WORKING PAPER SERIES

Experimental economics for machine learning - a methodological contribution

Dmitri Bershadskyy / Laslo Dinges / Marc-André Fiedler / Ayoub Al-Hamadi / Nina Ostermaier / Joachim Weimann

Working Paper No. 08/2023

OTTO VON GUERICKE UNIVERSITÄT MAGDEBURG

FACULTY OF ECONOMICS AND MANAGEMENT

Impressum (§ 5 TMG)

Herausgeber:

Otto-von-Guericke-Universität Magdeburg
Fakultät für Wirtschaftswissenschaft
Der Dekan

Verantwortlich für diese Ausgabe:

D. Bershadskyy, L. Dinges, M.-A. Fiedler
Otto-von-Guericke-Universität Magdeburg
Fakultät für Wirtschaftswissenschaft
Postfach 4120
39016 Magdeburg
Germany

http://www.fww.ovgu.de/femm

Bezug über den Herausgeber

ISSN 1615-4274

Experimental economics for machine learning - a methodological contribution

Dmitri Bershadskyy[a1], Laslo Dinges[b], Marc-André Fiedler[b], Ayoub Al-Hamadi[b], Nina Ostermaier[a], Joachim Weimann[a]

Abstract:

In this paper, we investigate how technology has contributed to experimental economics in the past and illustrate how experimental economics can contribute to technological progress in the future. We argue that with machine learning (ML) a new technology is at hand, where for the first time experimental economics can contribute to enabling substantial improvement of technology. At the same time, ML opens up new questions for experimental research because it can generate observations that were previously impossible. To demonstrate this, we focus on algorithms trained to detect lies. Such algorithms are of high relevance for research in economics as they deal with the ability to retrieve otherwise private information. We deduce that most of the commonly applied data sets for the training of lie detection algorithms could be improved by applying the toolbox of experimental economics. To illustrate this, we replicate the "lies in disguise-experiment" (Fischbacher & Föllmi-Heusi, 2013) with a modification regarding monitoring. The modified setup guarantees a certain level of privacy from the experimenter yet allows to record the subjects as they lie to the camera. Our results indicate the same lying behavior as in the original experiment despite monitoring. Yet, our experiment allows for an individual-level analysis and provides a video data set that can be used for lie detection algorithms.

Keywords: lying behavior, lie detection, experiment, technology, machine learning
JEL Codes: C90, C91, O30

---

[1] Corresponding author.
Dmitri Bershadskyy: ORCID: 0000-0002-9856-1548
dmitri.bershadskyy@ovgu.de

Marc-André Fiedler
marc-andre.fiedler@ovgu.de

Nina Ostermaier
nina.ostermaier@ovgu.de

Ayoub Al-Hamadi
ayoub.al-hamadi@ovgu.de

Joachim Weimann
joachim.weimann@ovgu.de

[a] Faculty of Economics and Management, Otto-von-Guericke University Magdeburg, Germany
[b] Faculty of Electrical Engineering and Information Technology, Otto-von-Guericke University Magdeburg, Germany
The authors gratefully acknowledge financial support by Deutsche Forschungsgemeinschaft (DFG) through project number 468478819. We have no conflict of interest to disclose.

1. Introduction

In a lot of experimental disciplines, scientific knowledge and technological progress are in a reciprocal exchange relationship that leads to a significant acceleration of both knowledge acquisition and technology development. Findings from basic research lead to the development of new, improved technology, which in turn can be used as an instrument for basic research. This reciprocal acceleration process is well-studied, especially in the natural sciences (Brooks, 1994; Hansson, 2015). In experimental economics research, the use of technology also plays an important role. Starting from calculators to computer tomographs, techniques of different complexity are used in experiments and increase the quality of economic experiments. However, the relationship between experiments and technology has so far been a kind of one-way street, because technical innovations are completely exogenous. Experimental economics has made no contribution of its own to the development of better technology. The reciprocal acceleration that we can observe in the natural sciences is therefore not taking place.

The central thesis of this paper is that with machine learning (ML) a new technology is at hand, where for the first time experimental economics can contribute to enabling substantial improvement of technology. At the same time, ML opens up new questions for experimental research because it can generate observations that were previously impossible (Camerer, 2019). Thus, for the first time, there could be fruitful interaction between technology and experimental research in economics.

Despite the high relevance of ML for experimental research, we do not observe that it is used in experiments to any significant extent. To understand why this is the case, we first need to clarify which conditions technologies have to fulfill to be applied and to gain acceptance in experimental research in economics. We will argue that a necessary condition is that a Pareto improvement is achieved by the use of the technology with respect to three quality characteristics of experimental research (internal and external validity, reproducibility). A sufficient condition for the use of a technology is that it is also available at a low cost, whereby the necessary learning effort for mastering the technology must also be included in the costs.

Based on this finding, we argue that the current quality of ML is not high enough to achieve the necessary Pareto improvement. We locate one cause of the insufficient quality in the fact that the datasets needed to train the ML algorithms are not of the necessary quality. At this point, a door opens for experimental economics, allowing it, in turn, to contribute to the advancement of technology. Experimental methodology is extremely well suited for generating data under controlled conditions. It can therefore not only be used to evaluate algorithm based systems

(Gardner et al., 1993; Gupta et al., 2018) but to generate exactly the data that is best suited to train ML algorithms (Marsden & Pingry, 2018). In this way, the experimental method can be used to optimize the algorithms, which cannot be achieved on the basis of the data sets available so far.

In this paper, we will first consider what conditions a technology must meet in order to be used in experimental research (Section 2). Then, we will use a concrete example (detecting lies using AI) to demonstrate that the availability of datasets for training ML algorithms is indeed the limiting factor for using ML in experimental research (Section 3). We will use the same example to demonstrate how an experiment can be used to generate optimal training datasets, and we will show that doing so can already generate interesting results for experimental research (Section 4).

2. When do technologies succeed?

The first economic experiments were conducted using pen and paper or even blackboard and chalk. This has changed because gradually more and more technological tools came into use. In the beginning, it was simple computer networks, later mobile devices, smartphones, and the Internet came along. On the other hand, some technological innovations have not managed to get beyond a niche existence. FMRI and eye tracking, for example, have not been able to establish themselves to this day. Why is that? What determines whether a new technology becomes an integral part of the experimental method or not? This question is important because only if it can be answered, it will be possible to assess whether and under what conditions techniques that make use of machine learning can become established in the laboratory.

To answer the question raised above, it is useful to briefly recall the purpose of economic experiments. The major goal of conducting economic experiments is to study human behavior in a controlled environment. Experimental economics focuses on people's decisions on the use of scarce resources and aims to establish causal links. The central factor is the proximity of economic experiments to formal models. Given that models rely on several strict assumptions, it is essential to ensure these assumptions are met when the theory is tested. This is why controlling the experimental environment is of major importance. This contrasts with classical empirical studies which do not rely on experimental control but on sophisticated statistics and econometrics to identify existing relationships.

