DOI: 10.52335/ekon/161838

BEATA WOŹNIAK-JĘCHOREK*

Experiments in modern economics – expansion and technological and institutional innovations in the U.S.[1]

1. Introduction

The experimental approach has been the protagonist of one of the most stunning methodological revolutions in the history of social science. Empirical research is more and more often confronted with questions regarding the validity of its causal claims. Experiments are one of the most powerful tools used to resolve this problem. Consequently, in just a few decades, the social sciences have been transformed from a discipline where experimental methods were considered impractical, ineffective, and largely irrelevant, to one where experimentation is considered a primary methodology (Guala, 2008).

Empirical research studies are designed to support arguments. In the case of an experiment understood as a study where researchers can observe behavior in an abstract environment that they control, the focus is on causal arguments (not simply correlation). Ideally, causality can be identified by exposing participants to different treatments (Charness, Gneezy, Kuhn, 2011). The goal here is to determine whether the intervention causes outcomes (however defined) to change, relative to what would have occurred without the intervention.

Describing more broadly the roles which experiments can play in empirical research, Roth argues that the first is "searching for facts", where the goal involves isolating the cause of some observed regularity. The second role entails "speak-

* Dr hab. Beata Woźniak-Jęchorek, prof. UEP – Poznań University of Economics and Business, Institute of Economics, Department of Macroeconomics and Development Research; ORCID: 0000-0002-5029-5885; e-mail: beata.wozniak-jechorek@ue.poznan.pl.

[1] This paper is a result of a 1-year Bekker scholarship at Northwestern University in the United States, financed by the National Agency for Academic Exchange in Poland. During this stay the author participated in Improving Evaluations of R&D in STEM Education 2021 Summer Institute, organized by the Institute for Policy Research at NU, and attended lectures on Experimental Political Science with James N. Druckman. The author is very grateful to Jamie Druckman for his helpful advice and insightful comments on this paper.

ing to theorists". Here the goal is to test the predictions (or the assumptions) of well-articulated formal theories. The third involves "whispering in the ears of princes". This facilitates "the dialogue between experimenters and policymakers" (Roth, 1995, p. 22).

The main aim of this paper is to outline the evolution of experimental economics, describe contemporary experimental methods, highlight the technological and institutional innovations that support experimentation, particularly in the United States, and identify the primary challenges that exist for the further development of this methodology.

The analysis conducted in this paper indicates that the role of experiments in modern economics is increasing, but still "the crucial question in social sciences is what experiments are for, why and when they should be used, and how they should be designed, implemented, and evaluated" (Druckman, 2021, p. 11). Furthermore, the data presented in this paper substantiates the notion that a majority of experiments originate from the United States. Approximately four times fewer papers based on the experimental approach were published by Chinese, British, and German scholars in the Scopus database. In CEE countries, this approach is less developed due to limited funding, cultural differences, different history of science, and different institutional infrastructure. The experimental approach in the U.S. was significantly influenced by the promotion of interdisciplinary research in American universities and the adoption of experiments as a universal language across diverse fields. Additionally, the emergence of crowdsourcing platforms facilitated less expensive data collection through surveys, further contributing to the growth of the experimental approach.

The paper is organized as follows. In the section 2, I present the expansion of experiments in economic sciences. I then focus, in the section 3, on the different modern types of experimental designs. The section 4 describes the technological and institutional innovations supporting experimentation. The final section 5 identifies the challenges in the contemporary experimental approach and section 6 concludes.

The motivation to write this paper lay in the conviction that this methodology can only be legitimized if its role, types, logic, and challenges are understood. Methodological issues are best covered within the context of the substantive research questions under investigation. However, a general overview is not only possible; it is required. This paper was also prepared for Eastern European students and scholars interested in contemporary experimental methodological literature.

2. Expansion of experiments in economics

The experimental approach has a long tradition. Comparative studies to determine whether a treatment or intervention has an effect have been carried out for decades (Roth, 1993; Hedges, Schauer, 2018). The basic principles of experimental research design were codified by R.A. Fisher in agricultural studies in the early 20th century (Fisher, 1921; Fisher, MacKenzie, 1923,

Fisher, 1935).[2] Like many methodological revolutions in science, the experimental turn in economics was primarily made possible not only by a change in philosophical perspective but by several innovations in scientific practice and theoretical commitment. At a very general level, economics was in the process of becoming a "tool-based" science in the mid-20th century, i.e. it was moving away from the old, discursive "moral science" of political economy and being transformed into a discipline where models, statistics, and mathematics were both instruments and, crucially, objects of investigation (Morgan, 2003). Theoretical models and computer simulations were first to the economist's basic toolkit, with laboratory experiments following shortly afterward.

The early contributors to experimental economics, e.g. Maurice Allais (of "Allais paradox" fame)[3] and Vernon Smith,[4] however, struggled to find an audience. Some areas, e.g. social dilemmas and bargaining experiments, were booming in psychology but had little impact on the economics literature (Leonard, 1994). Only the history of experimental economics in the 1980s and 1990s does the story of a booming research program become increasingly influential within the discipline and in social science at large. Powerful tools originated in psychology (Cook, Campbell, 1979), and created a boom through more sophisticated use of instrumental variables in economics starting in the 1990s (Angrist, Krueger, 2001), a little later in political science (Sovey, Green, 2011), and somewhat in parallel in computer science and statistics around 1995 (Pearl, 1995).

Together with game theorists, experimenters have also been increasingly involved in policymaking, notably by contributing to the design of new market institutions for the allocation of sensitive goods – from telecommunication licenses to space stations, airport slots, and physicians and surgeons (Roth, 2002). In 2002, Daniel Kahneman and Vernon Smith were awarded the Nobel Memorial Prize in Economic Sciences in recognition of their work as pioneers of experimental economics, effectively acknowledging this remarkable revolution officially. Some commentators, however, have seen two different styles of research in their work. Smith has developed novel experimental techniques to investigate traditional economic questions about the workings of markets ("experimental economics") and Kahneman, as a psychologist, has used the well-established experimental methods of his discipline to challenge conventional assumptions about the rationality of economic agents ("behavioral economics") (Bardsley et al., 2009).

Experimental economics experienced even more tremendous growth, both nationally and internationally, in the second decade of the 21st century (see Figure 1, which presents the growing number of papers based on the experimental approach

[2] His research on agricultural experiments used designs based on combinatorial principles, such as the Latin squares, to estimate the influence of some discrete factors (Fisher, 1935).

[3] The Allais paradox is a choice problem designed in 1953 to show an inconsistency of actual observed choices with the predictions of expected utility theory. According to Allais, von Neumann and Morgenstern's *Theory of Games and Economic Behavior* (1944) failed to capture the utility and probability functions of real decision makers.

[4] V. Smith had been experimenting since 1956, focusing on the properties of different market institutions and their effects on the convergence towards equilibrium (Smith, 1976, 1981).

in the Scopus database). The emergence of neuroeconomics, the significant growth in macroeconomic experiments (e.g. field experiments in development economics or experiments in political economy), and the substantial growth in experiments that support market design research or public goods and auctions have raised the status of the experimental approach in economic sciences and in the social sciences more generally. Recently, Nobel Memorial Prizes were awarded to Richard Thaler (2017) for incorporating psychologically realistic assumptions into analyses of economic decision-making, to Abhijit Banerjee, Esther Duflo, and Michael Kremer (2019) for their experimental approach to alleviating global poverty, and to David Card, Joshua Angrist, and Guido Imbens (2021) for providing the new insights into the labor market and for showing the conclusions that can be drawn about cause and effect from natural experiments. These were the most obvious signals that the experimental approach is now accepted as part of the discipline.

Covering the entire field of experimental economics in such a short paper is difficult. Bargaining, public goods provision, trust and reciprocity, workplace interaction, charity, choice under uncertainty, intertemporal choice and self-regulation, the neural circuity of social preferences, strategic thinking, new designs for spectrum auctions, and labor-market clearinghouses are only some examples of specific topics (eds. Kagel, Roth, 2016).

Experimental research is carried out by many economists around the world. Its results are routinely reported in major journals.[5] The dissemination of experimental methods has also been assisted by the *Journal of Experimental Economics*, *Experimental Economics,* and the *Journal of the Economic Science Association*. The last two are published by the Economic Science Association (ESA), a professional organization devoted to using controlled experiments to learn about economic behavior.

