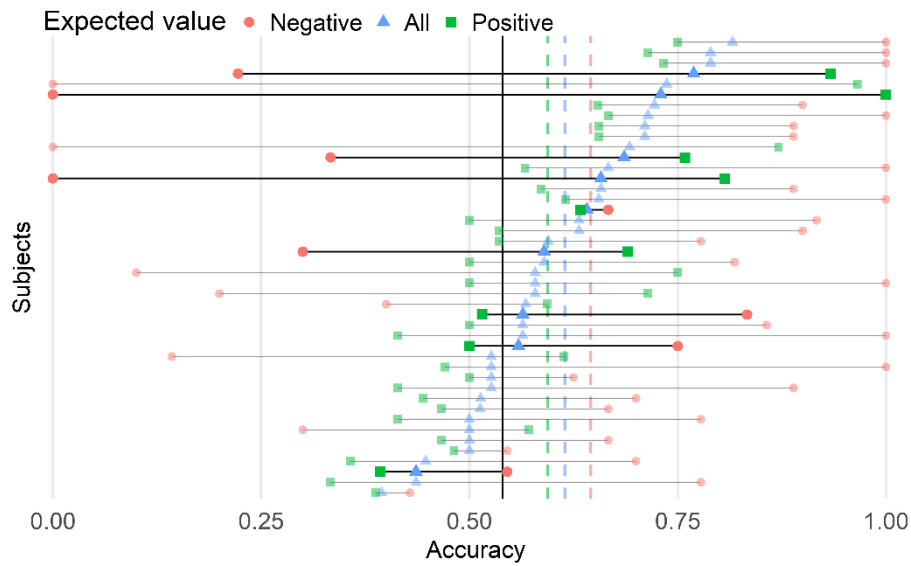
Notes: Out-of-sample accuracy for 44 subjects by positive expected value (PEV), negative expected value (NEV), and all lotteries (All) for playing a 50/50 gamble by an ANN model on the basis of predictors that relate to individual psychophysiological (P) choice-process data. The 9 subjects that play five or more NEV lotteries (NEVL) with respect to 45 NEVL that are included in the full experiment data are highlighted (bold; see Figure 2, section 2). Dashed lines correspond to average subject mean accuracy by lotteries' EV. The solid line is a naive forecast that achieves an accuracy of 54% on the test data by predicting all records as not-playing.

Figure A27 Out-of-sample accuracy results for RF for P



Notes: Out-of-sample accuracy for 44 subjects by positive expected value (PEV), negative expected value (NEV), and all lotteries (All) for playing a 50/50 gamble by an RF model on the basis of predictors that relate to individual psychophysiological (P) choice-process data. The 9 subjects that play five or more NEV lotteries (NEVL) with respect to 45 NEVL that are included in the full experiment data are highlighted (bold; see Figure 2, section 2). Dashed lines correspond to average subject mean accuracy by lotteries' EV. The solid line is a naive forecast that achieves an accuracy of 54% on the test data by predicting all records as not-playing.

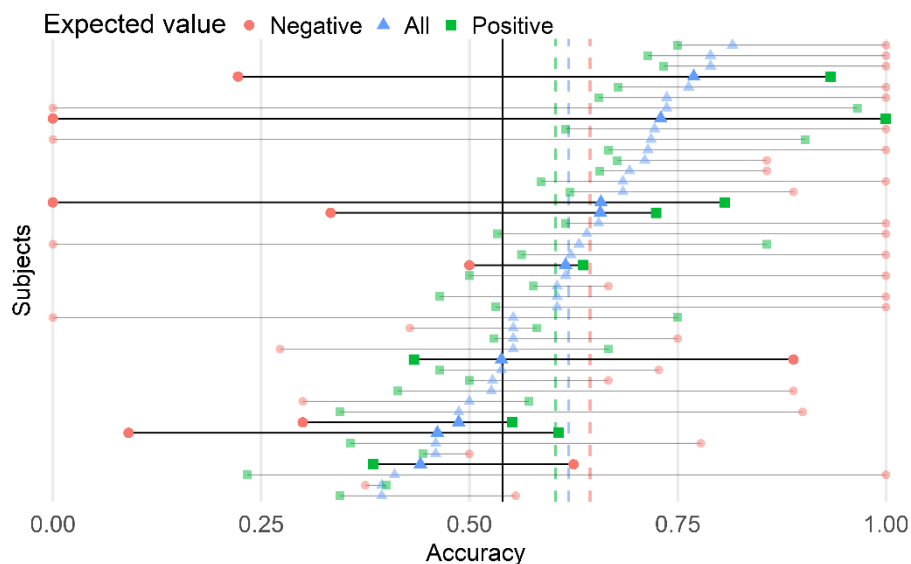
Figure A28 Out-of-sample accuracy results for GBM for P



Notes: Out-of-sample accuracy for 44 subjects by positive expected value (PEV), negative expected value (NEV), and all lotteries (All) for playing a 50/50 gamble by a GBM model on the basis of predictors that relate to individual psychophysiological (P) choice-process data. The 9 subjects that play five or more NEV lotteries (NEVL) with respect to 45 NEVL that are included in the full experiment data are highlighted (bold; see Figure 2, section 2). Dashed lines correspond to average subject mean accuracy by lotteries' EV. The solid line is a naive forecast that achieves an accuracy of 54% on the test data by predicting all records as not-playing.

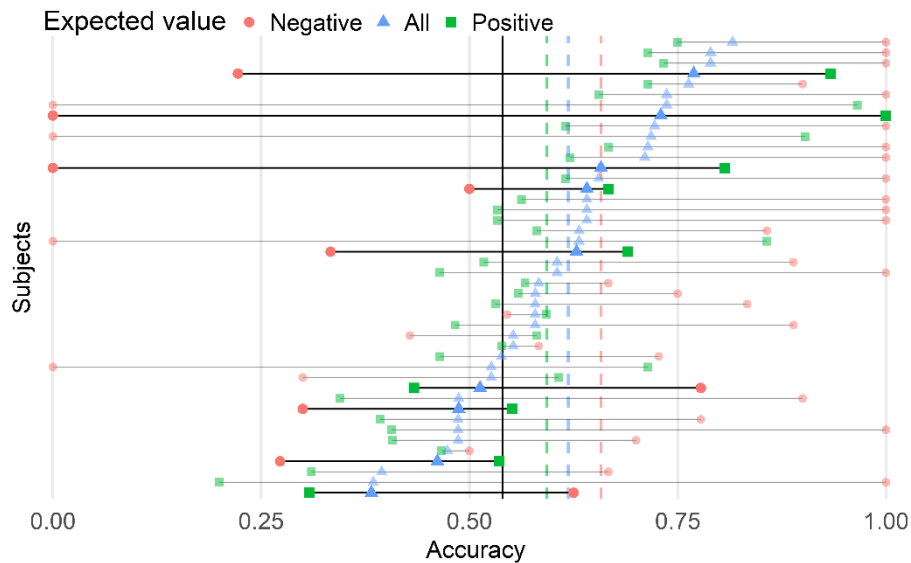
3.5.2 Attention data (A)

Figure A29 Out-of-sample accuracy results for the logistic regression model for A