This description of the tasks of experimental research suggests that the following three criteria should be used to assess the quality of experimental research: Internal validity, external validity, and reproducibility (Weimann & Brosig-Koch, 2019). The internal validity of an experiment ensures that experiments investigate exactly what they are supposed to investigate. For example, if an economic model is tested, it must be ensured that the experimental design reflects all incentives and restrictions that are assumed to be effective in the model. It is one of the outstanding strengths of economic experiments that they can ensure high internal validity due to the good controllability of many influencing factors in the laboratory. The *external* validity of experiments is a quality criterion because experimental research ultimately also serves to uncover causal relationships in reality (IJzerman et al., 2020). The *reproducibility* of an experiment is a criterion for its quality because reproducing experiments is an elementary and indispensable component of successful experimental research (Camerer et al., 2016). A single experiment allows one to make statements about the behavior of a certain group of people, at a certain place, at a certain time under certain boundary conditions. A generally valid causal connection cannot be asserted by a single experiment. Only the multiple repetitions of the experiment with other persons, at other places, at other times, and under other boundary conditions can provide evidence for the existence of a generally valid causal relationship. Therefore, the reproducibility of an experiment is a quality criterion. If a new technology can improve one of the three quality criteria without worsening another (i.e., achieving a Pareto improvement), then it fulfills the necessary condition for gaining acceptance in experimental research. What does this mean exactly?

The internal validity of an experiment depends very much on how accurately the variables of a model under test can be measured. Compared to an experiment conducted with pencil and paper, experiments conducted with a computer network are more accurate, faster, and less prone to error, even without specialized software. Thus, the computer improves the internal validity of an experiment. However, better internal validity can also be achieved by allowing technology to collect data that could not be observed without it. fMRI is a good example of this. Undoubtedly, this technology increases the internal validity of an experiment because it (e.g. by observing the BOLD response) allows inferences to be made about brain activity that would not be observable without fMRI. Internal validity of an experiment, however, involves not only accurate observation of the dependent variable but also control of other variables that may affect the subject's behavior. Computers and specialized software that was developed purposefully for experimental economics help here as well. A standardized experimental environment that can be achieved using computers and specialized software (e.g. z-Tree (Fischbacher, 2007), oTree (Chen et al., 2016)) implies a high level of control over what subjects can see and do in the experiment. Summing up, technology can contribute to establishing internal validity in terms of either allowing the measurement of the relevant variables or fostering control over other variables.

The external validity of experiments can also be increased by the use of certain technologies. This is especially true when it comes to technologies that also play an increasing role in the real world. Consider the case of communication. In reality, communication plays a central role in human decision-making and should be analyzed in laboratory experiments (Brandts et al., 2019). On the one hand, technology enables experimenters to implement communication in a controlled manner (thus contributing to internal validity). On the other hand, video-conference tools became an integral part of our society. Using similar or identical communication tools in laboratory experiments as in the real world increases parallelism (Brosig et al., 2003). Summing up, technology can contribute to increasing external validity.

Finally, let us consider the third quality criterion "replicability". Standardized experimental environments, e.g. z-Tree and oTree enable researchers to copy the experiment. This facilitates replications and increases their quality. Data repositories allow access to experimental data. None of it would be possible without technology (i.e. computers and networks). Self-evidently, there are non-technological factors that foster replicability (research culture – scientific appreciation of replications or publication criteria). We do not negate those, but simply stress the necessity of technology to improve replicability in experimental economics.

So far, we have only talked about the necessary conditions for the success of a new technology. What is missing is a consideration of the costs associated with the use of technology. For this, a preliminary consideration is necessary. We have argued above that the necessary condition for the success of a technology is that a Pareto improvement is achieved with respect to the quality criteria. Thus, it is clear that the use of fMRI, for example, does not fulfill this condition, because fMRI strengthens internal validity, but this is done at the expense of external validity and reproducibility. Thus, it is clear that fMRI will not gain widespread acceptance. However, one can also add the deterioration of external validity and reproducibility to the (opportunity) costs of the technology. In this case, its use makes sense at least in cases where the total cost is still smaller than the benefit of better internal validity. Concerning fMRI, this is only in relatively rare cases, because not only the opportunity costs are significant, but also the monetary costs and the learning effort for experimenters are immensely high.

In summary, the conditions for a technology to be successful can be described as follows. If the use of a new technology allows a Pareto improvement in the quality criteria and the monetary and learning costs are low, the sufficient conditions for the widespread use of the technology are met. If there is a tradeoff between the quality criteria, the use of the technology is per se associated with high opportunity costs, which reduce the chances of success. If this is compounded by high monetary and/or learning costs, this reduces the prospects of success even further.

3. Machine learning, lie detection, and the data set problem

3.1. A short introduction to machine learning

Machine learning (ML) is a core scientific field of artificial intelligence (AI) that provides machines with the ability to learn without being explicitly programmed (Samuel, 1959). Before ML, AI methods were only strict about solving low-level tasks in business and enterprise settings like automation or rule-based classification. Traditional rule-based methods generate predefined outputs based on particular rules programmed by humans. In contrast, ML models simulate human intelligence by learning the rules from the training data to solve complex problems, as shown in Fig. 1.[2]

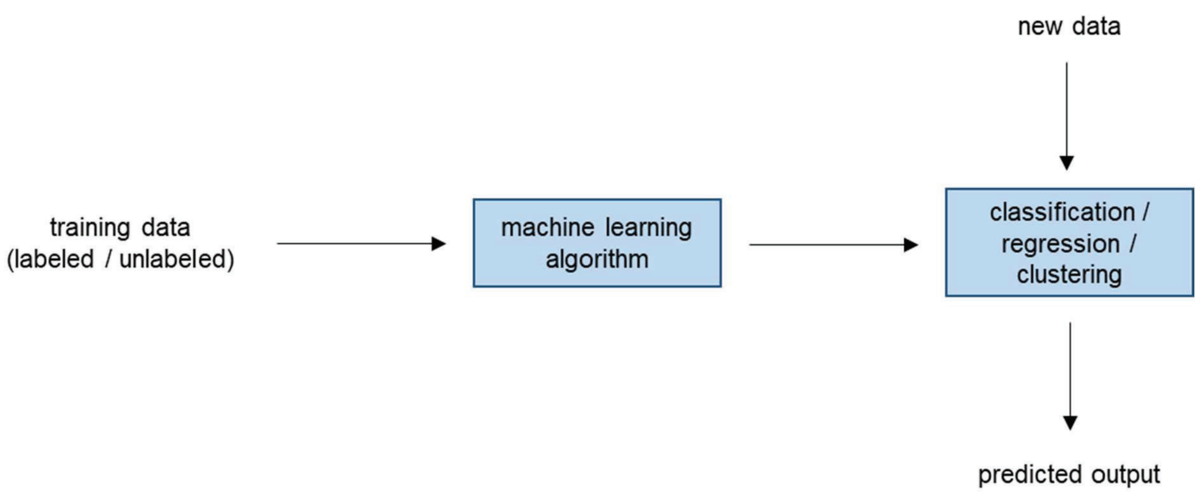

Fig. 1 A machine learning pipeline

The core concept in ML algorithms is the process of training. The data used for this purpose can be either labeled or unlabeled training. Labeled data comes with corresponding ground

[2] ML is considered one of the most interesting research topics and applies to different scientific areas, including bioinformatics (Kong et al., 2007; Mackowiak et al., 2015), medicine (Asadi et al., 2014; Kang et al., 2015; Zhang et al., 2017), meteorology (Aybar-Ruiz et al., 2016; Cramer et al., 2017; Rhee & Im, 2017), economics (Barboza et al., 2017; Bohanec et al., 2017; Zhao et al., 2017), robotics (Fiedler et al., 2021; Gastaldo et al., 2015; Strazdas et al., 2022), and food security (Fragni et al., 2018; Maione & Barbosa, 2019).

truths assigning each sample to a certain class or denoting specific information therein. In contrast, unlabeled data comes without any ground truths. Based on the type of data and the way of training, ML distinguishes between supervised and unsupervised learning (Mohammed et al., 2016).