The experimental approach differs, however, between the Western and Eastern worlds (see Figure 2). China, Germany, and the Netherlands have the highest proportions of papers based on the experimental approach in the Scopus database. Nonetheless, in terms of the number of papers utilizing the experimental approach, the United States is unparalleled and holds the top position globally[6]. The dominance of American experimental data is also confirmed in databases in the experimental branches of psychology and cognitive science. These data show that Western, and more specifically American, undergraduates are the basic target research population. A perusal of the top journals in six sub-disciplines of psychology from 2003 to 2007 revealed that 68% of subjects came from the U.S.,

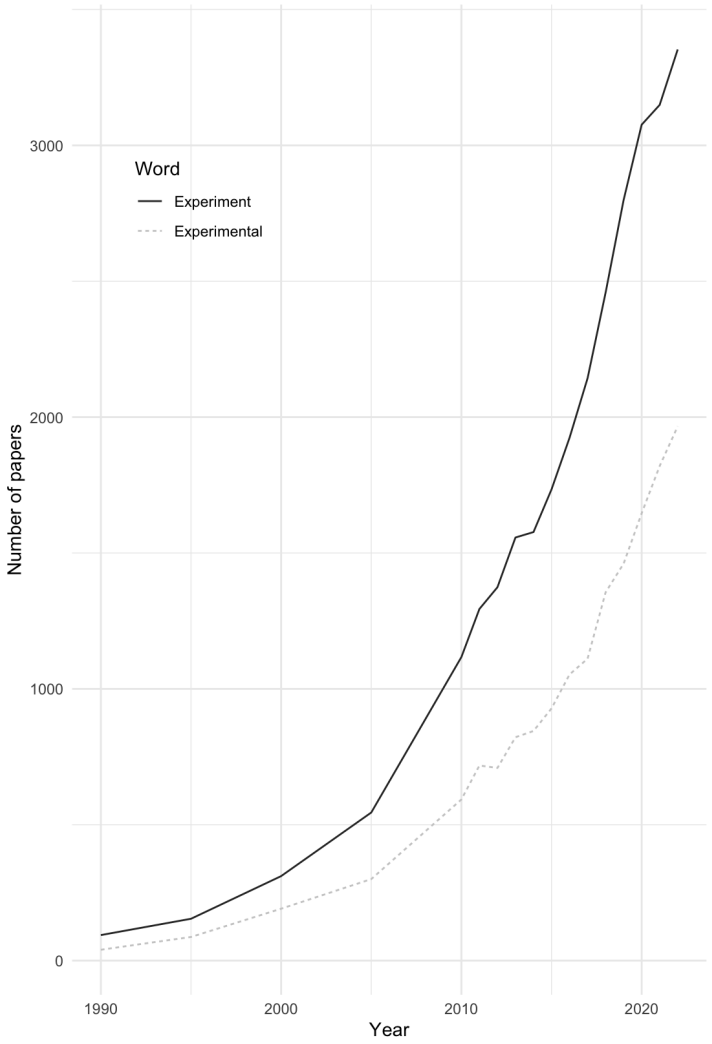
[5] The top journals in the Scopus database that publish experimental research are: *Journal of Economic Behavior and Organization* (1391 papers); *Journal of Economic Psychology* (544 papers); *Economic Letters* (518 papers); and *Experimental Economics* (351 papers). The organizations that provide the most funding for this kind of research are: National Natural Science Foundation of China (1607 papers); National Science Foundation in the U.S. (1100 papers); Deutsche Forschungsgemeinschaft (502); and the European Commission (458) (Scopus database).

[6] In the Scopus database, American academics have authored 12,851 papers with the term "experiment" and 6,527 papers with "experimental" appearing in the title, abstract, and/or keywords. In contrast, Chinese scholars have published only 4,609 and 2,073 such papers, respectively, while Polish scholars have contributed merely 308 and 187 publications to the database.

and a full 96% of subjects were from Western industrialized countries, specifically those in North America and Europe, as well as Australia and Israel (Arnett, 2008). From this, it follows that most of the data presented in the literature apply to a truly unusual group, viz. people from Western, Educated, Industrialized, Rich, and Democratic (WEIRD) societies (Henrich et al., 2010).

Figure 1

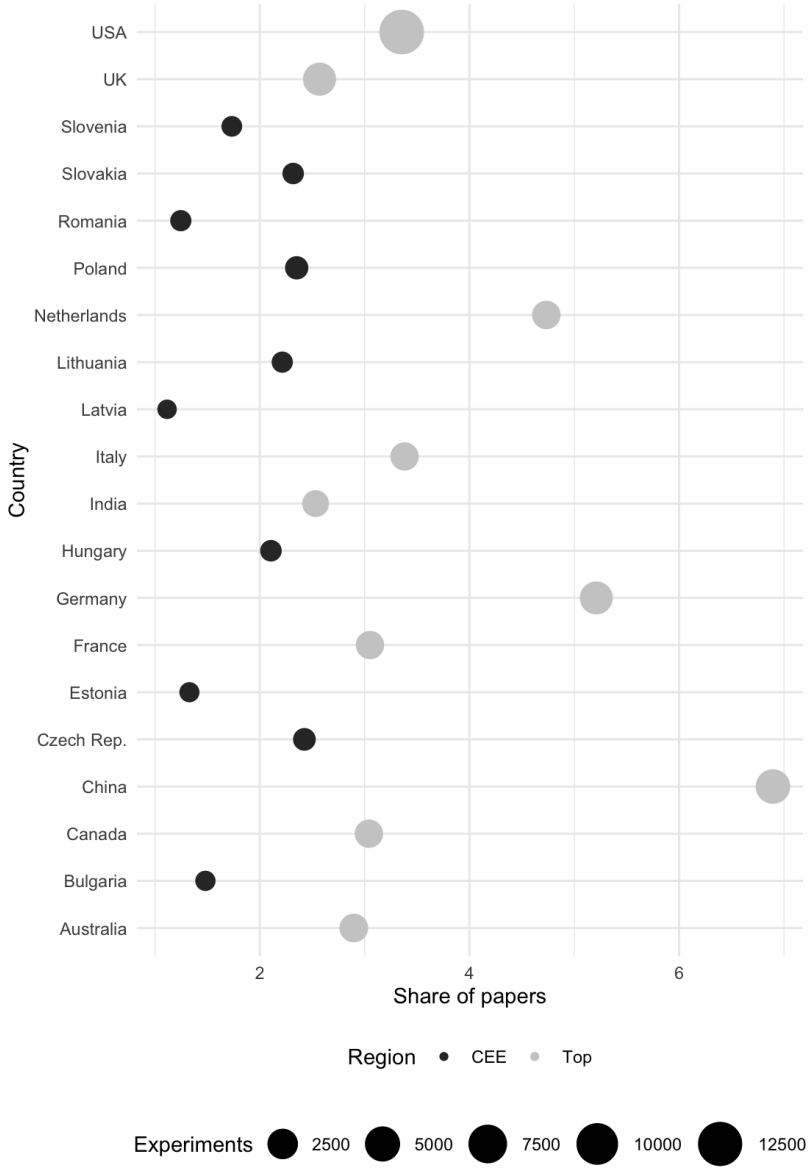
The Development of Experimental Research in Economics, Econometrics, and Finance



NB: Search based on the number of papers containing the words "experiment" or "experimental" in the article title, abstracts, and/or keywords.
Source: own elaboration based on Scopus database and journals classified in Economics, Econometrics, and Finance, data until 2022.

Figure 2

Share of papers on experiments in the total number of publications in the Top 10 and CEE countries



NB: Scopus papers based on Experimental Research in Economics, Econometrics, and Finance in the Top 10 countries and CEE countries by authors' affiliation. Search based on the number of papers containing the words "experiment" or "experimental" in the article title, abstracts, and/or keywords.
Source: own elaboration based on Scopus database and journals classified in Economics, Econometrics, and Finance.