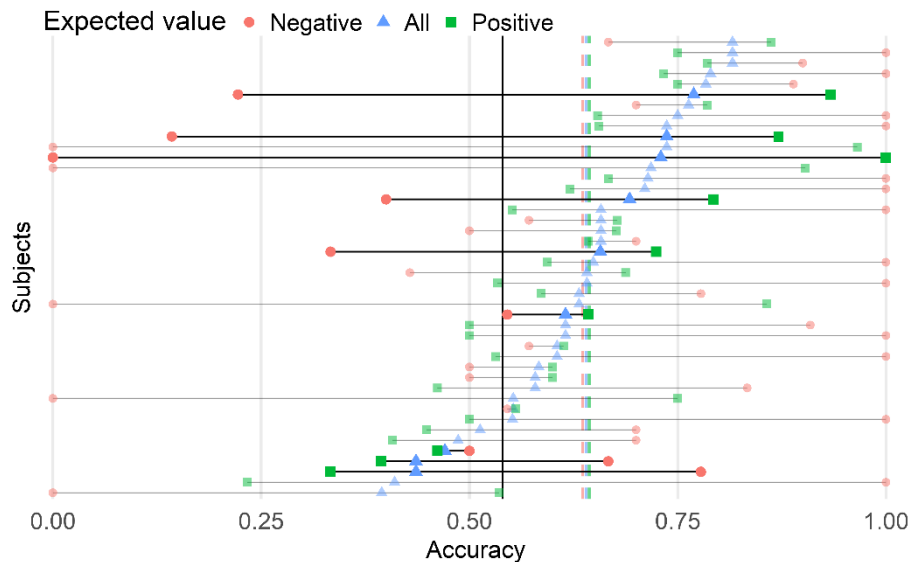
Notes: Out-of-sample accuracy for 44 subjects by positive expected value (PEV), negative expected value (NEV), and all lotteries (All) for playing a 50/50 gamble by a logistic regression model on the basis of predictors that relate to attention (A) choice-process data. The 9 subjects that play five or more NEV lotteries (NEVL) with respect to 45 NEVL that are included in the full experiment data are highlighted (bold; see Figure 2, section 2). Dashed lines correspond to average subject mean accuracy by lotteries' EV. The solid line is a naive forecast that achieves an accuracy of 54% on the test data by predicting all records as not-playing.

Figure A30 Out-of-sample accuracy results for the elastic regression model for A



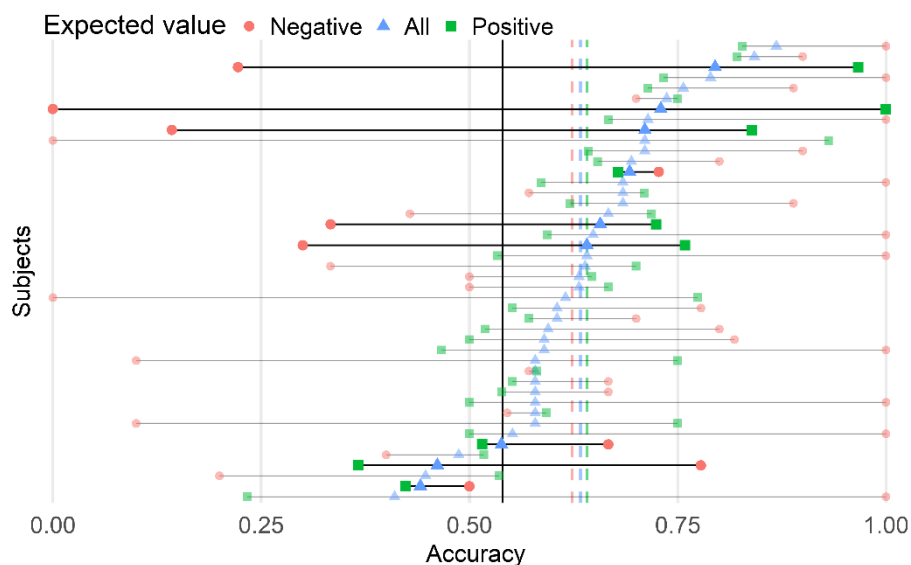
Notes: Out-of-sample accuracy for 44 subjects by positive expected value (PEV), negative expected value (NEV), and all lotteries (All) for playing a 50/50 gamble by an Elastic regression model on the basis of predictors that relate to attention (A) choice-process data. The 9 subjects that play five or more NEV lotteries (NEVL) with respect to 45 NEVL that are included in the full experiment data are highlighted (bold; see Figure 2, section 2). Dashed lines correspond to average subject mean accuracy by lotteries' EV. The solid line is a naive forecast that achieves an accuracy of 54% on the test data by predicting all records as not-playing.

Figure A31 Out-of-sample accuracy results for SVM for A



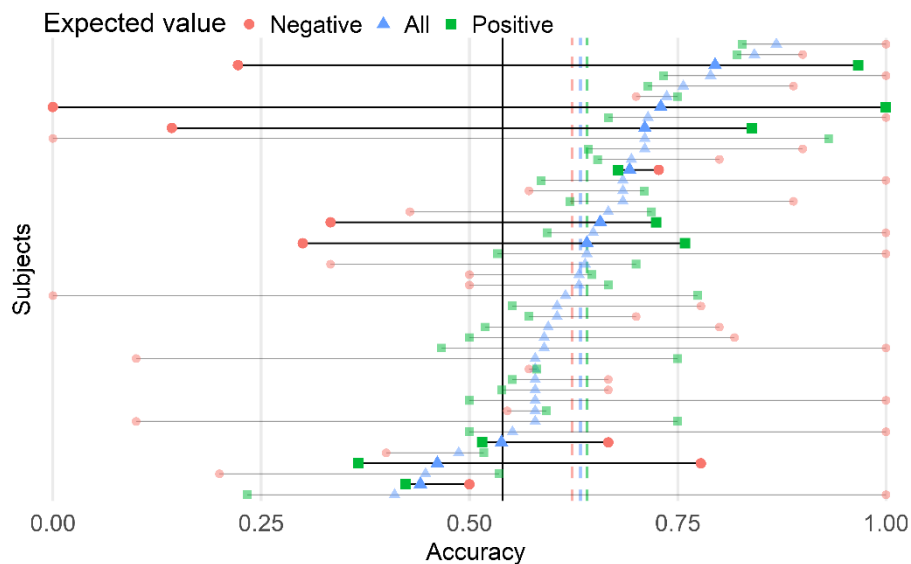
Notes: Out-of-sample accuracy for 44 subjects by positive expected value (PEV), negative expected value (NEV), and all lotteries (All) for playing a 50/50 gamble by an SVM model on the basis of predictors that relate to attention (A) choice-process data. The 9 subjects that play five or more NEV lotteries (NEVL) with respect to 45 NEVL that are included in the full experiment data are highlighted (bold; see Figure 2, section 2). Dashed lines correspond to average subject mean accuracy by lotteries' EV. The solid line is a naive forecast that achieves an accuracy of 54% on the test data by predicting all records as not-playing.

Figure A32 Out-of-sample accuracy results for SVM for A



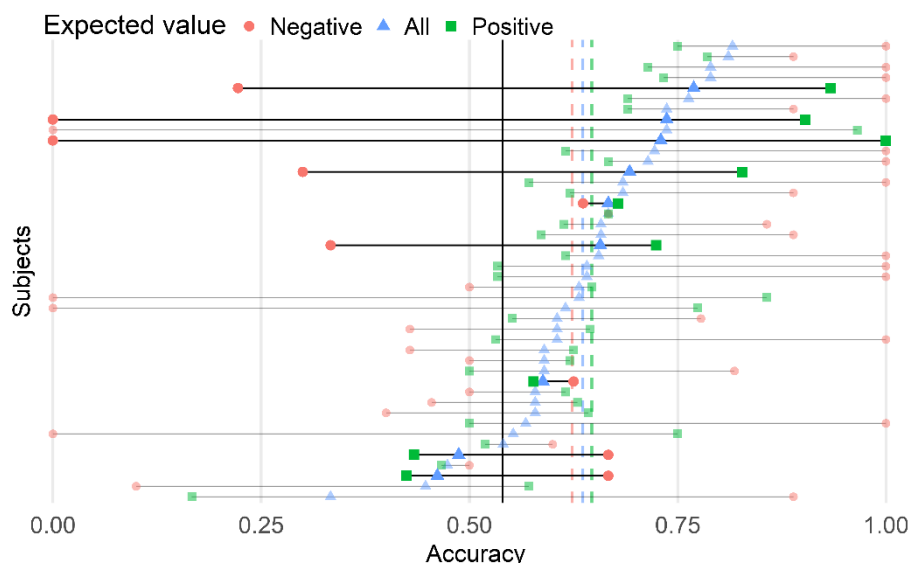
Notes: Out-of-sample accuracy for 44 subjects by positive expected value (PEV), negative expected value (NEV), and all lotteries (All) for playing a 50/50 gamble by an ANN model on the basis of predictors that relate to attention (A) choice-process data. The 9 subjects that play five or more NEV lotteries (NEVL) with respect to 45 NEVL that are included in the full experiment data are highlighted (bold; see Figure 2, section 2). Dashed lines correspond to average subject mean accuracy by lotteries' EV. The solid line is a naive forecast that achieves an accuracy of 54% on the test data by predicting all records as not-playing.