Supervised learning is considered the most fundamental type of ML, which uses labeled data to obtain an optimal model with a good performance by training ML algorithms. The ML model tries to find the relationship between input data and ground truths during training to emulate this into a mapping function. Its performance is afterward evaluated on other labeled data, referred to as the test set. Supervised learning is performed when specific goals are identified to be addressed using a certain set of inputs, thus it represents a task-driven approach (Sarker et al., 2020). The main drawback of this approach is that labeled data is challenging to obtain.

Unsupervised learning has the advantage of working with unlabeled data to obtain the underlying hidden structure, hence it is a data-driven approach (Han et al., 2012). The advantage is that the training data does not need to be to be extensively prepared by humans. Unsupervised learning models are considered computationally intensive, as they need extensive unlabeled data to output intended results. Their main drawback is that they can provide highly inaccurate results that need to be validated by humans.

For applications that aim to generate outputs for new data solely supervised training approaches can be considered since unsupervised training is only able to determine internal structures and dependencies of an already existing dataset. Due to these constraints concerning possible fields of application, supervised training methods are much more dominant in ML, leading as a consequence to the necessity for annotating the data. During this annotation process, one or more labels are assigned to each image or video sequence varying in type and quantity depending on the desired task. Thereby, these labels set the learning goal of the model and represent the optimal outputs. To accomplish this labeling, one relies on human annotators and experts in the respective field to ensure a high level of data quality.

3.2. Lie detection with machine learning

As mentioned earlier, in the following we will use lie detection as an example of an application of ML that can potentially generate an advance in experimental economics research while being improved by experimentally generated datasets. In principle, ML can detect lies if a lie is expressed in a certain way in the facial expressions of the person who says it. The decisive factor here is that the human face has a great many micro-expressions that cannot normally be consciously controlled. The technology would fail if each person's facial expressions moved in an individual way during a lie. However, if certain expressions were typical of a lie, then these could be detected with a well-trained ML algorithm.

Using a successful lie detection algorithm would mean using a new technology in an economic experiment. But would that represent an advance in experimental research in the sense of the considerations of the second paragraph? To answer this question, we need to apply the criteria mentioned in section two.

Internal validity

Similar to studies using fMRI technology, ML allows the observation of variables that cannot be observed without this technique. The human eye is unable to detect and correctly interpret micro expressions. Numerous experimental studies on lie detection have shown that humans are very poor at detecting lies (Ockenfels & Selten, 2000). ML can lead to significantly higher probabilities with which lies can be unmasked. This has two important implications. On the one hand, it makes it easier to experimentally reproduce models that assume a certain probability of detection. For example, consider the communication model of Kartik (2009) or models that depict corrupt behavior that is detected with a certain probability (Allingham & Sandmo, 1972). On the other hand, ML enables exploratory experiments that investigate how the existence of a technology that can detect lies affects behavior, for example, in negotiations. Will negotiators resort to this technology? How does the ability to detect lies with higher probability change the outcome of negotiations? In summary, it is clear that internal validity increases by a ML technology for lie detection and that the research horizon is significantly expanded with it.

External validity

Concerning external validity, the central question is whether such tools will be applied in the real world. Although forecasts are speculative by nature, there are a few indications that if such technology becomes accessible, it will be applied, as there is already some demand for it in the real world (Sánchez-Monedero & Dencik, 2020). Assuming, the technology will prevail, the application of such a tool in a laboratory experiment increases the external validity.

Replicability

Provided that the algorithm for lie detection is available as open source, it should be clear that the reproducibility of an experiment is given in the same way as it is the case for the replication of standard statistical procedures. Thus, it is clear that ML fulfills the necessary conditions for technical success concerning lie detection because it represents a Pareto improvement concerning the quality criteria considered so far. Finally, it remains to be clarified what costs are associated with the use of this technology.

Costs

Concerning acquisition costs of physical devices, the costs are comparatively low as normal HD Web-Cams suffice. Thus, the costs are similar to those for a webcam. Concerning other costs, the ultimate goal of developing such tools in science is to make them publicly available as open source. Yet, by definition, the existence of well-functioning open-source alternatives makes the financial costs for researchers very low. The learning costs are hardly known a priori but the application of the open-source tool is unlikely to be more complicated than currently applied code for statistical analysis in Python, R, or Stata. If commercial software provides an even more intuitive user interface, this will make the application even easier.

The monetary costs and the learning costs are thus so low that they certainly do not stand in the way of widespread use of the new technology. However, one could argue that the use of ML leads to opportunity costs because it is accompanied by a higher control effort and a correspondingly pronounced monitoring of the subjects' behavior. Behavior is monitored much more intensively than in laboratory experiments without the use of ML which potentially changes the behavior, e.g. effort or lying behavior (Boly, 2010; Jansen et al., 2018; Kajackaite & Gneezy, 2017). However, at least in the case of lie detection, this corresponds to reality when such technology is used. The awareness of being observed more closely than otherwise is exactly what is needed in terms of higher external validity. In this respect, there are no additional opportunity costs at this point.

Summing up, the application of ML on videos in experimental economics potentially combines some of the benefits of fMRI research (i.e., detecting variables that otherwise remain hidden to humans) with the low costs of common hardware and open-source software. As we discussed, the costs of this technology are moderate. Based on our prior discussion this means, if the quality of the algorithms is high enough, they are very likely to be applied in experimental economics. This leads to the essential question: how good are these algorithms?

3.3. The data set problem

ML has evolved significantly in recent years and in particular, deep learning with its deep neural networks (DNN) has emerged as the fastest-growing field in ML (Cao et al., 2018). ML models make decisions based on data fed into them during the training process. From this data, they identify patterns and derive features to perform the subsequent classification. The deeper a network architecture is, the more data is needed for training to adjust the weights according to the given task. The severe drawback of DNNs is that most of them are black boxes (Buhrmester et al., 2021). Thus, it is neither transparent for users how and why a decision was made nor for developers what exactly the network has learned. Therefore, the quality of the training data plays a key role, as it is responsible for the later behavior of the system.

In the following, we discuss what is important for high-quality data using the case of video-data-driven lie detection. First, we will discuss the data quality parameters of (i) knowing the ground truth and (ii) establishing endogenous lying. Then, we address parameters of (iii) reproducibility of datasets, (iv) deception of subjects, and (v) technological conditions.

First, if it is the central goal to train an algorithm to detect lies, then it is of paramount importance to obtain a data set that contains lies and truths as well as provides the ability for the researchers to distinguish between them with certainty. In the context of this paper we refer to this as knowing the ground truth. Therefore, when generating data for lie detection algorithms, it is important to obtain full control of the data generation process. This can be done through laboratory experiments. Such high level of control facilitates a key process for ML: annotations. As supervised learning is the dominant methodology in ML, raw data is of little utility without annotations (Garrow et al., 2021). Providing such annotations is a labor-intensive and error-prone task and usually relies on human annotators (Paullada et al., 2021). Their task is either the entirely manual creation of ground truths or the verification and refinement of annotations in a semi-automatic manner (Spasic & Nenadic, 2020). Yet, gathering data in a laboratory experiments leads to either fully automatic annotations or annotations that are very easy to conduct by humans. Therefore, this reduces labor-intensity and error-proneness of the data generation process.