There are also various possible reasons why experimental methods are comparatively less prevalent in Central and Eastern Europe (see Figure 3). For instance, researchers in CEE countries may face constraints in funding for their research projects. Additionally, there may exist cultural disparities in the approaches taken toward research in this region, resulting in greater emphasis being placed on alternative empirical research methods. Furthermore, the history of science in CEE has been shaped by multiple factors such as political and economic changes, which may have influenced the prioritization of particular research programs. Lastly, conducting experimental research often necessitates access to specialized equipment and facilities, which may be insufficiently developed in CEE countries to support such research activities.

Experimental economics is, however, assessed as a useful tool in modern comparative economics. This is clearly an argument in favor of their further development in CEE countries. Different historical perspectives and a variety of institutional backgrounds with differing qualities of formal and informal institutions can be tested – especially in policy-oriented experiments. Testing the causal effect of a precise intervention, such as a particular campaign message or a curriculum, and stimulating a particular "transition reality" should matter such as "institutions matter" – "the rules which determine the information states and individual incentives in the trading game" (Smith, 1994, p. 116). Therefore, it is worth building a better-grounded understanding of the modern types of experimental designs and their methodological challenges.

3. Modern types of experimental designs

The ideal experimental research design would be to take a given unit (e.g. person, country) and assess the impact of a variable (e.g. exposure = treatment; no exposure = control) at a single place and a single point in time: outcome (treatment, unit) versus outcome (control, unit) (Druckman, 2021, p. 46). This is the idea on which scholars developed laboratory, survey, and field experiments as the dominant types of experiments in the modern social sciences.

The first type, viz. laboratory experiment, is an experiment where the intervention occurs in a controlled setting (laboratory). Laboratory experiments are conducted in a secure location where the researcher carefully controls the subjects' experience. Subjects go to the same place, hear the same instructions, and are typically exposed to identical stimuli, save for key differences that manipulate the independent variable of interest (Arceneaux, 2010).

One novel design that has gained prominence in many social science applications over the last decade is lab-in-the-field studies. These are lab designs (administered and measured in a controlled setting), but where the intervention (e.g. treatment, payment structure) is conducted in naturally occurring situations. That is to say that the lab is constructed in the field to reach populations or contexts that would be difficult to reach in a conventional lab (e.g. the researcher travels to the participants to collect data in their town, such as at a community center, using portable computers or tablets) (Gilligan et al., 2014).

The second type is survey experiments which occur when the intervention comes as part of the survey, for example, in an experiment where some subjects receive a particular wording of a question while others receive a different wording. The surveys could be in-person, via the phone, or online. A researcher randomly assigns participants to at least two experimental conditions. The researcher then treats each condition differently. Due to random assignment, the researcher can assume that the only difference between conditions is the difference in treatment.

The most familiar types of survey experiments are vignette survey experiments and conjoint designs.[7] Vignette studies use short descriptions of situations or persons (vignettes) that are usually shown to respondents in order to elicit their judgments about these scenarios (Atzmüller, Steiner, 2010). In a typical vignette, the researcher varies only one component of the scenario. There also exists a factorial experiment, where the researcher varies several components of the scenario (Auspurg, Hinz, 2014). The respondents are asked to choose from or rate hypothetical profiles described with multiple attributes. The aim is to gauge the relative impact of each attribute on choice (Hainmueller et al., 2014). The point of the exercise is to evaluate the difference(s), if any, when aspects of the "story" change (Mutz, 2011). The most important conjoint advantage is that this type of experiment can vary many more factors than a vignette.

Finally, the third type of experiment is a field experiment in which participants receive the treatment in naturalistic settings, as part of their daily lives, typically without their knowing that it is an intervention. The key advantage of the field experiment is that the effects can be assessed in naturalistic contexts. The challenges are mostly connected with compliance (whether everyone in the treatment group receives the treatment), attrition (often, not every participant that begins a study persists; quitting may be related to both the intervention and the outcome), spillover (whether those in the treatment talk to those in the control) and external validity (generalization) (Gerber, Green, 2012). The naturalistic nature of these experiments makes them attractive to study market institutions (micro-finance), policy intervention, discrimination, mobilization, etc. Audit field experiments have become the most popular method for discrimination studies (Baert, 2017, Neumark, 2018; Neumark, Burn, Button, 2019). The point is here to "audit" a market for bias.[8]

---

[7] A list experiment (Kuklinski et al., 1997), a priming experiment (Macrae et al., 1994), and endorsement experiments (Cohen, 2003; Bullock, Imai, Shapiro 2011) can be used to measure sensitive topics.

[8] Researchers send out fictitious or real but controlled applications that are identical except for randomly varied dimensions of interest (e.g. race, religion, ethnicity, age, gender, disability, criminal record, immigration status, mental health, military service, parental status, physical appearance, sexual orientation, social class, employment status) and put applicants into a social situation (in person or not – via phone or online) such as e.g. a job interview, job offers, housing inquiry responses, response to roommate requests, doctor's appointment scheduling, responsiveness from bureaucrats or elected officials, responsiveness from professors, or the price paid for bargained goods. The response rate is then compared between groups (Gaddis, 2018).

Experiments can also be classified on the basis of randomization. Random assignment of individuals to treatment assures that, on average, there is a matching of all known and unknown variables. This profound advantage explains why randomized experiments have attained the status of the most powerful tool ("gold standard") for causal inference. Randomized experiments can be divided into random assignment experiments, random lottery natural experiments, and "as-if random" assignment experiments.

The first type, viz. random assignment experiment, constitutes the dominant approach in many disciplines. Here, the researcher randomly assigns values of the independent variable and then measures the outcome. This is a straightforward approach, as the researcher simply compares the average outcome values between the treatment and control. There are various types of this sort of design, as it involves an individually randomized experiment focusing on a situation in which individuals within one or more groups are randomly assigned to the intervention or comparison. The main types are block randomized experiments, where randomization occurs within groups, and cluster randomized experiments, where larger units are randomly assigned to receive an intervention or comparison. Importantly, the level of randomization is typically chosen on the basis of the type of intervention and the manner in which it is delivered, as well as concerns with the sensitivity of the anticipated treatment effects.

The second type, viz. random lottery natural experiment, assumes that "real world" events generate the random allocation of units (individuals) into treatment and control groups. However, as Titiunik states, the term natural experiment is sometimes used inconsistently. In one interpretation, it refers to an experiment where a treatment is randomly assigned by someone other than the researcher. In another interpretation, it refers to a study in which there is no controlled random assignment, but treatment is assigned by some external factor in a way that loosely resembles a randomized experiment – often described as an "as-if random" assignment (Titiunik, 2021). Druckman calls this type a "unit homogeneity natural experiment" ("as-if" randomization) (more Druckman, 2021).

Sometimes randomization cannot be used. In these situations, research methods in which the comparison groups are not assigned randomly have to be designed. Such designs are often designated quasi-experimental to reflect the fact that they mimic certain characteristics of experimental designs, such as the involvement of comparison groups.

Quasi-experiments can be "within-subject"[9] or "induced value". These involve observing the same unit over time and thus satisfying the assumptions that prior exposures and/or measurements do not matter (i.e. temporal stability and causal transience assumptions) or comparing non-randomly assigned units that are

[9] There also exists a "between-subject" designed experiment, where everyone is exposed to only one treatment classified among random assignment experiments. With these types of designs, if group assignment is random, causal estimates are obtained by comparing the behavior of those in one experimental condition with the behavior of those in another (Charness, Gneezy, Kuhn, 2011).

presumed to be identical in every way but for treatment assignment (Shadish, Cook, Campbell, 2002).

In a "within-subject" designed experiment, everyone is exposed to more than one of the treatments being tested, whether it be playing a game with two different parameter values, being treated and untreated, answering multiple questions, or performing tasks under more than one external stimulus. With such designs, if the multiple exposures are independent, causal estimates can be obtained by examining how individual behavior changes when the circumstances of the experiment changes (Charness, Gneezy, Kuhn, 2011).