Figure A33 Out-of-sample accuracy results for ANN for A



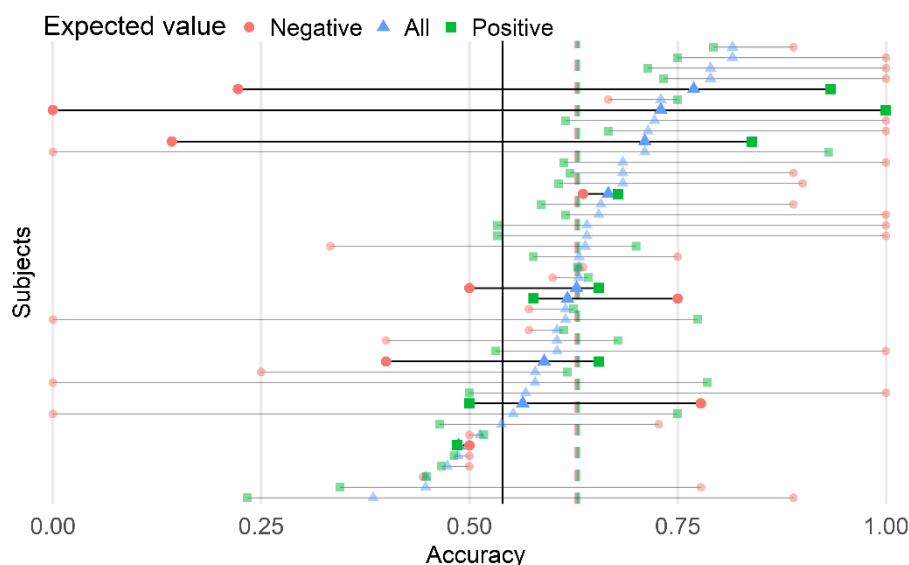
Notes: Out-of-sample accuracy for 44 subjects by positive expected value (PEV), negative expected value (NEV), and all lotteries (All) for playing a 50/50 gamble by an ANN model on the basis of predictors that relate to attention (A) choice-process data. The 9 subjects that play five or more NEV lotteries (NEVL) with respect to 45 NEVL that are included in the full experiment data are highlighted (bold; see Figure 2, section 2). Dashed lines correspond to average subject mean accuracy by lotteries' EV. The solid line is a naive forecast that achieves an accuracy of 54% on the test data by predicting all records as not-playing.

Figure A34 Out-of-sample accuracy results for RF for *A*



Notes: Out-of-sample accuracy for 44 subjects by positive expected value (PEV), negative expected value (NEV), and all lotteries (All) for playing a 50/50 gamble by an RF model on the basis of predictors that relate to attention (*A*) choice-process data. The 9 subjects that play five or more NEV lotteries (NEVL) with respect to 45 NEVL that are included in the full experiment data are highlighted (bold; see Figure 2, section 2). Dashed lines correspond to average subject mean accuracy by lotteries' EV. The solid line is a naive forecast that achieves an accuracy of 54% on the test data by predicting all records as not-playing.

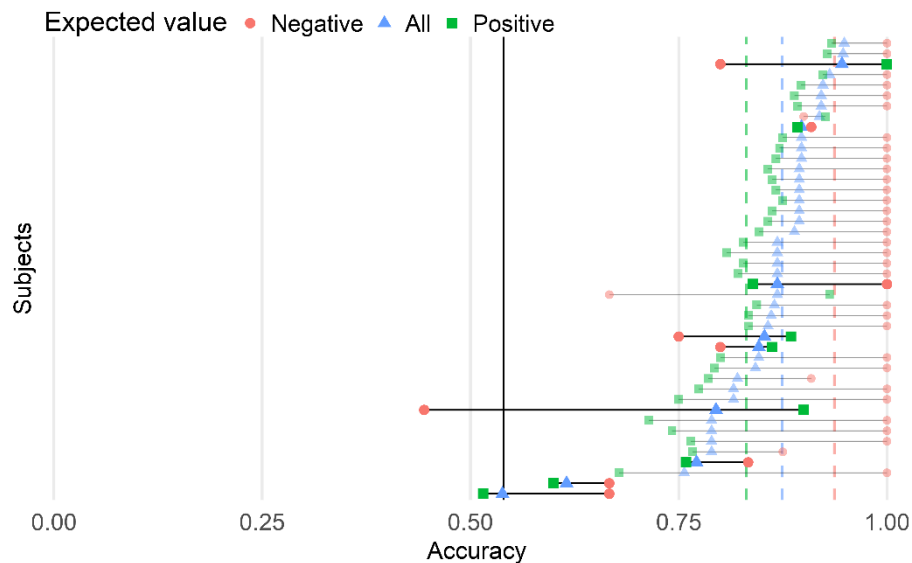
Figure A35 Out-of-sample accuracy results for GBM for *A*



Notes: Out-of-sample accuracy for 44 subjects by positive expected value (PEV), negative expected value (NEV), and all lotteries (All) for playing a 50/50 gamble by a GBM model on the basis of predictors that relate to attention (*A*) choice-process data. The 9 subjects that play five or more NEV lotteries (NEVL) with respect to 45 NEVL that are included in the full experiment data are highlighted (bold; see Figure 2, section 2). Dashed lines correspond to average subject mean accuracy by lotteries' EV. The solid line is a naive forecast that achieves an accuracy of 54% on the test data by predicting all records as not-playing.

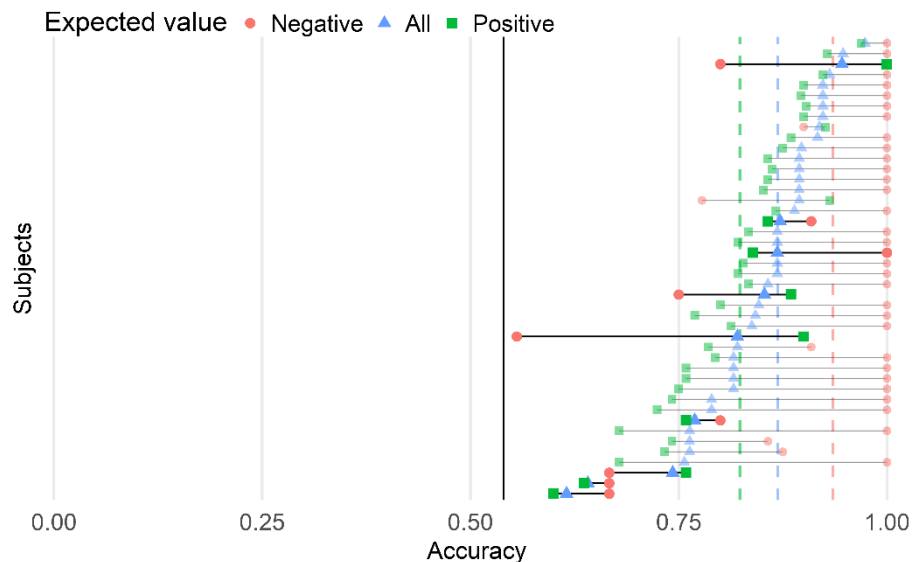
3.5.3 Lottery design, socioeconomic and past gambling characteristics

Figure A36 Out-of-sample accuracy results for the logistic regression model for *LSG*