Second, to establish high external validity of the data set, subjects should lie based on their own decisions in contrast to simply being asked to lie by the experimenter. This resembles much closer the real world. Therefore, a good data set on lying, should be based on an experiment, where subjects decide on their own whether to lie or not. To make this possible, it is crucial to establish the right financial incentives in the experiment. These incentives should resemble the real-world conflict between lying and telling the truth with respective consequences. This can be done by inducing value in the decisions (Smith, 1976, 1982) which is done through financial incentives. Such a carefully controlled laboratory experiment would provide top quality data as argued by Marsden and Pingry (2018).

Further, providing financial incentives contributes to the third quality parameter: reproducibility. A good economic experiment has financial incentives that dominate other subjective values. This and the careful documentation of the entire experimental design enable researchers to conduct identical experiments in different laboratories (Weimann & Brosig-Koch, 2019). This increases the validity of the data and enables other researchers to expand existing data bases from prior experiments with new data.

The fourth parameter is the no deception rule that is applied in experimental economics. This rule ensures that subjects trust the experimenter. The importance of the rules is highlighted when considering the opposite. If subjects do not trust the instructions because they are sometimes lied to, their behavior depends on their belief whether the instructions are true. This reduces experimental control and therefore, the quality of the data (Weimann & Brosig-Koch, 2019).

Finally, when it comes to data quality, it is important to control the technological environment of the experiment. For visual data, in particular, there is a multitude of possible environmental interferences such as varying illumination and others (Bae et al., 2015; Chi & Caldas, 2011; Zhu et al., 2016). In addition, camera characteristics can affect the quality (Golparvar-Fard et al., 2009; Hui et al., 2015). Furthermore, when storing large amount of data, the type of compression and possible accompanying losses should be taken under advisement (Guo et al., 2021; Israel et al., 2020). All these aspects are fundamental to the development of sophisticated algorithms capable of dealing with uncontrolled operations (Xu et al., 2021).

After establishing the quality parameters for high-quality data for lie-detection software, we briefly analyze data sets in current use. We identified six major data sets that are used for training lie-detection algorithms. Yet, none of these fulfills all the quality parameters we discussed (see Table 1). In some experiments, not even the experimenters know the truth. In others, the subjects are not incentivized to lie but are simply told to do so without any control mechanisms. Further, several experiments employ different types of deception. Thus, while it is possible to train algorithms on these data sets, these algorithms are likely to suffer from the same methodological flaws, inherent to the data generation process. Here, we argue, experimental economics can help. Due to the long history of providing controlled, standardized, well documented, and incentivized experiments in laboratories with good no-deception reputation, this methodology can contribute to providing automatically (or easily) annotated data sets for training algorithms. In the following chapter, we provide an example how a well-studied economic experiment on lying can be used to generate such a video data set.

Table 1. Summary of currently used data sets on lie detection

Paper	True state objectively known[3]	Incentivization[4]	Is incentivization related to the content of the lie	Avoiding subject deception
DDPM	No	Yes	No	Yes
Box of Lies	Yes	No	No	Yes
Real-life trial data	No	Yes	Yes	No subjects
MU3D	No	No	No	Yes
Bag-of-lies	Yes	No	No	Yes
Silesian	Yes	No	No	No

4. The Experiment

4.1. Experimental design

The aim of the experiment was to observe subjects who, of their own free will, either tell the truth or lie in order to gain an advantage. In order to know whether a subject is currently lying or speaking the truth, it was necessary to know the respective "truth". To achieve this goal, we suitably modified the well-known experiment of Fischbacher and Föllmi-Heusi (2013) (FFH). Subjects had to roll a die once. The number of points on the dice is documented on a camera (Camera 1), which is placed on the left-hand side of their desk. Afterward, they are connected to the experimenter via video chat (using Camera 2[5]) and tell her their respective number of points. The payoff is based on the number they tell the experimenter into the camera. In line with Fischbacher and Föllmi-Heusi (2013), for rolling a number between 1 and 5, the subjects receive a payment of 1 to 5€ respectively. If they roll a 6, the payoff is 0€.

[3] This refers to some data sets relying on the subjects themselves defining their statements as true or false. Thus, the experimenter does not objectively know the truth.
[4] Unless clearly stated, we assume no incentivization took place. The real-life trial data constitutes an exemption as people recorded were incentivized through real-life consequences of the judicial system.
[5] To observe the dice, we used a simple Logitech HD Pro Webcam C920 for Camera 1. As Camera 2 which is important for the video data set, we used Logitech BRIO 4K UHD.

Concerning the technical setup, Camera 2 is placed on a monitor in front of the subject. It is used for video chat. Camera 1 is downside-faced placed, mounted at a height of 40cm, and focuses on an area of 20cm*30cm where subjects roll a die. Camera 1 is not connected to the experimenter, i.e., she does not check the individual dice roll and video chat testimonies of experimental subjects. This is important since prior research indicates that the chance of being caught reduces lying behavior (Kajackaite & Gneezy, 2017). Therefore, it is crucial that the subjects believe this setup. Since Charness et al. (2022) observe that a lot of subjects are not aware of the no-deception policy in economic experiments and Frollová et al. (2021) indicate the importance of subjects' perception of information in laboratory experiments with lying, this rule is explicitly stated and verified by a certificate of the GfeW ethics commission, which is presented to the subjects along with the instructions in the experiment[6]. To ensure the anonymity of data, the research process is split up among the authors as it is displayed in Figure 2. Author 4 conducted the experiment in the laboratory but is not informed about the true dice rolls. Author 2 received the videos of the dice and annotated them[7]. Using a pseudonymized identifier, Author 2 merged two sets to analyze individual lying behavior, and sends it to Author 1. Author 1 does neither have any personal information on the subjects nor their videos. This results in a similar situation as in the original paper yet includes lying information on the individual level.

For conducting and recording the videoconferences we used MTV (Bershadskyy et al., 2022).[8] In total, we recruited 148 students. Out of these, 47 students were in the treatment that directly replicated FFH (without cameras) and 101 students belonged to our variation of FFH (with cameras). The students agreed to be recorded in line with the data regulation of the laboratory. The experiments were conducted between June and October 2022. Each experiment lasted at most 10 minutes. During the sessions, there was no communication between subjects. Because of the Covid-19 pandemic and hygienic regulations, people were told that communication with the experimenter during the session takes place via video chat. The experimenter guided the subjects to their places. Throughout the whole experiment, we used the same two sound-insulated booths with the same lightning. At their places, the subjects received instructions (see

[6] To further ensure that all our subjects believe this, we excluded all subjects who study psychology from our sample. The reason is that psychology students conduct a lot of experiments yet experiments in psychology can include deception of the subjects followed by a debriefing after the experiment.

[7] Note that it would have been better if Author 2 worked in a different city as there remains a chance of recognizing somebody from the video on campus. Yet, Author 2 does not work in the laboratory. Therefore, the probability of a random meeting is lower than for experimenters who have a more regular contact to subjects. Our results indicate that such a split suffices.

[8] The tool allows not only for the typical recording of videos but enables researchers to distinguish the audio channels between different speakers. This generates a richer data sample for the engineers.

Appendix A) for the experiment. Subjects were fully informed about the experimental setup, there was no deception.