In "induced value" experiments, the researcher tests the unit homogeneity assumption by using e.g. financial incentives. In these experiments, the researcher induces pre-specified characteristics in the participants so that their "innate characteristics become largely irrelevant" (Friedman, Sunder, 1994, p. 13). This is effected by offering an award medium (i.e., money). Induced value experiments were introduced by V. Smith in the 1970s. According to the theory of induced valuation, "control is the essence of experimental methodology, and in experimental exchange studies one must be able to state that, as between two experiments, individual values (e.g., demand or supply) either do or do not differ in a specified way. Such control can be achieved by using a reward structure to induce prescribed monetary value on actions" (Smith, 1976, p. 275). Induced value experiments are still widely used and expected in current economics.

4. Technological and institutional innovations supporting experimentation

In the last couple of years, the experimental methods presented above have grown from nothing to become the primarily accepted methods in many social science disciplines. This has been made possible by massive technological advances in computing technology, such as crowdsourcing platforms, Internet panels, social media access to behavioral data, elite samples via e-mail, public data repositories, and the ability to analyze high-dimensional data. These advances have all facilitated data access and analysis (Bond et al., 2012; Druckman, 2021).

While academics have been trying to improve their data collection practices, several unresolved issues remain, e.g. inconsistent data collection standards, complexity, lack of training in data collection, lack of quality assurance processes, changes to definitions and policies, and maintaining data comparability. New IT infrastructure and institutional support have been created to rectify these problems.

A few examples of this technology are worth noting. Currently, data can be easily retrieved from different crowdsourcing platforms, thereby outsourcing tasks (e.g. survey participation) to a workforce who can perform it virtually. Platforms such as Amazon Mechanical Turk (MTurk), Prolific, and Lucid offer access to a global, on-demand workforce, most of whom speak in English.

Data from experiments can (and more often should - according to the policies of international journals) be shared in public data repositories, such as Dataverse and

GitHub (Druckman, 2021, p. 22). There exist several Dataverse repositories installed in universities and organizations around the world, e.g. the Austrian Social Science Data Archive (AUSSDA), British Columbia Research Libraries' Data Service (ABACUS), and CIRAD Dataverse in France. The Dataverse Project was, however, housed and developed at Harvard University and is open to all scientific data from all disciplines worldwide as a repository for sharing, citing, analyzing, and preserving research data (https://dataverse.harvard.edu). Another example is GitHub – the largest and most advanced development platform in the world with more than 200 million repositories. This platform connects developers and companies operating in English, especially from the U.S., Asia, and Western Europe (https://github.com).

Platforms, repositories, and the widely used z-Tree package designed by Urs Fischbacher (2007) have greatly reduced the costs and the investment in skills necessary to run computerized experiments.

To provide financing for the survey experiments, the Time-sharing Experiments for the Social Sciences (TESS) platform was created in the U.S. This interdisciplinary initiative was established in 2001 with support from the National Science Foundation. TESS capitalizes on economies of scale to enable scholars from across the social sciences, on a competitive basis, to conduct survey experiments on probability-based samples of adults in the U.S. (NORC's AmeriSpeak® Panel) (Mutz, 2011). Since its founding, TESS has supported more than 400 experiments. Many of them are published in disciplinary flagship journals and *Science* and the *Proceedings of the National Academy of Science.* TESS also makes raw data from all experiments publicly available, regardless of whether the results are published.

In order to protect institutional liability in academia, Institutional Review Boards (IRBs) were introduced to ensure that experimental studies meet ethical standards. While several scholars have pointed out that IRBs do not protect the individual researcher from opprobrium (King, Sands, 2015), these committees are treated as an element of experimental credibility.

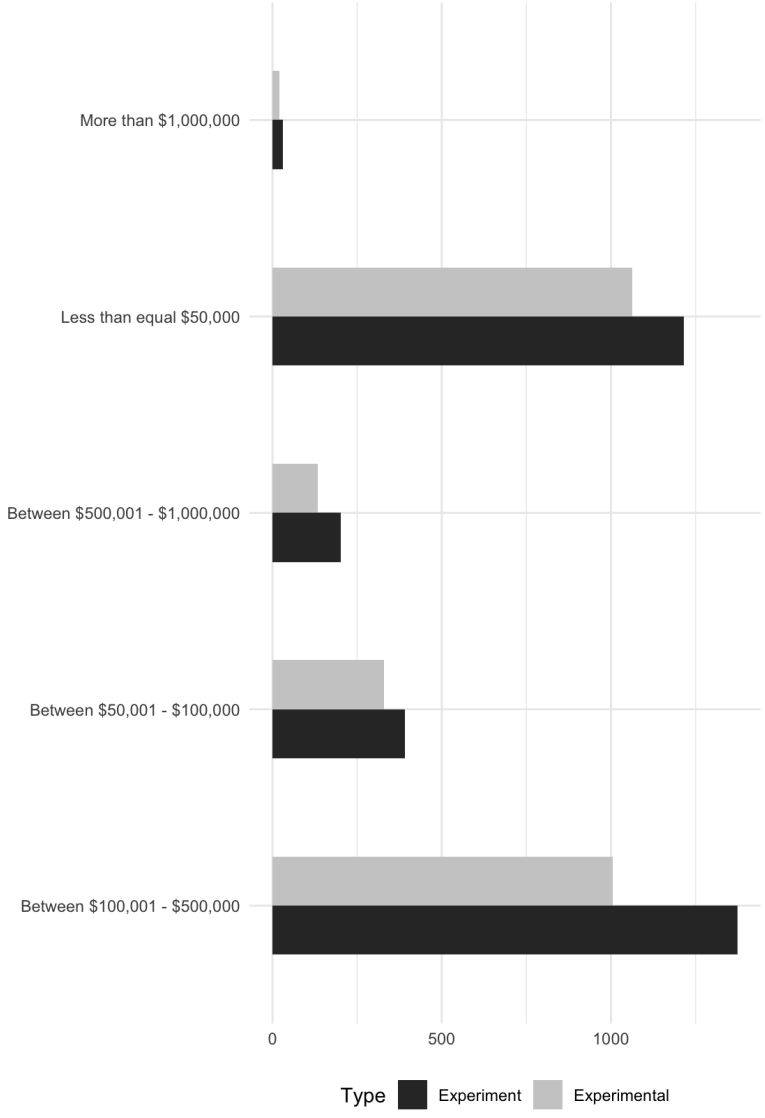
The other institutional requirement is the preregistration of the experimental study. There are several options, e.g. the Wharton Credibility Lab of the University of Pennsylvania, which has created the "Aspredicted" (https://aspredicted.org/index.php) platform or the Center for Open Science (https://www.cos.io/initiatives/prereg). Preregistration simply means specifying the research plan and separating *hypothesis-generating* (exploratory) from *hypothesis-testing* (confirmatory) research. Both are important in the research process, but the same data cannot be used to generate and test a hypothesis, which can happen unintentionally and vitiate the credibility of the results. Addressing this problem through planning improves the quality and transparency of research.

The development of experimentation in the U.S. would not have been possible without the support of funding organizations. The American public funding organization – the National Science Foundation – provides data on the number of standard grants and amounts awarded for Social, Behavioral, and Economic Sciences. Querying this database in 2021 revealed 3213 grants for projects having "experiment" in the keywords and 2551 grants for projects having "experimental" in the keywords. Most grants were between $100,001-500,000 and less than $50,000

(see Figure 3). The data show the scale of experimental projects conducted in the U.S. and their average cost. The number of grants below $50,000 may seem surprising. The low costs of these projects are due to access to cheap sources of data acquisition from crowdsourcing platforms, Internet panels, social media contacts, and elite samples via e-mail.

Figure 3

Standard Grants for Social, Behavioral, and Economic Sciences in the U.S.



NB: current award amount available in August 2021.
Source: own elaboration based on data from National Science Foundation (https://www.nsf.gov/awardsearch)

5. Challenges in the development of an experimental approach

There are several challenges that researchers and scientists face when conducting experiments. These include the increasing complexity of systems with many variables and interactions. This can make it difficult to design experiments that accurately test a specific hypothesis. Experiments can also be expensive and time-consuming to conduct, and researchers often have limited resources, including funding and personnel. Some experiments may raise ethical concerns, such as those involving human subjects or animals. Researchers must ensure that they follow all relevant ethical guidelines when conducting their experiments. There is also a growing concern about the reproducibility of scientific experiments, as many studies are difficult or impossible to replicate. This can undermine confidence in the scientific community and make it difficult to build on previous research. Finally, with the increasing amount of data being generated in scientific experiments, researchers must have effective systems in place for managing and analyzing this data. Bearing in mind that experiments are difficult to conduct, it is worth focusing more on the challenges related to the appropriate research design, realism, ethics, interdisciplinarity, publication bias, transparency, and replication.