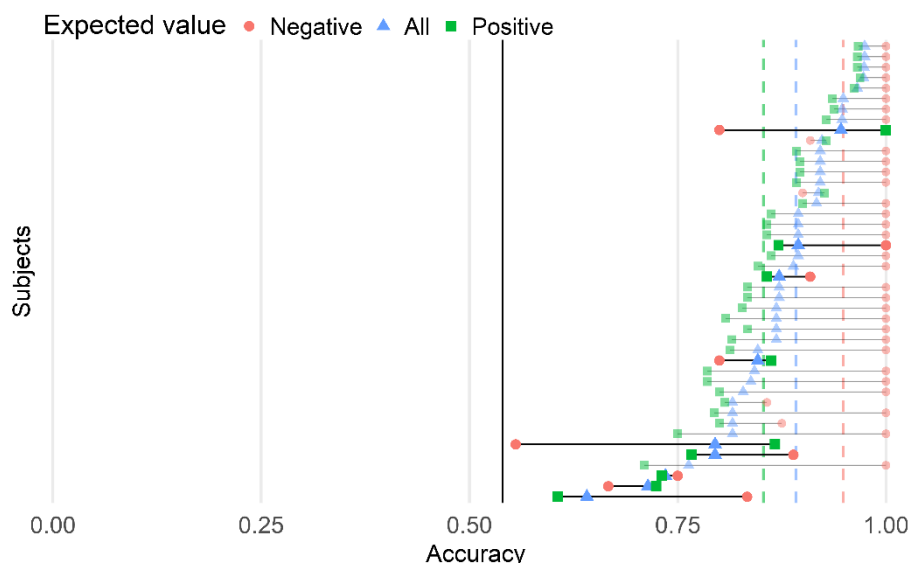
Notes: Out-of-sample accuracy for 44 subjects by positive expected value (PEV), negative expected value (NEV), and all lotteries (All) for playing a 50/50 gamble by a logistic regression model on the basis of information that is related to lottery design variables, socioeconomic characteristics, and past gambling behavior (*LSG*). Subjects that play five or more NEVL with respect to the full experiment data are highlighted (bold; see Figure 2, section 2). Dashed lines correspond to average subject mean accuracy by lotteries' EV. The solid line is a naive forecast that achieves an accuracy of 54% on the test data by predicting all records as not-playing.

Figure A37 Out-of-sample accuracy results for the elastic regression model for *LSG*



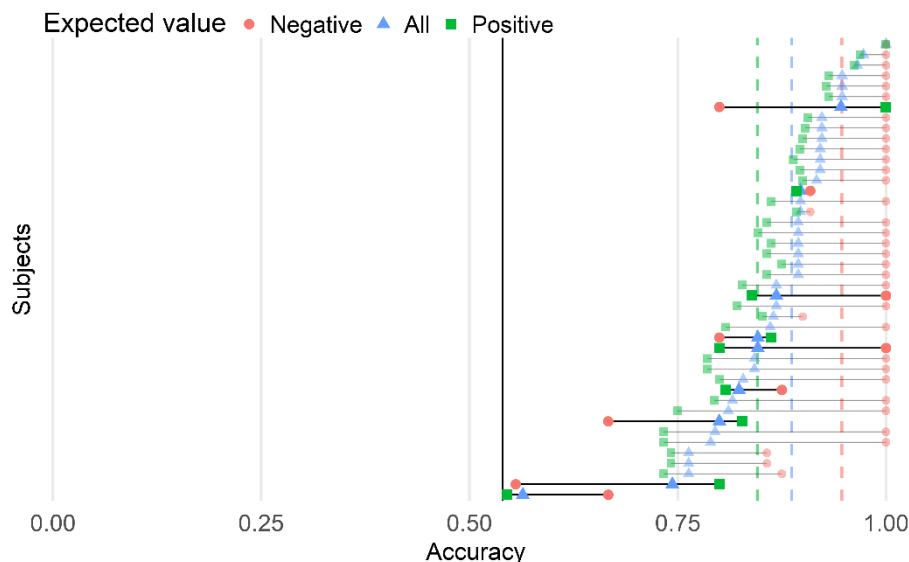
Notes: Out-of-sample accuracy for 44 subjects by positive expected value (PEV), negative expected value (NEV), and all lotteries (All) for playing a 50/50 gamble by an Elastic regression model on the basis of information that is related to lottery design variables, socioeconomic characteristics, and past gambling behavior (*LSG*). Subjects that play five or more NEVL with respect to the full experiment data are highlighted (bold; see Figure 2, section 2). Dashed lines correspond to average subject mean accuracy by lotteries' EV. The solid line is a naive forecast that achieves an accuracy of 54% on the test data by predicting all records as not-playing.

Figure A38 Out-of-sample accuracy results for SVM for *LSG*



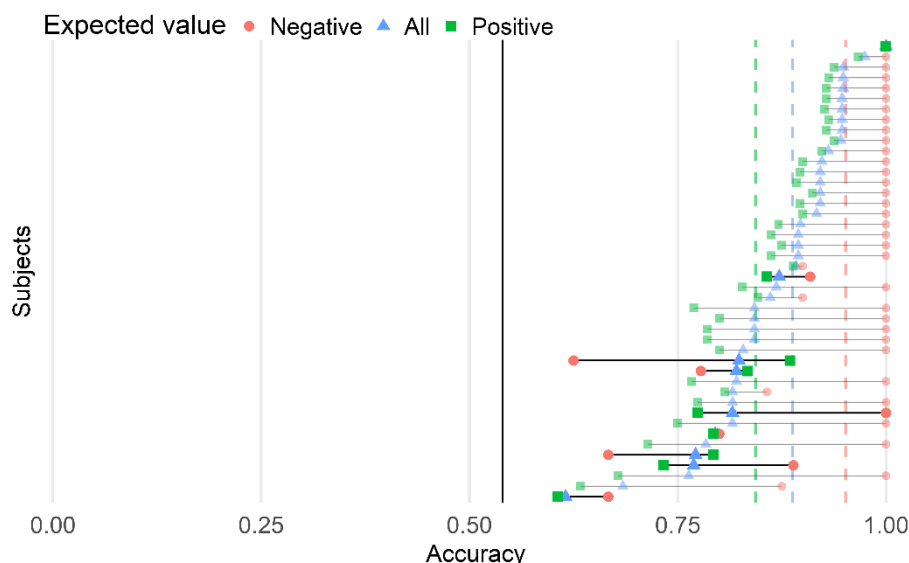
Notes: Out-of-sample accuracy for 44 subjects by positive expected value (PEV), negative expected value (NEV), and all lotteries (All) for playing a 50/50 gamble by an SVM model on the basis of information that is related to lottery design variables, socioeconomic characteristics, and past gambling behavior (*LSG*). Subjects that play five or more NEVL with respect to the full experiment data are highlighted (bold; see Figure 2, section 2). Dashed lines correspond to average subject mean accuracy by lotteries' EV. The solid line is a naive forecast that achieves an accuracy of 54% on the test data by predicting all records as not-playing.

Figure A39 Out-of-sample accuracy results for ANN for *LSG*



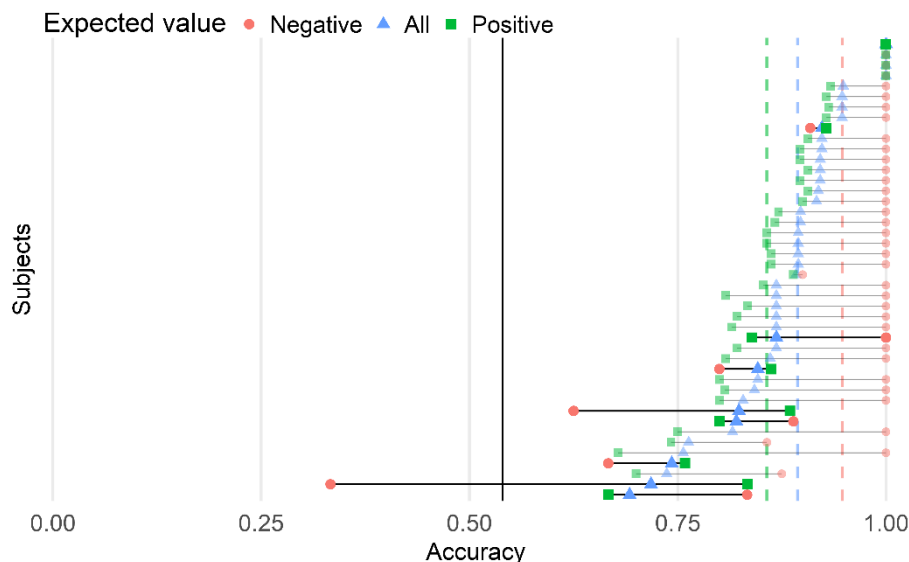
Notes: Out-of-sample accuracy for 44 subjects by positive expected value (PEV), negative expected value (NEV), and all lotteries (All) for playing a 50/50 gamble by an ANN model on the basis of information that is related to lottery design variables, socioeconomic characteristics, and past gambling behavior (*LSG*). Subjects that play five or more NEVL with respect to the full experiment data are highlighted (bold; see Figure 2, section 2). Dashed lines correspond to average subject mean accuracy by lotteries' EV. The solid line is a naive forecast that achieves an accuracy of 54% on the test data by predicting all records as not-playing.