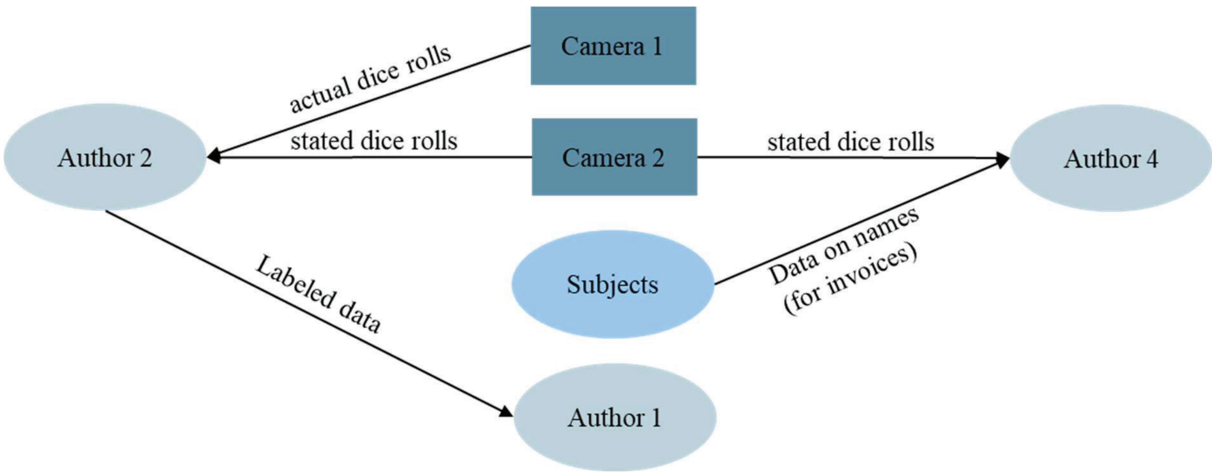

Fig. 2 Roles of authors to establish a high level of anonymity for the subjects

4.2 Experimental Results

In this section we will present the results from our experiment, starting with the same type of analysis that was possible in FFH (2013). After that, we will present additional findings that became possible due to our additional observation.

After cleaning the data, 96 out of 101 observations[9] can be evaluated. Looking at the stated dice rolls, we observe the same pattern as in the original paper (see Figure 3). 37.5% of subjects report having rolled a 5 (payoff maximizing number) and 4.2% report to have rolled a six (payoff minimizing number). To provide additional evidence that this effect is not confounded by a selected sample as compared to the original research, we further conducted a small replication of the original experiment from FFH without monitoring. The results obtained from the control group of 47 students are in line with the original findings and support the argument that the group lying behavior did not change due to monitoring. The average requested payoffs were 3.43€ in the control and 3.68€ in the video treatment (MW-Test: p= 0.2224).

[9] For some of the videos it is unclear whether there was a technical issue, the subjects removed the box, or the subjects simply told a number without having rolled the dice at all. Despite this being a very conservative approach, we decided to exclude all of these cases.

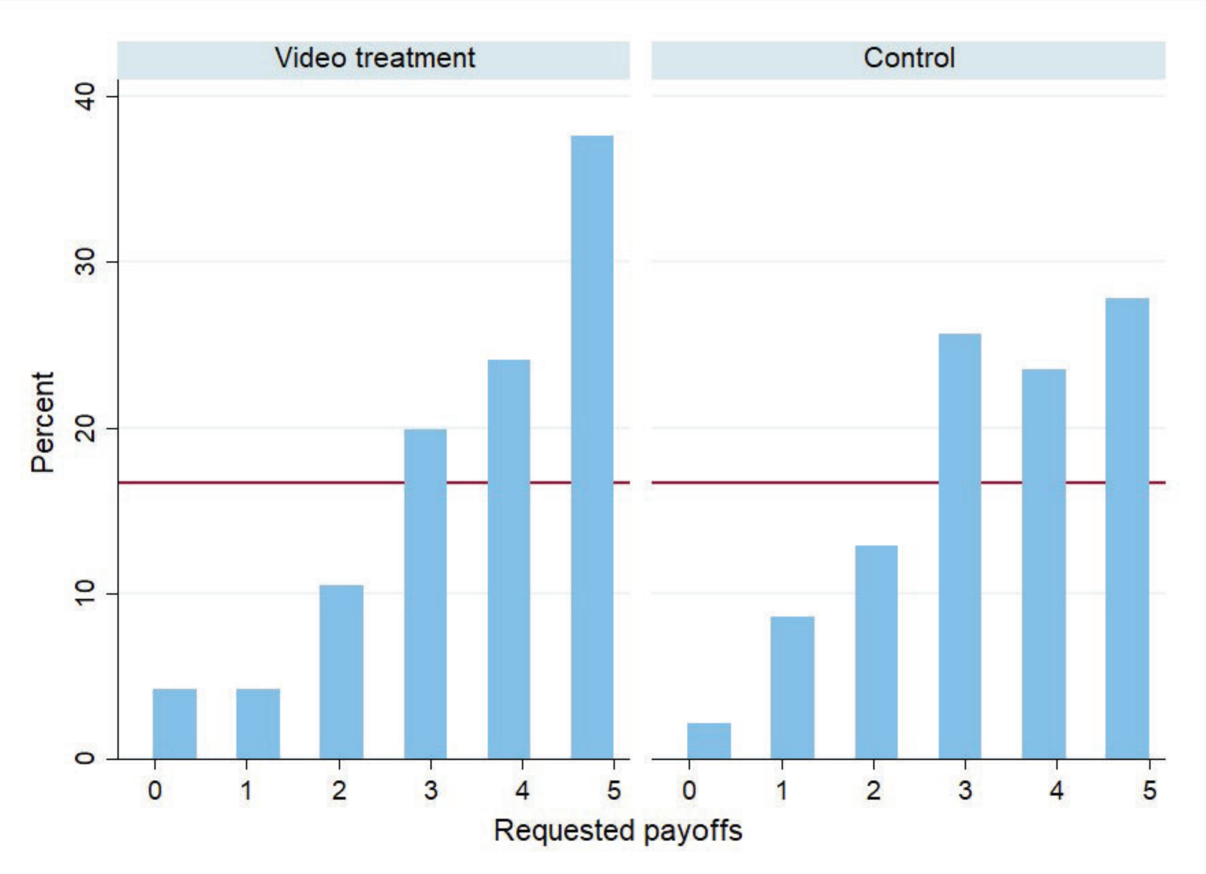

Fig. 3 Histogram of requested payoffs in video-treatment and control group

Note: The red line represents the 16.67% benchmark which is the probability of rolling one specific number.

Thus, after we indicated that monitoring did not change the results, we focus on the information obtained from the process of labeling the data and investigate individual-level lying behavior. This allows us to distinguish between payoffs that the subjects requested and those they would have deserved due to actual dice rolls. Starting with the information from the labeling, there are two interesting observations. First, two subjects did not roll the dice but visibly put them with their hand such that it shows the number 5. We consider this to qualify as a lie. These subjects were recoded to have deserved a payment of zero euros.[10] Second, two subjects rolled more than once to obtain a payoff-increasing number. This is also considered a lie since the subjects were explicitly told to report the number they rolled with their first roll. In these cases, the deserved payment was coded as the payment that would have resulted from the first roll.

[10] Please note, that the results do not change if we were to put any other number (e.g., the average payoff of 2.5€ or would have excluded these observations).

After displaying relevant information from labeling and how it was handled, we turn to statistical analysis. First, we investigate the differences between requested and deserved payoffs. On average subjects requested a payoff of 3.68€ while the actual dice would have yielded a payoff of 2.56€. The difference is highly significant using Wilcoxon signed-rank test ($p<0.0001$). Second, we focus on the question of how many subjects lied. Given our definitions, 37.50% of all subjects lied[11]. Out of 36 individuals who lied to increase their payoff, 23 individuals lied to the fullest extent (stating to have rolled a 5). Four subjects lied to state number 3 and nine subjects lied to state number 4. This provides further evidence for partial lying behavior. Further, following the analysis strategy from (Lilleholt et al., 2020), we can exclude all those subjects who rolled number 5, since they did not have any incentive to lie. This implies that 43.9% of subjects with a financial incentive to lie, reported higher numbers than they rolled. All of this information is considered to be unobtainable in the classical setup with anonymity. Still, some of these results come with a limitation, which shall be discussed in the next chapter.