Good research design

Each type of experiment offers many possibilities but also has several limitations and is only useful in specific circumstances. There is a large literature on the pros and cons of each type of experiment.[10] During the initial testing of an intervention, it is difficult to quantify its benefits and determine whether they outweigh the costs. As Druckman states "experiments are useful only if there exists a substantively grounded question, a well-defined target population, carefully constructed measures, and clear points of comparisons. Many extant experiments fail to explicitly consider these issues" (Druckman, 2021, p. 13).

Therefore, before collecting data, the experimenter should scrupulously specify the goal of the experiment, identify the target population, and justify the sampling approach.[11] Next, it is important to construct valid[12] and accurate[13] measures,[14]

---

[10] E.g., in case of conjoint see McFadden 2017; Ben-Akiva et al. 2019; in case of audit experiment see Bischof et al., 2021, in case of field experiment see Teele, 2014.

[11] There is a large literature on sampling (e.g. Groves et al., 2009; Beimer, 2010; Blair and Blair, 2015). Identifying the population is often overlooked (Westreich et al., 2019, p. 439). Sometimes scholars also sample the context (settings), topic, and measures (e.g. treatments/outcomes) (Shadish, Cook, Campbell, 2002, pp. 23, 69–72).

[12] Validity concerns the extent to which the measure/quantification reflects the abstracted concept. Shadish, Cook and Campbell (2002) define internal validity, external validity, statistical conclusion validity and construct validity.

[13] Accuracy means that measures need be also reliable and unbiased.

[14] Most experiments seek to measure the causal effect of a treatment (i.e. the independent variable) on an outcome (i.e. the dependent variable).

posit reasonable assumptions, and select an appropriate research design.[15] After that, the key comparison groups have to be specified and theoretically justified, the research design has to be compared with those of related studies, and the implications of choosing a particular design for generalization have to be considered. Finally, if heterogeneous effects are expected, the experimenter should think about blocking and increasing the sample size. The final step is to consider all these challenges in the context of cost and research budget.

The sensitivity of a research design is determined by many factors, e.g. sample size, reliability and validity of measures, experimental error, subject variability, strength and integrity of treatment, and type of statistical analysis (Lipsey, 1990, pp. 9–10). There are three basic ways (matching, randomization, and statistical adjustment) that can be employed to create appropriate research designs for detecting causal relationships. All of them start with the idea of controlling variation. From a statistical standpoint, empirical research always evaluates observed relationships ("estimates") to variation that is presumably a consequence of other variables that are not directly of research interest. The former is often called systematic variation while the latter is often called noise or error variation. This variation might be between groups, e.g. those receiving the intervention and those not, or within groups. Of course, what counts as systematic and what counts as irrelevant, or error variation depends on the research problem (Cronbach, Meehl, 1955).

Of these three methods of increasing the research design sensitivity, randomization is the most popular, although there are a variety of theoretical objections to randomized experiments. Some scientists argue that experiments depend on an oversimplified theory of causation or epistemology, or they frequently fail and do not find effective causality. Others argue that they only work in specific circumstances or when certain assumptions are adopted, or that they are simply unethical (see e.g. Hegtvedt, 2014; McDermott, Hatemi, 2020).

Realism

The design, implementation, and analysis of experiments raise a variety of distinct epistemological and methodological challenges. One of them is the question about representation and isolation. As Mäki suggests "an experiment is an arrangement seeking to isolate a fragment of the world by controlling for causally relevant things outside that fragment. Models are substitute systems that are directly examined to indirectly acquire information about their target systems. It is suggested that many theoretical models are ('thought') experiments and that many ordinary experiments are ('material') models. The major difference between the two is that

[15] According to this characteristic, an experiment is a study where an intervention provides the primary mechanism by which a researcher uses a procedure to resolve the Fundamental Problem of Causal Inference (FPCI), where "it is impossible to observe the value of (the treatment outcome) and (the control outcome) on the same unit…" (Holland, 1986, p. 947). To solve this problem, Holland proposed using the scientific solution and the statistical approach (see Holland, 1986; Druckman, 2021).

the controls affecting the required isolation are based on material manipulation in one case, and on assumptions in the other" (Mäki, 2005).

The question of representation is closely related to concerns over realism and the contribution of a particular experiment to the general knowledge (there is a large literature on this topic, e.g. Hacking, 1984; Resnik, 1994). There are two ways to assess this realism. First, experimental realism refers to whether "an experiment is realistic, if the situation is involving to the subjects if they are forced to take it seriously, [and] if it has an impact on them" (Aronson et al., 1985, p. 485). Second, mundane realism concerns "the extent to which events occurring in the research setting are likely to occur in the normal course of the subjects' lives, that is, in the 'real world' (Aronson et al., 1985, p. 485). Much of the debate about experiments revolve around mundane realism, but experimental realism is far more important. The failure of participants to take the study and treatments seriously compromises the basis of the causal inference, which renders the experimental results meaningless (see e.g. Dickhaut et al., 1972). Moreover, as a given experiment aims to identify the effect of a given variable in isolation, it inevitably appears different from the multidimensional nature of "real life". Designing valid and accurate treatments and focusing on experimental realism should therefore play a crucial role in experimentation (Druckman, 2021).

Interdisciplinarity

Many of the important research questions facing society today are complex and cannot be adequately addressed by a single discipline. Interdisciplinary research allows scientists to bring a variety of perspectives and approaches to bear on these questions. The experimental approach was a response to a growing need to transform the social sciences and dismantle disciplinary boundaries in order to deal with the pressing social challenges of the 21st century. Thinking collectively about complex problems requires crossing boundaries both horizontally (across disciplines) and vertically (across experts, policymakers, practitioners, and the public). Increasing the exchange of methods and knowledge across fields has resulted in a need for a common language across disciplines. Experiments have become this interdisciplinary language, which has its origins in individual disciplines, e.g. an understanding of culture from anthropology, rational choice theory from economics, ideology from political science, laboratory experimental methods from psychology, and social networks from sociology (Buyalskaya, Gallo, Camerer, 2021).

Lélé and Norgaard (2005) have identified four major types of barriers to interdisciplinarity. First, there is the problem of values being embedded in all types of inquiry and at all stages: in the choice of questions, theoretical positions, variables, and style of research. Second, researchers in different disciplines may study the same phenomenon but differ in their theories or explanatory models (and underlying assumptions). The third type of barrier is the differences in epistemology and hence in specific methods, notions of adequate proof, and other fundamental assumptions of different fields. Finally, how society interacts with and organizes

academia influences the production of interdisciplinary research (Lélé, Norgaard, 2005, p. 968). This last argument seems to be especially important in many countries.

As Schoenberger (2001) has pointed out, the relative importance or validity of a direction of inquiry or approach is not determined simply by some objective recognition by academics of its ability to generate more valid knowledge than another approach. Forces at work in the larger society outside academia shape the perception of importance gained by a certain discipline, or by a particular kind of interdisciplinary crossing. The society also influences the institutional arrangements within academia that create incentives or disincentives for interdisciplinary knowledge production (Lélé, Norgaard, 2005, p. 968).

These incentives appear to be most developed in the U.S. Several factors may contribute to the trend toward interdisciplinary research programs in American academia. The U.S. has a culture of collaboration and a strong research infrastructure that facilitates the formation of interdisciplinary research teams. There is also a strong demand from society for research that addresses real-world problems and has practical applications. Interdisciplinary research is well-suited to addressing these types of problems and is more likely to receive funding and support from policymakers and the public. Therefore, funding agencies, such as the National Science Foundation, often encourage interdisciplinary research and provide funding specifically for interdisciplinary projects.

Ethics

For five decades, research ethics have been guided by a set of criteria known as the Belmont Report and the principles of beneficence, respect for persons, and justice. The idea behind these principles is that participation in research should always be voluntary, that researchers should always secure informed consent from participants (where informed implies being adequately briefed about potential risks), and that researchers should have to be able to say whether the findings of the study are relevant to the population on which it is to be carried out to ensure that they are not burdening the poor, the old, the very young, or the weak unless necessary for the scientific findings (Teele, 2021).