Figure A40 Out-of-sample accuracy results for RF for *LSG*



Notes: Out-of-sample accuracy for 44 subjects by positive expected value (PEV), negative expected value (NEV), and all lotteries (All) for playing a 50/50 gamble by an RF model on the basis of information that is related to lottery design variables, socioeconomic characteristics, and past gambling behavior (*LSG*). Subjects that play five or more NEVL with respect to the full experiment data are highlighted (bold; see Figure 2, section 2). Dashed lines correspond to average subject mean accuracy by lotteries' EV. The solid line is a naive forecast that achieves an accuracy of 54% on the test data by predicting all records as not-playing.

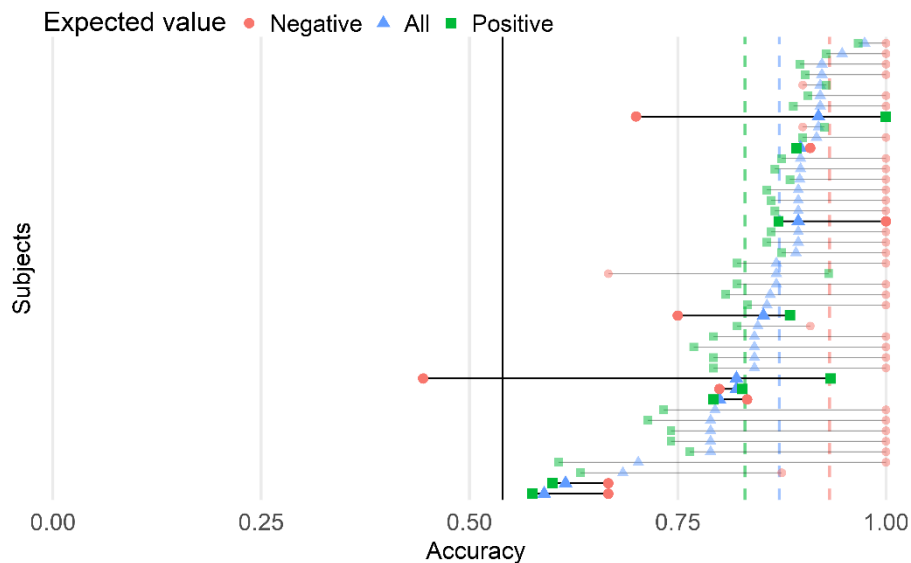
Figure A41 Out-of-sample accuracy results for GBM for *LSG*



Notes: Out-of-sample accuracy for 44 subjects by positive expected value (PEV), negative expected value (NEV), and all lotteries (All) for playing a 50/50 gamble by a GBM model on the basis of information that is related to lottery design variables, socioeconomic characteristics, and past gambling behavior (*LSG*). Subjects that play five or more NEVL with respect to the full experiment data are highlighted (bold; see Figure 2, section 2). Dashed lines correspond to average subject mean accuracy by lotteries' EV. The solid line is a naive forecast that achieves an accuracy of 54% on the test data by predicting all records as not-playing.

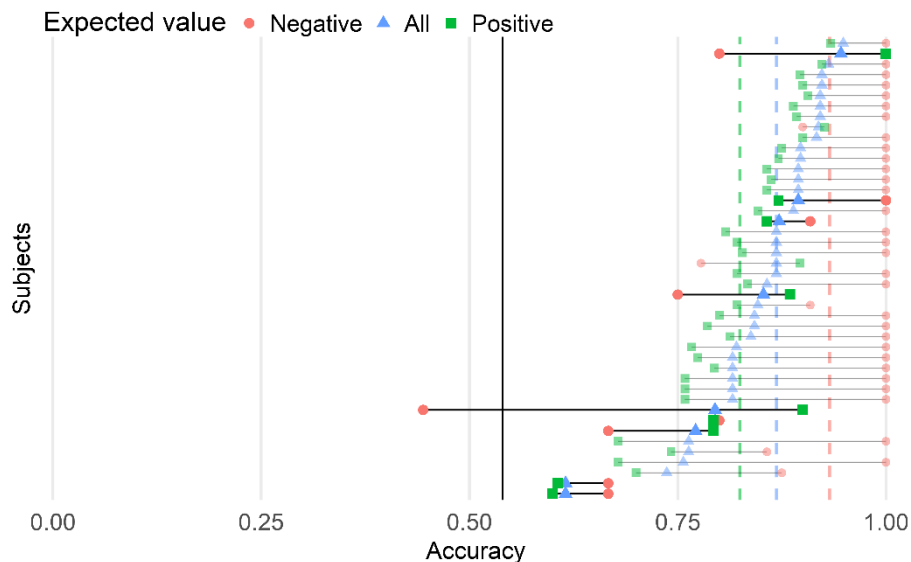
3.5.3 Lottery, socioeconomic, gambling, psychological, and attention data (LSGPA)

Figure A42 Out-of-sample accuracy results for the logistic regression model for *LSGPA*



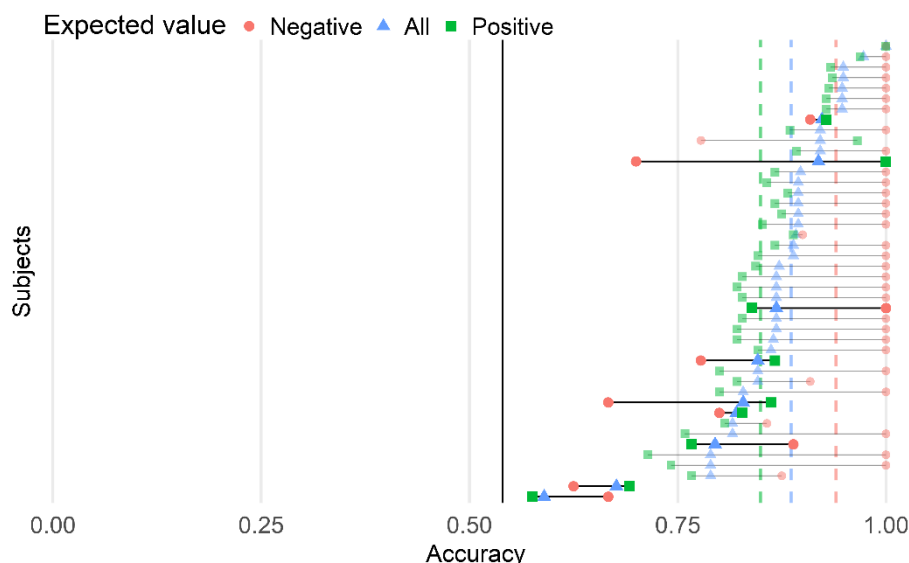
Notes: Out-of-sample accuracy for 44 subjects by positive expected value (PEV), negative expected value (NEV), and all lotteries (All) for playing a 50/50 gamble by a logistic regression model on the basis of information that is related to lottery design variables, socioeconomic characteristics, and past gambling behavior (*LSG*), information on psycho-physiological reactions (*P*) and attention data (*A*). Subjects that play five or more NEVL with respect to the full experiment data are highlighted. Dashed lines correspond to average subject mean accuracy by lotteries' EV. The solid line is a naive forecast that achieves an accuracy of 54% on the test data by predicting all records as not-playing.