Our experiment led to the generation of a new data set that will be presented in a companion paper. We consider this to be the first lie detection video data set that uses incentivized subjects without deceiving them and making labeling the data easy. In short, the data set consists of the videos that remained after removing questionable video files. It provides an opportunity to assess and improve prior prediction models used for lie detection. It is being used in ongoing research with ML algorithms achieving an accuracy of 67% (Dinges et al., 2023).

5 Conclusion

The goal of this article is to investigate whether and how experimental economics can benefit from and simultaneously contribute to the development of certain machine learning algorithms. In so doing, we focus on algorithms designed to detect lies from simple conversational data (mostly facial expressions). Lies are an omnipresent element of our daily life, daily work, and politics. In bargaining situations, lies can be used to manipulate the other side or to simply keep private information private. A tool that can interfere with these goals in real-time can substantially change the nature of bargaining. Given the amount of literature on the topic of bargaining and private information, these topics are evidently of high relevance. Thus, research on technologies that affect these topics should be of high relevance, too.

[11] Please note, that one out of the total of 37 lying subjects lied in the other direction. The individual rolled a five but stated a four.

Already these technologies exist. Yet, they rely on disputable data sets. Experimental economics with its rich experimental toolbox can contribute to the development of high-quality data sets that are required to develop good lie detecting algorithms. Likewise, given certain quality of the algorithms, our methodological discussion clearly shows, that these tools will be applied in experimental economics. This result stems from the discussion on the general role of technology in experimental economics which is divided into four major points: internal validity, external validity, replicability, and costs. Our investigation suggests, that given high quality of algorithms, this technology will fulfill the necessary and sufficient conditions for application in experimental economics. Yet, we conclude, that before experimental economics can use such algorithms, it first can support their development.

We, further, provide a practical example of how this process can work. We apply a classical experiment on lying from experimental economics (FFH). We modify the design such that it can provide a high-quality video data set on lying. Then, we show, that this modification does not affect the experimental results. Instead, it provides a richer statistical analysis on the individual level and contributes to future research ideas (e.g., the role of evasive lies). In total, we consider the benefits of such experiments to be substantial. It can deal with classical economic questions on the role of privacy on lying behavior or partial lying and simultaneously contribute to engineering high-quality algorithms by providing the required data sets. Our discussion of necessary and sufficient conditions for a technology to be applied in experimental economics indicates that such machine learning algorithms will be employed in the future. However, to achieve the required quality, there is a need for profound cooperation between experimenters and engineers to provide the high-quality data necessary to improve the algorithms.

List of References

Allingham, M. G., & Sandmo, A. (1972). Income tax evasion: a theoretical analysis. *Journal of Public Economics*, *1*(3–4), 323–338. https://doi.org/10.1016/0047-2727(72)90010-2

Asadi, H., Dowling, R., Yan, B., & Mitchell, P. (2014). Machine learning for outcome prediction of acute ischemic stroke post intra-arterial therapy. *PloS One*, *9*(2), e88225. https://doi.org/10.1371/journal.pone.0088225 PM - 24520356

Aybar-Ruiz, A., Jiménez-Fernández, S., Cornejo-Bueno, L., Casanova-Mateo, C., Sanz-Justo, J., Salvador-González, P., & Salcedo-Sanz, S. (2016). A novel Grouping Genetic Algorithm–Extreme Learning Machine approach for global solar radiation prediction from numerical weather models inputs. *Solar Energy*, *132*, 129–142. https://doi.org/10.1016/j.solener.2016.03.015

Bae, H., Golparvar-Fard, M., & White, J. (2015). Image-Based Localization and Content Authoring in Structure-from-Motion Point Cloud Models for Real-Time Field Reporting Applications. *Journal of Computing in Civil Engineering*, *29*(4). https://doi.org/10.1061/(ASCE)CP.1943-5487.0000392

Barboza, F., Kimura, H., & Altman, E. (2017). Machine learning models and bankruptcy prediction. *Expert Systems with Applications*, *83*, 405–417. https://doi.org/10.1016/j.eswa.2017.04.006

Bershadskyy, D., Ghadwal, S., & Greif, J. (2022). *MTV-Magdeburg Tool for Videoconferences* (FEMM Working Paper Series).

Bohanec, M., Kljajić Borštnar, M., & Robnik-Šikonja, M. (2017). Explaining machine learning models in sales predictions. *Expert Systems with Applications*, *71*, 416–428. https://doi.org/10.1016/j.eswa.2016.11.010

Boly, A. (2010). On the incentive effects of monitoring: evidence from the lab and the field. *Experimental Economics 2010 14:2*, *14*(2), 241–253. https://doi.org/10.1007/S10683-010-9265-1

Brandts, J., Cooper, D. J., & Rott, C. (2019). Communication in laboratory experiments. In A. Schram & A. Ule (Eds.), *Handbook of Research Methods and Applications in Experimental Economics.* (pp. 401–426). Edward Elgar Publishing.

Brooks, H. (1994). The relationship between science and technology. *Research Policy*, *23*(5), 477–486. https://doi.org/10.1016/0048-7333(94)01001-3

Brosig, J., Weimann, J., & Ockenfels, A. (2003). The Effect of Communication Media on Cooperation. *German Economic Review*, *4*(2), 217–241. https://doi.org/10.1111/1468-0475.00080

Buhrmester, V., Münch, D., & Arens, M. (2021). Analysis of Explainers of Black Box Deep Neural Networks for Computer Vision: A Survey. *Machine Learning and Knowledge Extraction*, *3*(4), 966–989. https://doi.org/10.3390/make3040048

Camerer, C. F. (2019). Artificial Intelligence and Behavioral Economics. In A. Agrawal, J. Gans, & J. Goldfarb (Eds.), *The Economics of Artificial Intelligence: An Agenda* (pp. 587–610). University of Chicago Press. https://doi.org/10.7208/9780226613475-026

Camerer, C. F., Dreber, A., Forsell, E., Ho, T. H., Huber, J., Johannesson, M., Kirchler, M., Almenberg, J., Altmejd, A., Chan, T., Heikensten, E., Holzmeister, F., Imai, T., Isaksson, S., Nave, G., Pfeiffer, T., Razen, M., & Wu, H. (2016). Evaluating replicability of laboratory experiments in economics. *Science*, *351*(6280), 1433–1436. https://doi.org/doi/10.1126/science.aaf0918

Cao, C., Liu, F., Tan, H., Song, D., Shu, W., Li, W., Zhou, Y., Bo, X., & Xie, Z. (2018). Deep Learning and Its Applications in Biomedicine. *Genomics, Proteomics & Bioinformatics*, *16*(1), 17–32. https://doi.org/10.1016/j.gpb.2017.07.003

Charness, G., Samek, A., & van de Ven, J. (2022). What is considered deception in experimental economics? *Experimental Economics*, *25*(2), 385–412. https://doi.org/10.1007/S10683-021-09726-7/TABLES/6

Chen, D. L., Schonger, M., & Wickens, C. (2016). oTree-An open-source platform for laboratory, online, and field experiments. *Journal of Behavioral and Experimental Finance*, *9*, 88–97. https://doi.org/10.1016/j.jbef.2015.12.001

Chi, S., & Caldas, C. H. (2011). Automated Object Identification Using Optical Video Cameras on Construction Sites. *Computer-Aided Civil and Infrastructure Engineering*, *26*(5), 368–380. https://doi.org/10.1111/j.1467-8667.2010.00690.x

Cramer, S., Kampouridis, M., Freitas, A. A., & Alexandridis, A. K. (2017). An extensive evaluation of seven machine learning methods for rainfall prediction in weather derivatives. *Expert Systems with Applications*, *85*, 169–181. https://doi.org/10.1016/j.eswa.2017.05.029

Dinges, L., Fiedler, M.-A., Al-Hamadi, A., Hempel, T., Abdelrahman, A., Weimann, J., & Bershadskyy, D. (2023). *Automated Deception Detection from Videos: Using End-to-End Learning Based High-Level Features and Classification Approaches*.