In practice, there are three ways to classify experiments in terms of the type of participation they rely on and the quality and amount of information that participants are given. The first is the "gold standard". This requires that participation be voluntary and information about the treatments involved, along with the potential risks, are articulated as clearly as possible. The "silver standard" emerges when the research is voluntary, but the participants are not given complete information (e.g. medical experiments with placebos, lab experiments with games, psychological experiments, survey research with experiments embedded). The third is the "bronze standard". Here, neither voluntary participation nor full information apply. This produces a situation in which consent is "virtual" (e.g. audit experiment, informational experiment, correspondence experiments, radio and television experiments, experiments where randomization is at a higher level of

aggregation). An overwhelming number of today's experimental designs in many fields are in the bronze realm and deny the Belmont principles.

Moreover, Teele (2014) has argued that there is no ethical tradeoff between respect for persons and measurement, and that a new ethical framework, specially designed for experimental, virtual studies, is needed. Virtual consent, a situation in which the researcher consents on behalf of the subject population, could only be ethical if a reasonable person would agree to participate if he/she had been informed (Teele, 2021). In practice, however, researchers follow the argumentation that ethics only requires an assessment of the potential risks and benefits to the participants and society, and if the benefits outweigh the risks, then the experiment should be carried out.[16].

Publication bias, transparency, and replication

Publication bias occurs when the publication of study results is based on the direction or significance of the findings. One pernicious form of publication bias is the greater likelihood of statistically significant results being published than statistically insignificant results, holding fixed research quality. Selective reporting of scientific findings is often referred to as the "file drawer" problem. Such a selection process increases the likelihood that published results reflect Type I errors rather than true population parameters, biasing effect sizes upwards of.[17] Further, it constrains efforts to assess the state of knowledge in a field or on a particular topic because null results are largely unobservable to the scholarly community (Franco, Malhotra, Simonovits, 2014). To solve the publication bias issue, academia has proposed such solutions as the two-stage review (the first stage is for the design and the second for the results), pre-analysis plans, and requirements to preregister studies. All these proposals should be complemented by incentives not to bury statistically insignificant results in "file drawers". The movement toward open-access journals may provide space for such articles.

The open science movement (free online access to scientific and scholarly research literature, especially peer-reviewed journal articles and their preprints) concentrates its efforts on transforming scientific practice to enhance the transparency, productivity, and reproducibility of research (Elliott, Resnik, 2019). Transparency makes research more comprehensible, allows it to be subjected to public scrutiny, and enables future research to build on it.[18] The situation with

[16] When defending experiments with random assignment, it is worth stressing that this kind of experiment generally requires a smaller sample size to obtain a design with adequate sensitivity. Furthermore, because randomized experiments have higher internal validity than other designs, the validity of the evidence provided by randomized designs is also higher, meaning the benefits are greater than would be achieved by other designs.

[17] Berinsky, Druckman, and Yamamoto (2021) use three types of publication biases: (1) file drawer bias; (2) a "repeat study" bias against the publication of replication studies; and (3) a "gotcha bias" where replication results that run contrary to a prior study are more likely to be published.

[18] See website https://topfactor.org/ which rates the transparency policies of journals.

transparency is, however, a classic collective action problem. Many individual researchers lack strong incentives to be more transparent, even though the credibility of science would benefit if everyone were more transparent. As Nosek et al. (2015) state, "unfortunately, there is no centralized means of aligning individual and communal incentives via universal scientific policies and procedures. Universities, granting agencies, and publishers each create different incentives for researchers."

The other challenge is the reproducibility of science. Scientific claims should not gain credence because of the status or authority of their originator but because of the replicability of their supporting evidence (Casadevall, Fang, 2010, p. 4972). If an experimental result has succeeded in revealing a real process or effect, then that success should be replicated when the experiment is repeated, whether by the same experimenter in the same lab ("repeatability") or by others, elsewhere, using equivalent procedures ("reproducibility"). Replications can correct initial study publication bias and facilitate research transparency by requiring that the authors make their study materials public (e.g. Nosek et al., 2015). There have been several large-scale replication studies (Mullinix et al., 2015; Open Science Collaboration, 2015; Camerer et al., 2016, 2018; Coppock et al., 2018; Coppock, 2019) and ongoing debate about the existence of a "replication crisis" (cf. Baker, 2016; Fanelli, 2018). Sometimes replication can be hard to achieve, either because of the difficulty of replicating pertinent conditions or simply through a lack of institutional rewards to the replicating experimenters (Norton, 2015, p. 230). In the last case, reproducibility is not well understood because the incentives for individual scientists prioritize novelty over replication, and journal reviewers and editors may dismiss a new test of a published idea as unoriginal. As Open Science Collaboration states, "innovation points out paths that are possible; replication points out paths that are likely; progress relies on both. Replication can increase certainty when findings are reproduced and promote innovation when they are not" (2015). Therefore, the promotion of reproducibility must be seen as a core principle of scientific progress.

6. Conclusion

Lipsey once stated that "experimental sciences, such as chemistry and some branches of psychology, have an advantage because it is possible to produce relevant evidence through controlled laboratory experiments. Other sciences, such as astronomy and economics, cannot do this" (Lipsey, 1979, p. 8). In microeconomics, the focus was on explaining and predicting the values of statistics of aggregate market behavior, particularly prices and total quantities traded. Macroeconomics worked at an even higher level of aggregation. Thus, the useful predictions of economics operated at a level at which, it was thought, direct experimental tests would be enormously costly and perhaps even unethical (Bardsley et al., 2009).

Half a century later, scientists began to use large-scale experiments to answer broad questions about cause and effect in economic life. From natural experi-

ments where situations arise in real life that resemble randomized experiments, to laboratory experiments where the variable can be controlled in isolation or in various combinations, the benefits of experiments are now well-known and commonly accepted in social science.

The widespread use of experimental methods in economics and the social sciences generally has been made possible by the development of technological innovations. Advanced statistical methods, the use of new measurement techniques, and the explosion of new data sources have changed the perspective in modern research. Nonetheless, the surge in experimental research has not been evenly distributed across the globe, as the employment of experimental methods in Eastern Europe seems to have remained stagnant. It is evident that there is a necessity to increase awareness of the advantages of these methods in this region. To achieve this goal, various requirements must be met, such as the provision of institutional support from funding organizations, the establishment of interdisciplinary research programs, access to cost-effective data, transparency, and the reduction of publication bias.

References

Angrist, J. D., Krueger, A. B. (2001). Instrumental variables and the search for identification: From supply and demand to natural experiments. *Journal of Economic Perspective*, 15, 69–85.

Arceneaux, K. (2010). The Benefits of Experimental Methods for the Study of Campaign Effects. *Political Communication*, *27*(2), 199–215. https://doi.org/10.1080/10584601003709407.

Arnett, J. (2008). The neglected 95%: Why American psychology needs to become less American. *American Psychologist*, *63*(7), 602–14.

Arsonson, E., Brewer, M. B., Carlsmith, J. M. (1985). Experimentation in Social Psychology. W: Lindzey, G., Aronson, E.(eds.), *Handbook of Social Psychology*, 3rd Edition. New York: Random House.

Atzmüller, C., Steiner, P. M. (2010). Experimental vignette studies in survey research. *Methodology: European Journal of Research Methods for the Behavioral and Social Sciences*, *6*(3), 128–138. https://doi.org/10.1027/1614-2241/a000014.

Auspurg, K., Hinz, T. (2014). *Factorial Survey Experiments*. Vol. 175. Sage Publications.

Baert, S. (2017). Hiring Discrimination: An Overview of (Almost) All Correspondence Experiments Since 2005. *IZA Discussion Paper Series*, 10738.

Baker, M. (2016). *Is There a Reproducibility Crisis?. Nature*, 533, 452–455.

Bardsley, N., Cubitt, R., Loomes, G., Moffatt, P., Starmer, Ch., Sugden, R. (2009). *Experimental Economics: Rethinking the Rules*. Princeton University Press.