Figure A43 Out-of-sample accuracy results for the elastic regression model for *LSGPA*



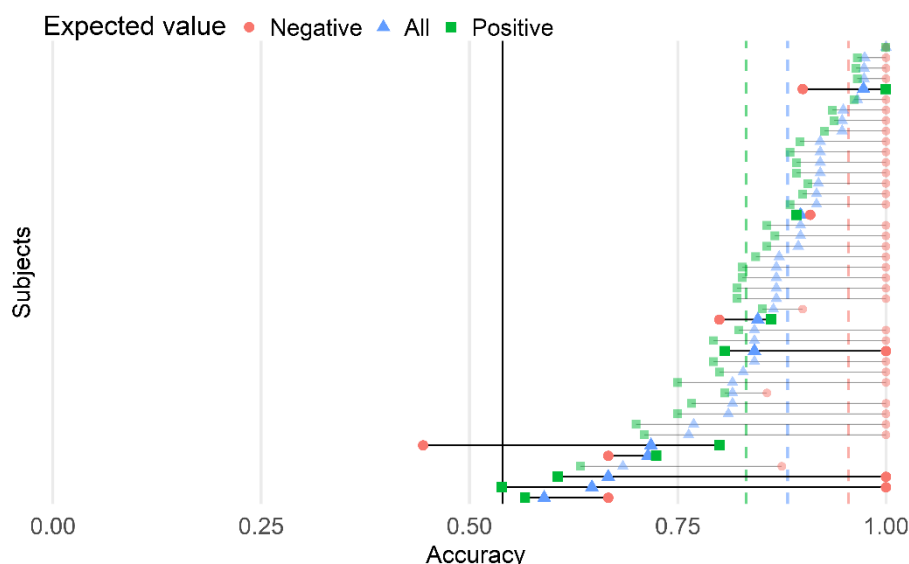
Notes: Out-of-sample accuracy for 44 subjects by positive expected value (PEV), negative expected value (NEV), and all lotteries (All) for playing a 50/50 gamble by an Elastic regression model on the basis of information that is related to lottery design variables, socioeconomic characteristics, and past gambling behavior (*LSG*), information on psycho-physiological reactions (*P*) and attention data (*A*). The 9 subjects that play five or more NEV lotteries (NEVL) with respect to 45 NEVL that are included in the full experiment data are highlighted. Dashed lines correspond to average subject mean accuracy by lotteries' EV. The solid line is a naive forecast that achieves an accuracy of 54%.

Figure A44 Out-of-sample accuracy results for SVM for *LSGPA*



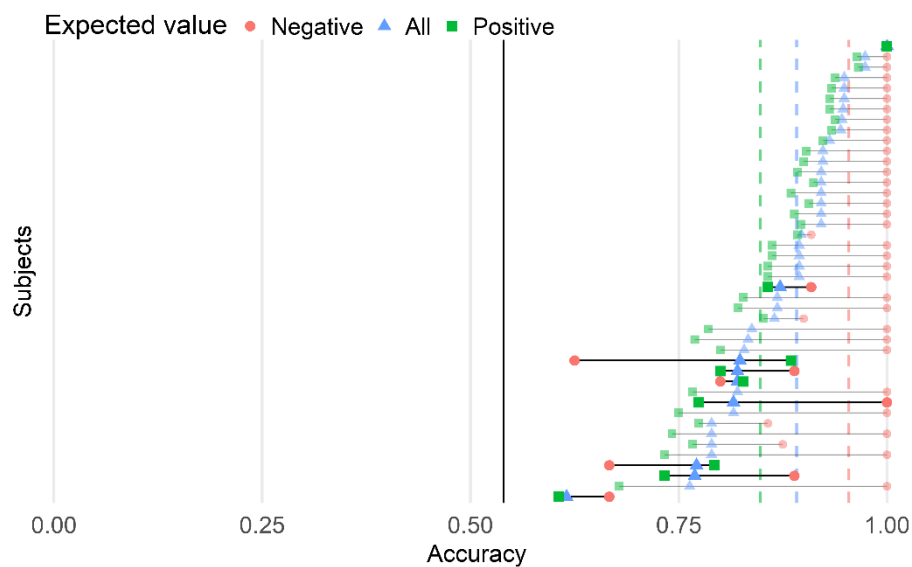
Notes: Out-of-sample accuracy for 44 subjects by positive expected value (PEV), negative expected value (NEV), and all lotteries (All) for playing a 50/50 gamble by an SVM model on the basis of information that is related to lottery design variables, socioeconomic characteristics, and past gambling behavior (*LSG*), information on psychophysiological reactions (*P*) and attention data (*A*). The 9 subjects that play five or more NEV lotteries (NEVL) with respect to 45 NEVL that are included in the full experiment data are highlighted (bold; see Figure 2, section 2). Dashed lines correspond to average subject mean accuracy by lotteries' EV. The solid line is a naive forecast that achieves an accuracy of 54% on the test data by predicting all records as not-playing.

Figure A45 Out-of-sample accuracy results for ANN for *LSGPA*



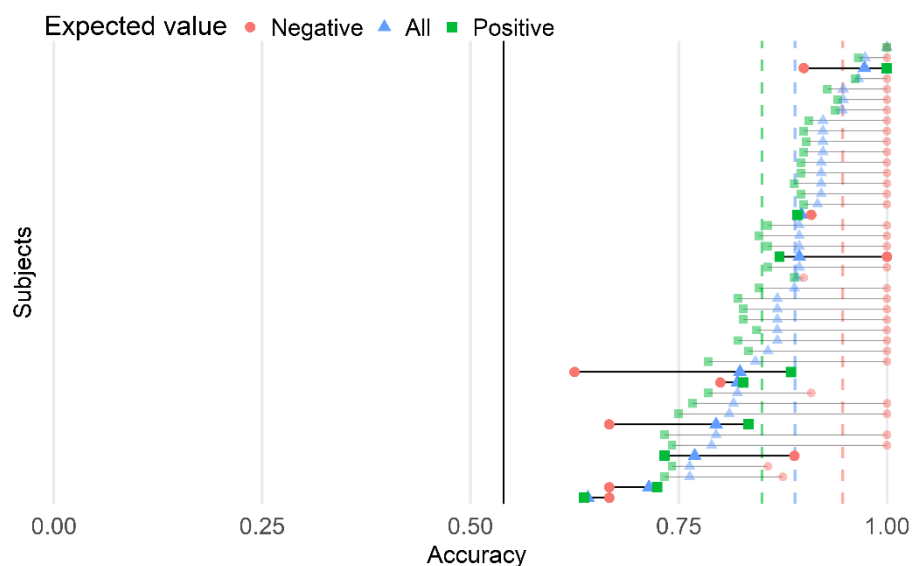
Notes: Out-of-sample accuracy for 44 subjects by positive expected value (PEV), negative expected value (NEV), and all lotteries (All) for playing a 50/50 gamble by an ANN model on the basis of information that is related to lottery design variables, socioeconomic characteristics, and past gambling behavior (*LSG*), information on psychophysiological reactions (*P*) and attention data (*A*). The 9 subjects that play five or more NEV lotteries (NEVL) with respect to 45 NEVL that are included in the full experiment data are highlighted (bold; see Figure 2, section 2). Dashed lines correspond to average subject mean accuracy by lotteries' EV. The solid line is a naive forecast that achieves an accuracy of 54% on the test data by predicting all records as not-playing.

Figure A46 Out-of-sample accuracy results for RF for *LSGPA*



Notes: Out-of-sample accuracy for 44 subjects by positive expected value (PEV), negative expected value (NEV), and all lotteries (All) for playing a 50/50 gamble by an RF model on the basis of information that is related to lottery design variables, socioeconomic characteristics, and past gambling behavior (*LSG*), information on psychophysiological reactions (*P*) and attention data (*A*). The 9 subjects that play five or more NEV lotteries (NEVL) with respect to 45 NEVL that are included in the full experiment data are highlighted (bold; see Figure 2, section 2). Dashed lines correspond to average subject mean accuracy by lotteries' EV. The solid line is a naive forecast that achieves an accuracy of 54% on the test data by predicting all records as not-playing.

Figure A47 Out-of-sample accuracy results for GBM for *LSGPA*



Notes: Out-of-sample accuracy for 44 subjects by positive expected value (PEV), negative expected value (NEV), and all lotteries (All) for playing a 50/50 gamble by a GBM model on the basis of information that is related to lottery design variables, socioeconomic characteristics, and past gambling behavior (*LSG*), information on psychophysiological reactions (*P*) and attention data (*A*). The 9 subjects that play five or more NEV lotteries (NEVL) with respect to 45 NEVL that are included in the full experiment data are highlighted (bold; see Figure 2, section 2). Dashed lines correspond to average subject mean accuracy by lotteries' EV. The solid line is a naive forecast that achieves an accuracy of 54% on the test data by predicting all records as not-playing.