Fiedler, M.-A., Werner, P., Khalifa, A., & Al-Hamadi, A. (2021). SFPD: Simultaneous Face and Person Detection in Real-Time for Human-Robot Interaction. *Sensors (Basel, Switzerland)*, *21*(17). https://doi.org/10.3390/s21175918 PM - 34502809

Fischbacher, U. (2007). z-Tree: Zurich toolbox for ready-made economic experiments. *Experimental Economics*, *10*(2), 171–178. https://doi.org/10.1007/s10683-006-9159-4

Fischbacher, U., & Föllmi-Heusi, F. (2013). Lies in Disguise—An Experimental Study on Cheating. *Journal of the European Economic Association*, *11*(3), 525–547. https://doi.org/10.1111/JEEA.12014

Fragni, R., Trifirò, A., Nucci, A., Seno, A., Allodi, A., & Di Rocco, M. (2018). Italian tomato-based products authentication by multi-element approach: A mineral elements database to distinguish the domestic provenance. *Food Control*, *93*, 211–218. https://doi.org/10.1016/j.foodcont.2018.06.002

Frollová, N., Vranka, M., & Houdek, P. (2021). A qualitative study of perception of a dishonesty experiment. *Journal of Economic Methodology*, *28*(3), 274–290. https://doi.org/10.1080/1350178X.2021.1936598

Gardner, C. L., Marsden, J. R., & Pingry, D. E. (1993). The design and use of laboratory experiments for DSS evaluation. *Decision Support Systems*, *9*(4), 369–379. https://doi.org/10.1016/0167-9236(93)90047-7

Garrow, C. R., Kowalewski, K.-F., Li, L., Wagner, M., Schmidt, M. W., Engelhardt, S., Hashimoto, D. A., Kenngott, H. G., Bodenstedt, S., Speidel, S., Müller-Stich, B. P., & Nickel, F. (2021). Machine Learning for Surgical Phase Recognition: A Systematic Review. *Annals of Surgery*, *273*(4), 684–693. https://doi.org/10.1097/SLA.0000000000004425 PM - 33201088

Gastaldo, P., Pinna, L., Seminara, L., Valle, M., & Zunino, R. (2015). A tensor-based approach to touch modality classification by using machine learning. *Robotics and Autonomous Systems*, *63*, 268–278. https://doi.org/10.1016/j.robot.2014.09.022

Golparvar-Fard, M., Peña-Mora, F., Arboleda, C. A., & Lee, S. (2009). Visualization of Construction Progress Monitoring with 4D Simulation Model Overlaid on Time-Lapsed

Photographs. *Journal of Computing in Civil Engineering*, *23*(6), 391–404. https://doi.org/10.1061/(ASCE)0887-3801(2009)23:6(391)

Guo, Y., Liu, X., Peng, S., Jiang, X., Xu, K., Chen, C., Wang, Z., Dai, C., & Chen, W. (2021). A review of wearable and unobtrusive sensing technologies for chronic disease management. *Computers in Biology and Medicine*, *129*, 104163. https://doi.org/10.1016/j.compbiomed.2020.104163 PM - 33348217

Gupta, A., Kannan, K., & Sanyal, P. (2018). Economic experiments in information systems. *MIS Quarterly: Management Information Systems*, *42*(2), 595–606. https://doi.org/10.25300/MISQ/2018/13049

Han, J., Pei, J., & Kamber, M. (2012). Data mining. In *The Morgan Kaufmann series in data management systems* (3rd ed.). Morgan Kaufmann. http://www.worldcat.org/oclc/1310206920

Hansson, S. O. (2015). *Preview* (S. O. Hansson (ed.); Vol. 18, pp. 3–9). Springer Netherlands. https://doi.org/10.1007/978-94-017-9762-7_1

Hui, L., Park, M.-W., & Brilakis, I. (2015). Automated Brick Counting for Façade Construction Progress Estimation. *Journal of Computing in Civil Engineering*, *29*(6). https://doi.org/10.1061/(ASCE)CP.1943-5487.0000423

IJzerman, H., Lewis, N. A., Przybylski, A. K., Weinstein, N., DeBruine, L., Ritchie, S. J., Vazire, S., Forscher, P. S., Morey, R. D., Ivory, J. D., & Anvari, F. (2020). Use caution when applying behavioural science to policy. In *Nature Human Behaviour* (Vol. 4, Issue 11, pp. 1092–1094). Nature Research. https://doi.org/10.1038/s41562-020-00990-w

Israel, S. A., Irvine, J. M., & Tate, S. (2020). Machine Learning, Compression, and Image Quality. In *2020 IEEE Applied Imagery Pattern Recognition Workshop (AIPR)* (pp. 1-7 TS-CrossRef). IEEE. https://doi.org/10.1109/AIPR50011.2020.9425256

Jansen, A. M., Giebels, E., van Rompay, T. J. L., & Junger, M. (2018). The influence of the presentation of camera surveillance on cheating and pro-social behavior. *Frontiers in Psychology*, *9*(OCT), 1937. https://doi.org/10.3389/FPSYG.2018.01937/BIBTEX

Kajackaite, A., & Gneezy, U. (2017). Incentives and cheating. *Games and Economic Behavior*, *102*, 433–444. https://doi.org/10.1016/j.geb.2017.01.015

Kang, J., Schwartz, R., Flickinger, J., & Beriwal, S. (2015). Machine Learning Approaches for Predicting Radiation Therapy Outcomes: A Clinician's Perspective. *International Journal* *of Radiation Oncology, Biology, Physics*, *93*(5), 1127–1135. https://doi.org/10.1016/j.ijrobp.2015.07.2286 PM - 26581149

Kartik, N. (2009). Strategic Communication with Lying Costs. *Review of Economic Studies*, *76*(4), 1359–1395. https://doi.org/10.1111/j.1467-937X.2009.00559.x

Kong, L., Zhang, Y., Ye, Z.-Q., Liu, X.-Q., Zhao, S.-Q., Wei, L., & Gao, G. (2007). CPC: assess the protein-coding potential of transcripts using sequence features and support vector machine. *Nucleic Acids Research*, *35*(Web Server issue), W345-9. https://doi.org/10.1093/nar/gkm391 PM - 17631615

Lilleholt, L., Schild, C., & Zettler, I. (2020). Not all computerized cheating tasks are equal: A comparison of computerized and non-computerized versions of a cheating task. *Journal of Economic Psychology*, *78*, 102270. https://doi.org/10.1016/J.JOEP.2020.102270