Ben-Akiva, M., McFadden, D., Train, K. (2019). Foundations of stated preference elicitation: consumer behavior and choice-based conjoint analysis. *Foundations and Trends in Econometrics*, 10, 1–144.

Berinsky, A. J., Druckman, J. N., Yamamoto, T. (2021). Publication Biases in Replication Studies. *Political Analysis*, 29, 370–384.

Biemer, P. P. (2010). Total Survey Error: Design, Implementation, and Evaluation. *Public Opinion Quarterly*, 74, 817–848.

Bischof, D., Cohen, G., Cohen, S., Foos, F., Kuhn, P. M., Nanou, K., Visalvanich, N., Vivyan, N. (2021). Advantages, Challenges and Limitations of Audit Experiments with Constituents. *Political Studies Review*, 1–9, SAGE. https://doi.org/10.1177/14789299211037865.

Blair, E., Blair, J. (2015). *Applied Survey Sampling.* Los Angeles, CA: Sage Publications.

Bond, R. M., Fariss, Ch. J., Jones, J. J., Kramer, A. D. I., Marlow, C., Settle, J. E., Fowler, J. H. (2012). A 61-Million-Person Experiment in Social Influence and Political Mobilization. *Nature*, 489, 295–298.

Bullock, W., Imai, K., Shapiro, J. N. (2011). Statistical Analysis of Endorsement Experiments: Measuring Support for Militant Groups in Pakistan. *Political Analysis*, *19*(4). Cambridge University Press, 363–84.

Buyalskaya, A., Gallo, M., Camerer, C. F. (2021). The golden age of social science. *PNAS*, *118*(5).

Callard, F., Fitzgerald, D. (2015). *Rethinking interdisciplinarity across the Social Sciences and Neurosciences*. London: Palgrave Macmillan.

Camerer, C.F., Dreber, A., Holzmeister, F., Ho, T.H., Huber, J., Johannesson, M., Altmejd, A. (2018). Evaluating the replicability of social science experiments in Nature and Science between 2010 and 2015. *Nature Human Behavior*, 2, 637–644.

Casadevall, A., Fang, F. C. (2010). Editorial: Reproducible Science. *Infection and Immunity*, 78, 4972–4975.

Charness, G., Gneezy, U., Kuhn, M. A. (2011). Experimental methods: Between-subject and within-subject design. *Journal of Economic Behavior & Organization*, *81*(1), 1–8.

Cohen, G. L. (2003). Party over Policy: The Dominating Impact of Group Influence on Political Beliefs. *Journal of Personality and Social Psychology*, *85*(5), 808.

Cook, T. D., Campbell, D. T. (1979). *Quasi-Experimentation: Design & Analysis Issues for Field Settings*. Houghton Mifflin.

Coppock, A. (2019). Generalizing from Survey Experiments Conducted on Mechanical Turk: A replication Approach. *Political Science Research and Methods*, 7, 613–628.

Coppock, A., Leeper, T. J., Mullinix, K. J. (2018). Generalizability of Heterogenous Treatment Effect Estimates across Samples. *Proceedings of the National Academy of Sciences*, 115, 12441–12446.

Cronbach, L. J., Meehl, P. E. (1955). Construct validity in psychological tests. *Psychological Bulletin*, *52*(4), 281–302.

Dickhaut, J. W., Livingstone, J. L., Watson, D. J. H. (1972). On the Use of Surrogates in Behavioral Experimentation. *The Accounting Review*, 47, Supplement: 455–471.

Druckman, J. N. (2021). *Experimental Thinking: A Primer on Social Science Experiments*. New York: Cambridge University Press, Forthcoming.

Druckman, J. N., Green, D. P. (2021). A New Era of Experimental Political Science. W: Druckman, J. N., Green, D. P. (eds.). *Cambridge Handbook of Advances in Experimental Political Science*. New York: Cambridge University Press.

Elliott, K. C., Resnik, D. B. (2019). Making Open Science Work for Science and Society. *Environmental Health Perspective*, *127*(7).

Fanelli, D. (2018). Is Science Really Facing a Reproducibility Crisis, and Do We Need it to?. *Proceedings of the National Academy of Science*, 114, 3714–3719.

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

Fisher, R.A. (1921). Studies in Crop Variation. I. An examination of the yield of dressed grain from Broadbalk. *Journal of Agricultural Science*, *11*(2), 107–135.

Fisher, R. A. (1935). *The design of experiments.* Oliver & Boyd.

Fisher, R. A. MacKenzie, W. A. (1923). Studies in Crop Variation. II. The manurial response of Different potato varieties. *Journal of Agricultural Science*, *13*(3), 311–320.
Franco, A., Malhotra, N., Simonovits, G. (2014). Publication Bias in the Social Sciences: Unlocking the File Drawer. *Science*, *345*(6203), 1502–5.
Friedman, D., Sunder, S. (1994). *Experimental Economics: A Primer for Economists*. New York: Cambridge University Press.
Gaddis, S. M. (ed.) (2018). *Audit studies: Behind the scenes with theory, method, and nuance*. Cham, Switzerland: Springer International Publishing.
Gerber, A. S., Green, D. P. (2012). *Field Experiments: Design, Analysis, and Interpretation*. New York: W. W. Norton.
Gilligan, M. J., Pasquale, B. J., Samii, C. (2014). Civil War and Social Cohesion: Lab-in-the-Field Evidence from Nepal. *American Journal of Political Science*, 58, 604–619.
Groves, R. M., Fowler Jr., F. J., Couper, M. P., Lepkowski, J. M., Singer, E., Tourangeau, R. (2009). *Survey Methodology.* Hoboken, New Jersey: John Wiley & Sons.
Guala, F. (2008). History of experimental economics. W: Durlauf, S., Blume, L., *The New Palgrave Dictionary of Economics*. Palgrave-Macmillan.
Hacking, I. (1984). Experimentation and Scientific Realism. W: Leptin, J. (ed.), *Scientific Realism*. Berkeley and Los Angeles: University of California Press.
Hainmueller, J., Hopkins, D. J., Yamamoto, T. (2014). Causal Inference in Conjoint Analysis: Understanding Multidimensional Choices via Stated Preference Experiments. *Political Analysis*, 22, 1–30.
Hedges, L. V., Schauer, J. (2018). The history of randomized trials in education in the USA. *Education Research*, 60, 265–275.
Hegtvedt, K. A. (2014). Ethics and Experiments. W: Webster, M., Sell, J., *Laboratory Experiments in the Social Sciences* (pp. 23–51), "Academic Press".
Henrich, J., Heine, S. J., Norenzayan, A. (2010). The Weirdest People in the World?. *Behavioral and Brain Sciences*, 33 (April), 61–83.
Holland, P. W. (1986). Statistics and causal inference. *Journal of the American Statistical Association*, 81, 945–960.
Kagel, J. H., Roth, A. E. (2016). *The Handbook of Experimental Economics*, Vol. 2. Princeton University Press.
King, G., Sands, M. (2015). *How Human Subjects Research Rules Mislead You and Your University, and What to Do About It*. Working Paper, Harvard University.
Kuklinski, J. H., Sniderman, P. M., Knight, K., Piazza, T., Tetlock, P. E., Lawrence, G. L., Mellers, B. (1997). Racial Prejudice and Attitudes Toward Affirmative Action. *American Journal of Political Science* JSTOR, 402–19.
Lélé, S., Norgaard, R. B. (2005). Practicing Interdisciplinarity. *BioScince*, *55*(11), 967–975.
Leonard, R. (1994). *Laboratory strife: higgling as an experimental science in economics and social psychology.* W: De Marchi, N. B., Morgan, M. S. (eds.), *Higgling* (pp. 343–369). Durham: Duke University Press.
Lipsey, M. W. (1990). *Design Sensitivity: Statistical Power for Experimental Research*. SAGE Publications.
Lipsey, R. G. (1979). *An Introduction to Positive Economics*. Weidenfeld and Nicolson.
Macrae, C. N., Bodenhausen, G. V., Milne, A. B., Jetten, J. (1994*).* Out of Mind but Back in Sight: Stereotypes on the Rebound. *Journal of Personality and Social Psychology*, *67*(5),808.
Mäki, U. (2005). Models are experiments, experiments are models. *Journal of Economic Methodology*, *12*(2), 303–315. https://doi.org/10.1080/13501780500086255.