References

- Benedek, M., & Kaernbach, C. (2010). A continuous measure of phasic electrodermal activity. *Journal of Neuroscience Methods*, 190(1), 80–91.
- Calaway, R., Corporation, M., Weston, S., & Tenenbaum, D. (2017). “doParallel”: foreach parallel adaptor for the parallel package. *R Package Version 1.0.11*.
- Friedman, J., Hastie, T., & Tibshirani, R. (2010). Regularization Paths for Generalized Linear Models via Coordinate Descent. *Journal of Statistical Software*, 3(1), 1–22. *R Package Version 2.1.4*.
- Greenwell, B., Boehmke, B., Cunningham, J., & Developers, G. (2018). gbm: Generalized Boosted Regression Models. *R Package Version 2.1.4*.
- Gu, Z., Eils, R., & Schlesner, M. (2016). Complex heatmaps reveal patterns and correlations in multidimensional genomic data. *Bioinformatics*, 32(18), 2847–2849.
- Gu, Z., Gu, L., Eils, R., Schlesner, M., & Brors, B. (2014). Circlize implements and enhances circular visualization in R. *Bioinformatics. R Package Version 1.99.0*.
- Karatzoglou, A., Smola, A., Hornik, K., & Zeileis, A. (2004). kernlab - An S4 Package for Kernel Methods in R. *Journal of Statistical Software*, 11(9), 1–20. *R Package Version 1.99.0*.
- Kuhn, M. (2018). caret: Classification and Regression Training. *Contributions from Jed Wing, Steve Weston, Andre Williams, Chris Keefer, Allan Engelhardt, Tony Cooper, Zachary Mayer, Brenton Kenkel, the R Core Team, Michael Benesty, Reynald Lescarbeau, Andrew Ziem, Luca Scrucca, Yuan Tang, Can Candan and Tyler Hunt. R package version 6.0-80*.
- Neuwirth, E. (2014). RColorBrewer: ColorBrewer palettes. *R Package Version 1.1-2*.
- R Core Team. (2018). R: A Language and Environment for Statistical Computing. *R Foundation for Statistical Computing (Version 3.5.1)*.
- Sarkar, D. (2008). Lattice multivariate data visualization with R. *Use R!*, xvii, p. 265. *R Package Version 0.20-35*.
- Venables, W. N., & Ripley, B. D. (2002). *Modern Applied Statistics with S. Fourth Edition*. New York: Springer. *R Package Version 7.3-12*.

Wickham, H. (2016). *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag, New York. *Package Version 3.0.0*.

Wickham, H. (2017). tidyverse: Easily Install and Load the “Tidyverse.” *R Package Version 1.2.1. R Package Version 1.2.1*.

Wickham, H., Francois, R., Henry, L., & Müller, K. (2018). dplyr: A Grammar of Data Manipulation. *R Package Version 0.7.6*.

Wright, M. N., & Ziegler, A. (2017). A Fast Implementation of Random Forests for High Dimensional Data in C++ and R. *Journal of Statistical Software*, 77(1), 1–17. *R Package Version 0.1.10*

Hamburg Contemporary Economic Discussions

(Download: <https://www.wiso.uni-hamburg.de/en/fachbereich-vwl/professuren/maennig/research/hceds.html>)

- 66 MUELLER, S. Q. / RING, P. / SCHMIDT, M.: Forecasting economic decisions under risk: The predictive importance of choice-process data, 2019.
- 65 MUELLER, S. Q.: Pre- and within-season attendance forecasting in Major League Baseball: A random forest approach, 2018.
- 64 KRUSE, F. K. / MAENNIG, W.: Suspension by choice – determinants and asymmetries, 2018.
- 63 GROTHE, H. / MAENNIG, W.: A 100-million-dollar fine for Russia's doping policy? A billion-dollar penalty would be more correct! Millionenstrafe für Russlands Doping-Politik? Eine Milliarden-Strafe wäre richtiger! 2017.
- 62 MAENNIG, W., / SATTARHOFF, C. / STAHLECKER, P.: Interpretation und mögliche Ursachen statistisch insignikanter Testergebnisse - eine Fallstudie zu den Beschäftigungseffekten der Fußball-Weltmeisterschaft 2006, 2017.
- 61 KRUSE, F. K. / MAENNIG, W.: The future development of world records, 2017.
- 60 MAENNIG, W.: Governance in Sports Organizations, 2017.
- 59 AHLFELDT, G. M. / MAENNIG, W. / FELIX J. RICHTER: Zoning in reunified Berlin, 2017.
- 58 MAENNIG, W.: Major Sports Events: Economic Impact, 2017.
- 57 MAENNIG, W.: Public Referenda and Public Opinion on Olympic Games, 2017.
- 56 MAENNIG, W. / WELLBROCK, C.: Rio 2016: Sozioökonomische Projektion des Olympischen Medaillenrankings, 2016.
- 55 MAENNIG, W. / VIERHAUS, C.: Which countries bid for the Olympic Games? Economic, political, and social factors and chances of winning, 2016.
- 54 AHLFELDT, G. M. / MAENNIG, W. / STEENBECK, M.: Après nous le déluge? Direct democracy and intergenerational conflicts in aging societies, 2016.
- 53 LANGER, V. C. E.: Good news about news shocks, 2015.
- 52 LANGER, V. C. E. / MAENNIG, W. / RICHTER, F. J.: News Shocks in the Data: Olympic Games and their Macroeconomic Effects – Reply, 2015.
- 51 MAENNIG, W.: Ensuring Good Governance and Preventing Corruption in the Planning of Major Sporting Events – Open Issues, 2015.

Hamburg Contemporary Economic Discussions

(Download: <https://www.wiso.uni-hamburg.de/en/fachbereich-vwl/professuren/maennig/research/hceds.html>)

- 50 MAENNIG, W. / VIERHAUS, C.: Who Wins Olympic Bids? 2015 (3rd version).
- 49 AHLFELDT, G. M. / MAENNIG, W. / RICHTER, F.: Urban Renewal after the Berlin Wall, 2013.
- 48 BRANDT, S. / MAENNIG, W. / RICHTER, F.: Do Places of Worship Affect Housing Prices? Evidence from Germany, 2013.
- 47 ARAGÃO, T. / MAENNIG, W.: Mega Sporting Events, Real Estate, and Urban Social Economics – The Case of Brazil 2014/2016, 2013.
- 46 MAENNIG, W. / STEENBECK, M. / WILHELM, M.: Rhythms and Cycles in Happiness, 2013.
- 45 RICHTER, F. / STEENBECK, M. / WILHELM, M.: The Fukushima Accident and Policy Implications: Notes on Public Perception in Germany, 2014 (2nd version).
- 44 MAENNIG, W.: London 2012 – das Ende des Mythos vom erfolgreichen Sportsoldaten, 2012.
- 43 MAENNIG, W. / WELLBROCK, C.: London 2012 – Medal Projection – Medaillenvorausberechnung, 2012.
- 42 MAENNIG, W. / RICHTER, F.: Exports and Olympic Games: Is there a Signal Effect? 2012.
- 41 MAENNIG, W. / WILHELM, M.: Becoming (Un)employed and Life Satisfaction: Asymmetric Effects and Potential Omitted Variable Bias in Empirical Happiness Studies, 2011.
- 40 MAENNIG, W.: Monument Protection and Zoning in Germany: Regulations and Public Support from an International Perspective, 2011.
- 39 BRANDT, S. / MAENNIG, W.: Perceived Externalities of Cell Phone Base Stations – The Case of Property Prices in Hamburg, Germany, 2011.
- 38 MAENNIG, W. / STOBERNACK, M.: Do Men Slow Down Faster than Women? 2010.
- 37 DU PLESSIS, S. A. / MAENNIG, W.: The 2010 World Cup High-frequency Data Economics: Effects on International Awareness and (Self-defeating) Tourism, 2010.