Mackowiak, S. D., Zauber, H., Bielow, C., Thiel, D., Kutz, K., Calviello, L., Mastrobuoni, G., Rajewsky, N., Kempa, S., Selbach, M., & Obermayer, B. (2015). Extensive identification and analysis of conserved small ORFs in animals. *Genome Biology*, *16*, 179. https://doi.org/10.1186/s13059-015-0742-x PM - 26364619

Maione, C., & Barbosa, R. M. (2019). Recent applications of multivariate data analysis methods in the authentication of rice and the most analyzed parameters: A review. *Critical Reviews in Food Science and Nutrition*, *59*(12), 1868–1879. https://doi.org/10.1080/10408398.2018.1431763 PM - 29363991

Marsden, J. R., & Pingry, D. E. (2018). Numerical data quality in IS research and the implications for replication. *Decision Support Systems*, *115*, A1–A7. https://doi.org/10.1016/j.dss.2018.10.007

Mohammed, M., Khan, M. B., & Bashier, E. B. M. (2016). *Machine Learning*. CRC Press TS - CrossRef. https://doi.org/10.1201/9781315371658

Ockenfels, A., & Selten, R. (2000). An Experiment on the Hypothesis of Involuntary Truth-Signalling in Bargaining. *Games and Economic Behavior*, *33*(1), 90–116. https://doi.org/10.1006/game.1999.0762

Paullada, A., Raji, I. D., Bender, E. M., Denton, E., & Hanna, A. (2021). Data and its (dis)contents: A survey of dataset development and use in machine learning research. *Patterns (New York, N.Y.)*, *2*(11), 100336. https://doi.org/10.1016/j.patter.2021.100336 PM - 34820643

Rhee, J., & Im, J. (2017). Meteorological drought forecasting for ungauged areas based on machine learning: Using long-range climate forecast and remote sensing data. *Agricultural and Forest Meteorology*, *237–238*, 105–122. https://doi.org/10.1016/j.agrformet.2017.02.011

Samuel, A. L. (1959). Some Studies in Machine Learning Using the Game of Checkers. *IBM Journal of Research and Development*, *3*(3), 210–229. https://doi.org/10.1147/rd.33.0210

Sánchez-Monedero, J., & Dencik, L. (2020). The politics of deceptive borders: ‘biomarkers of deceit’ and the case of iBorderCtrl. *Https://Doi.Org/10.1080/1369118X.2020.1792530*, *25*(3), 413–430. https://doi.org/10.1080/1369118X.2020.1792530

Sarker, I. H., Kayes, A. S. M., Badsha, S., Alqahtani, H., Watters, P., & Ng, A. (2020). Cybersecurity data science: an overview from machine learning perspective. *Journal of Big Data*, *7*(1). https://doi.org/10.1186/s40537-020-00318-5

Smith, V. L. (1976). Experimental Economics: Induced Value Theory. *The American Economic Review*, *66*(2), 274–279. http://www.jstor.org/stable/1817233

Smith, V. L. (1982). Microeconomic Systems as an Experimental Science. *American Economic Review*, *72*(5), 923–955. https://econpapers.repec.org/RePEc:aea:aecrev:v:72:y:1982:i:5:p:923-55

Spasic, I., & Nenadic, G. (2020). Clinical Text Data in Machine Learning: Systematic Review. *JMIR Medical Informatics*, *8*(3), e17984. https://doi.org/10.2196/17984 PM - 32229465

Strazdas, D., Hintz, J., Khalifa, A., Abdelrahman, A. A., Hempel, T., & Al-Hamadi, A. (2022). Robot System Assistant (RoSA): Towards Intuitive Multi-Modal and Multi-Device Human-Robot Interaction. *Sensors (Basel, Switzerland)*, *22*(3). https://doi.org/10.3390/s22030923 PM - 35161671

Weimann, J., & Brosig-Koch, J. (2019). *Methods in Experimental Economics*. Springer International Publishing. https://doi.org/10.1007/978-3-319-93363-4

Xu, S., Wang, J., Shou, W., Ngo, T., Sadick, A.-M., & Wang, X. (2021). Computer Vision Techniques in Construction: A Critical Review. *Archives of Computational Methods in Engineering*, *28*(5), 3383–3397. https://doi.org/10.1007/s11831-020-09504-3

Zhang, B., He, X., Ouyang, F., Gu, D., Dong, Y., Zhang, L., Mo, X., Huang, W., Tian, J., & Zhang, S. (2017). Radiomic machine-learning classifiers for prognostic biomarkers of advanced nasopharyngeal carcinoma. *Cancer Letters*, *403*, 21–27.

https://doi.org/10.1016/j.canlet.2017.06.004

Zhao, Y., Li, J., & Yu, L. (2017). A deep learning ensemble approach for crude oil price forecasting. *Energy Economics*, *66*, 9–16. https://doi.org/10.1016/j.eneco.2017.05.023

Zhu, Z., Park, M.-W., Koch, C., Soltani, M., Hammad, A., & Davari, K. (2016). Predicting movements of onsite workers and mobile equipment for enhancing construction site safety. *Automation in Construction*, *68*, 95–101. https://doi.org/10.1016/j.autcon.2016.04.009

Appendix A

Translated Instructions: Camera-treatment

Ethical review by the Society for Experimental Economic Research (GfeW)

The GfeW with identification code (fnax4PA2) certified this experiment. This is a confirmation that all the information in the instructions is true. A copy of the audit is enclosed.

Instructions

For the following experiment, you will receive a payout that depends on the number of points you score with a single dice roll.

To comply with hygiene measures, communication in this experiment will be via the screen. You can now take off the mask for this. For communication, please put on the headset in front of you.

Your task is to throw a fair die **once** into the marked area to your left. The die may only be rolled once and must be within the marked area.

After you have rolled the dice, you can click on the "Next" button and you will be connected to the experimenter via video chat. She will ask you what number you rolled. Please answer in a complete sentence:

I rolled the number X.

Your payoff depends only on the number of dots you speak into the camera. The experimenter does **not** know what number you actually rolled. Even after the experiment, she will not know.

The following table shows you the payout for each number rolled:

Number of eyes on dice	1	2	3	4	5	6
Payoff	1€	2€	3€	4€	5€	0€

Say you rolled a number between 1 and 5, your payout is €1 to €5 accordingly. If you roll a 6, your payout is 0€.

Then, when prompted, you can get up and collect the corresponding money from the experimenter and leave the lab. To do this, put the mask back on.

From now on, please follow the instructions on the screen.

Translated Instructions: Control group

For the following experiment you will receive a payoff that depends on the number of points you score with a single roll of the dice.

Your task is to roll a fair die **once** in the marked area to your left. The die may only be rolled once and must be within the marked area. You have 30 seconds to roll the dice.

Your payoff depends only on the number of points you enter in the input field. The experimenter does **not** know what number you actually rolled. Even after the experiment, she will not know.

The following table shows you the payoff for each number rolled:

Number of eyes on dice	1	2	3	4	5	6
Payoff	1€	2€	3€	4€	5€	0€

Say you rolled a number between 1 and 5, your payoff will be €1 to €5 accordingly. If you roll a 6, your payoff is €0.

After you have rolled the dice, please click on "Next" and enter your rolled number. Once the experimenter has called your booth number, please step out of the booth. You can then collect your payoff and the experiment is finished after that.

Otto von Guericke University Magdeburg
Faculty of Economics and Management
P.O. Box 4120 | 39016 Magdeburg | Germany

Tel.: +49 (0) 3 91/67-1 85 84
Fax: +49 (0) 3 91/67-1 21 20

www.fww.ovgu.de/femm

ISSN 1615-4274