McDermott, R., Hatemi, P. K. (2020). Ethics in field experimentation: A call to establish new standards to protect the public from unwanted manipulation and real harms. *PNAS*, December 1, *117*(48), 30014–30021.

McFadden, D. (2017). Stated preference methods and their applicability to environmental use and non-use valuations. W: McFadden, D., Train, K. (eds.), *Contingent valuation of environmental goods: a comprehensive critique*. Cheltenham, UK Edward Elgar Publishing Limited.

Morgan, M. S. (2003). Economics. W: Porter, T., Ross, D. (eds.), *The Cambridge History of Science*, Vol. 7, "The Modern Social Sciences" (pp. 275–305). Cambridge: Cambridge University Press.

Mullinix, K. J., Leeper, T. J., Druckman, J. N., Freese, J. (2015). The Generalizability of Survey Experiments. *Journal of Experimental Political Science*, 2, 109–138.

Mutz, D. C. (2011). *Population-Based Survey Experiments*. Princeton University Press.

Neumark, D. (2018). Experimental research on labor market discrimination. *Journal of Economic Literature*, 56, 799–866.

Neumark, D., Burn, I., Button, P. (2019). Is it harder for older workers to find jobs? New and Improved Evidence from a Field Experiment. *Journal of Political Economy*, *127*(2).

Norton, J. D. (2015). Replicability of Experiment. *Theoria*, 30(2), 229–248.

Nosek, B. A., Alter, G., Banks, G. C., Borsboom, D., Bowman, S. D., Breckler, S. J., Buck, S., Chambers, C. D., Chin, G., Christensen, G., Contestabile, M., Dafoe, A., Eich, E., Freese, J., Glennerster, R., Goroff, D., Green, D. P., Hesse, B., Humphreys, M., Ishiyama, J., Karlan, D., Kraut, A., Lupia, A., Mabry, P., Madon, T., Malhorta, N., Mayo-Wilson, E., McNutt, M., Miguel, E., Levy Paluck, E., Simonsohn, U., Soderberg, C., Spellman, B. A., Turitto, J., VandenBos, G., Vazire, S., Wagenmakers, E. J., Wilson, R., Yarkoni, T. (2015). Promoting an Open Research Culture. *Science*, 348, 1422–1425.

Open Science Collaboration. (2015). Estimating the Reproducibility of Psychological Science. *Science*, 349, aac4716.

Pearl, J. (1995). Causal diagrams for empirical research. *Biometrika*, 82, 669–688.

Resnik, D. B. (1994). Hacking's Experimental Realism. *Canadian Journal of Philosophy*, 24(3), 395–411. https://doi.org/10.1080/00455091.1994.10717376.

Roth, A. E. (1993). The Early History of Experimental Economics. *Journal of the History of Economic Thought*, *15*(2), Fall, 184–209.

Roth, A. E. (1995). Introduction to Experimental Economics. W: Kagel, J. H., Roth, A. E. (eds.), *The Handbook of Experimental Economics*. Princeton: Princeton University Press.

Roth, A. E. (2002). The Economist as Engineer: game theory, experimentation, and Partnerships. *Perspectives on Politics*, 18, 10–26.

Schoenberger, E. (2001). Interdisciplinarity and social power. *Progress in Human Geography*, 25, 365–382.

Shadish, W. R., Cook, T. D., Campbell, D. T. (2002). *Experimental and Quasi-Experimental Designs for Generalized Causal Inferences*. Boston: Houghton Mifflin.

Smith, V. L. (1976). Experimental Economics: Induced Value Theory. *American Economic Review*, 66, 274–279.

Smith, V. L. (1981). Experimental economics at Purdue. W: Smith, V. L., *Papers in Experimental Economics* (pp. 154–158). Cambridge: Cambridge University Press.

Smith, V. L. (1994). *Economics in the Laboratory. Journal of Economic Perspective*, (8).

Sovey, A. J., Green, D. P. (2011). Instrumental variables estimation in political science: A readers' guide. *American Journal of Political Science*, 55, 188–200.

Teele, D. (2021). Virtual Consent: The Bronze Standard for Experimental Ethics. W: Druckman, J. N., Green, D. P. (eds.), *Cambridge Handbook of Advances in Experimental Political Science*. New York: Cambridge University Press.
Teele, D. L. (ed.) (2014). *Field experiments and their critics. Essays on the Uses and Abuses of Experimentation in the Social Sciences*. New Haven & London: Yale University Press.
Titiunik, R. (2021). Natural experiments. W: Druckman, J. N., Green, D. P. (eds.), *Advances in Experimental Political Science*. New York: Cambridge University Press.
Von Neumann, J., Morgenstern, O. (1944). *The Theory of Games and Economic Behavior*. Princeton: Princeton University Press.
Westreich, D., Edwards, J. K., Lesko, C. R., Cole, S. R., Stuart, E. A. (2019). Target Validity and the Hierarchy of Study Designs. *American Journal of Epidemiology*, 188, 438–443.

EXPERIMENTS IN MODERN ECONOMICS – EXPANSION AND TECHNOLOGICAL AND INSTITUTIONAL INNOVATIONS IN THE U.S.

Abstract

Experimental economics emerged during the mid-20th century and was created through a combination of the experimental methodology employed in psychology and new advancements in economic theory. Early studies utilizing experimental approaches were conducted on bargaining behavior, social dilemmas, individual decision-making, and market institutions, but experienced a lengthy period of underground growth prior to flourishing in the 21st century. The contemporary state of experimental economics is characterized by a surge of new data sources, the adoption of innovative measurement techniques, the implementation of underutilized experimental designs, advancements in statistical methodologies, increased discussions on robustness and generalizability, and the extensive application of experiments to various fields of study (Druckman, Green, 2021). The main aim of this paper is to outline the evolution of experimental economics, describe contemporary experimental methods, highlight the technological and institutional innovations that support experimentation, particularly in the United States, and identify the primary challenges that exist for the further development of this methodology. It is argued that experimental methods are more commonly employed in the U.S. due to factors such as access to low-cost data collection tools, institutional support, and the emphasis on interdisciplinary research.

Keywords: experiments, methodology, experimental economics

JEL: B41, C90, C91, C92, C93

EKSPERYMENTY WE WSPÓŁCZESNEJ EKONOMII. ROZWÓJ METOD EKSPERYMANTALNYCH WRAZ Z INNOWACJAMI TECHNOLOGICZNYMI I INSTYTUCJONALNYMI W USA

Streszczenie

Współczesna ekonomia eksperymentalna narodziła się w połowie XX wieku wraz z adaptacją metod eksperymentalnych, stosowanych dotąd w psychologii, do analizy problemów ekonomicznych. Początkowo podejście eksperymentalne wykorzystywano na marginesie

głównego nurtu ekonomii w celu zrozumienia zachowań uczestników negocjacji przetargowych, procesów indywidualnego podejmowania decyzji oraz badania instytucji rynkowych. Prawdziwy rozkwit tej metodologii w naukach społecznych nastąpił dopiero w XXI wieku dzięki rozwojowi nowych źródeł pozyskiwania danych i innowacyjnych technik pomiarowych, postępowi w metodach statystycznych oraz intensyfikacji dyskusji na temat możliwości generalizowania wyników tego typu badań (Druckman, Green, 2021). Głównym celem artykułu jest przedstawienie rozwoju podejścia eksperymentalnego w ekonomii wraz z charakterystyką współczesnych typów eksperymentów oraz wskazaniem innowacji technologicznych i instytucjonalnych wspierających wykorzystanie tej metodologii w naukach społecznych, zwłaszcza w USA. W artykule podjęto również próbę identyfikacji głównych wyzwań stojących przed dalszym rozwojem podejścia eksperymentalnego w ekonomii. Artykuł dowodzi, że metody eksperymentalne są częściej stosowane w USA ze względu na dostęp do tanich źródeł pozyskiwania danych, wsparcia instytucjonalnego i promowania interdyscyplinarności na amerykańskich uczelniach.

Słowa kluczowe: eksperymenty, metodologia, ekonomia eksperymentalna

JEL: B41, C90, C91, C92, C93