Hamburg Contemporary Economic Discussions

(Download: <https://www.wiso.uni-hamburg.de/en/fachbereich-vwl/professuren/maennig/research/hceds.html>)

- 36 BISCHOFF, O.: Explaining Regional Variation in Equilibrium Real Estate Prices and Income, 2010.
- 35 FEDDERSEN, A. / MAENNIG, W.: Mega-Events and Sectoral Employment: The Case of the 1996 Olympic Games, 2010.
- 34 FISCHER, J.A.V. / SOUSA-POZA, A.: The Impact of Institutions on Firms Rejuvenation Policies: Early Retirement with Severance Pay versus Simple Lay-Off. A Cross-European Analysis, 2010.
- 33 FEDDERSEN, A. / MAENNIG, W.: Sectoral Labor Market Effects of the 2006 FIFA World Cup, 2010.
- 32 AHLFELDT, G.: Blessing or Curse? Appreciation, Amenities, and Resistance around the Berlin “Mediaspree”, 2010.
- 31 FALCH, T. / FISCHER, J.A.V.: Public Sector Decentralization and School Performance: International Evidence, 2010.
- 30 AHLFELDT, G. / MAENNIG, W. / ÖLSCHLÄGER, M.: Lifestyles and Preferences for (Public) Goods: Professional Football in Munich, 2009.
- 29 FEDDERSEN, A. / JACOBSEN, S. / MAENNIG, W.: Sports Heroes and Mass Sports Participation – The (Double) Paradox of the “German Tennis Boom”, 2009.
- 28 AHLFELDT, G. / MAENNIG, W. / OSTERHEIDER, T.: Regional and Sectoral Effects of a Common Monetary Policy: Evidence from Euro Referenda in Denmark and Sweden, 2009.
- 27 BJØRNSKOV, C. / DREHER, A. / FISCHER, J.A.V. / SCHNELLENBACH, J.: On the Relation Between Income Inequality and Happiness: Do Fairness Perceptions Matter? 2009.
- 26 AHLFELDT, G. / MAENNIG, W.: Impact of Non-Smoking Ordinances on Hospitality Revenues: The Case of Germany, 2009.
- 25 FEDDERSEN, A. / MAENNIG, W.: Wage and Employment Effects of the Olympic Games in Atlanta 1996 Reconsidered, 2009.

Hamburg Contemporary Economic Discussions

(Download: <https://www.wiso.uni-hamburg.de/en/fachbereich-vwl/professuren/maennig/research/hceds.html>)

- 24 AHLFELDT, G. / FRANKE, B. / MAENNIG, W.: Terrorism and the Regional and Religious Risk Perception of Foreigners: The Case of German Tourists, 2009.
- 23 AHLFELDT, G. / WENDLAND, N.: Fifty Years of Urban Accessibility: The Impact of Urban Railway Network on the Land Gradient in Industrializing Berlin, 2008.
- 22 AHLFELDT, G. / FEDDERSEN, A.: Determinants of Spatial Weights in Spatial Wage Equations: A Sensitivity Analysis, 2008.
- 21 MAENNIG, W. / ALLMERS, S.: South Africa 2010: Economic Scope and Limits, 2008.
- 20 MAENNIG, W. / WELLBROCK, C.-M.: Sozio-ökonomische Schätzungen Olympischer Medaillengewinne: Analyse-, Prognose- und Benchmarkmöglichkeiten, 2008.
- 19 AHLFELDT, G.: The Train has Left the Station: Real Estate Price Effects of Mainline Realignment in Berlin, 2008.
- 18 MAENNIG, W. / PORSCHE, M.: The Feel-good Effect at Mega Sport Events – Recommendations for Public and Private Administration Informed by the Experience of the FIFA World Cup 2006, 2008.
- 17 AHLFELDT, G. / MAENNIG, W.: Monumental Protection: Internal and External Price Effects, 2008.
- 16 FEDDERSEN, A. / GRÖTZINGER, A. / MAENNIG, W.: New Stadia and Regional Economic Development – Evidence from FIFA World Cup 2006 Stadia, 2008.
- 15 AHLFELDT, G. / FEDDERSEN, A.: Geography of a Sports Metropolis, 2007.
- 14 FEDDERSEN, A. / MAENNIG, W.: Arenas vs. Multifunctional Stadia – Which Do Spectators Prefer? 2007.
- 13 AHLFELDT, G.: A New Central Station for a Unified City: Predicting Impact on Property Prices for Urban Railway Network Extension, 2007.
- 12 AHLFELDT, G.: If Alonso was Right: Accessibility as Determinant for Attractiveness of Urban Location, 2007.

Hamburg Contemporary Economic Discussions

(Download: <https://www.wiso.uni-hamburg.de/en/fachbereich-vwl/professuren/maennig/research/hceds.html>)

- 11 AHLFELDT, G., MAENNIG, W.: Assessing External Effects of City Airports: Land Values in Berlin, 2007.
- 10 MAENNIG, W.: One Year Later: A Re-Appraisal of the Economics of the 2006 Soccer World Cup, 2007.
- 09 HAGN, F. / MAENNIG, W.: Employment Effects of the World Cup 1974 in Germany.
- 08 HAGN, F. / MAENNIG W.: Labour Market Effects of the 2006 Soccer World Cup in Germany, 2007.
- 07 JASMAND, S. / MAENNIG, W.: Regional Income and Employment Effects of the 1972 Munich Olympic Summer Games, 2007.
- 06 DUST, L. / MAENNIG, W.: Shrinking and Growing Metropolitan Areas – Asymmetric Real Estate Price Reactions? The Case of German Single-family Houses, 2007.
- 05 HEYNE, M. / MAENNIG, W. / SUESSMUTH, B.: Mega-sporting Events as Experience Goods, 2007.
- 04 DU PLESSIS, S. / MAENNIG, W.: World Cup 2010: South African Economic Perspectives and Policy Challenges Informed by the Experience of Germany 2006, 2007.
- 03 AHLFELDT, G. / MAENNIG, W.: The Impact of Sports Arenas on Land Values: Evidence from Berlin, 2007.
- 02 FEDDERSEN, A. / MAENNIG, W. / ZIMMERMANN, P.: How to Win the Olympic Games – The Empirics of Key Success Factors of Olympic Bids, 2007.
- 01 AHLFELDT, G. / MAENNIG, W.: The Role of Architecture on Urban Revitalization: The Case of “Olympic Arenas” in Berlin-Prenzlauer Berg, 2007.
- 04/2006 MAENNIG, W. / SCHWARTHOFF, F.: Stadium Architecture and Regional Economic Development: International Experience and the Plans of Durban, October 2006.

Hamburg Contemporary Economic Discussions

(Download: <https://www.wiso.uni-hamburg.de/en/fachbereich-vwl/professuren/maennig/research/hceds.html>)

- 03/2006 FEDDERSEN, A. / VÖPEL, H.: Staatliche Hilfen für Profifußballclubs in finanziellen Notlagen? – Die Kommunen im Konflikt zwischen Imageeffekten und Moral-Hazard-Problemen, September 2006.
- 02/2006 FEDDERSEN, A.: Measuring Between-season Competitive Balance with Markov Chains, July 2006.
- 01/2006 FEDDERSEN, A.: Economic Consequences of the UEFA Champions League for National Championships – The Case of Germany, May 2006.
- 04/2005 BUETTNER, N. / MAENNIG, W. / MENSSEN, M.: Zur Ableitung einfacher Multiplikatoren für die Planung von Infrastrukturkosten anhand der Aufwendungen für Sportstätten – eine Untersuchung anhand der Fußball-WM 2006, May 2005.
- 03/2005 SIEVERS, T.: A Vector-based Approach to Modeling Knowledge in Economics, February 2005.
- 02/2005 SIEVERS, T.: Information-driven Clustering – An Alternative to the Knowledge Spillover Story, February 2005.
- 01/2005 FEDDERSEN, A. / MAENNIG, W.: Trends in Competitive Balance: Is there Evidence for Growing Imbalance in Professional Sport Leagues? January 2005.

ISSN 1865-2441 (PRINT)
ISSN 1865-7133 (ONLINE)

ISBN 978-3-942820-46-2 (PRINT)
ISBN 978-3-942820-47-9 (ONLINE)

Ha mbur

Contemporary Economic Discussions