
17. Experimental sociology*

Klarita Gërkhani and Luis Miller†

1. INTRODUCTION: A BRIEF HISTORY OF EXPERIMENTAL SOCIOLOGY

Over the last 30 years, social sciences have witnessed an increase in the popularity of experimental research. This popularity is evident, for instance, in the number of articles using experimental designs published in the main general-interest journals of various social science disciplines (Gereke & Gërkhani 2020). However, upon a closer look, the tendencies in economics and political science, on the one hand, and sociology, on the other, are remarkably different for at least two reasons. First, in the former disciplines, the adoption of the experimental method has been more recent, but experiments have become increasingly mainstream. On the contrary, sociologists have been conducting experiments for more than a century, but the experimental method has never become central in the discipline. Second, sociological research employs a wide variety of experimental designs, something that is consistent with its theoretical and empirical ‘pluralism’ (for the concepts of theoretical and empirical pluralism in sociology, see the chapter by Raub, De Graaf & Gërkhani on rigorous sociology in this *Handbook*). In contrast, in economics, laboratory and natural field experiments have dominated the scene and, in political science, the survey experiment is clearly the most commonly used experimental design.

For most of the past century, economists and political scientists claimed that the experimental method could not be used in their field. In the beginning of the twentieth century, the president of the American Political Science Association, A.L. Lowell, argued that political science was an observational, not an experimental science (Lowell 1910, p. 7). Similarly, the Nobel Prize winner in Economics, Milton Friedman, stated in the 1950s that economics must rely on ‘experiments that happen to occur’ (Friedman 1953, p. 150). In contrast, a few decades later, one could read that in political science ‘The number and influence of experimental studies are growing rapidly’ (Druckman et al. 2011, p. 3) and in economics, that ‘Quite a substantial body of replicable experimental evidence has been gathered on a growing number of topics’ (Roth 1995, p. 3). This is not the way things evolved in sociology.

In sociology, the most intense debate about the use of the experimental methods occurred in the 1930s and 1940s (see Barrera et al. forthcoming, for a more detailed history of and elaboration on experimental sociology). But, even then, ‘only a minute portion of Sociology [was] experimental’ (Greenwood 1945, p. 5). Moreover, contrary to the other social sciences, one can find several definitions of experimental research in

* K. Gërkhani, N.D. de Graaf and W. Raub (eds) (2022), *Handbook of Sociological Science: Contributions to Rigorous Sociology*, Cheltenham, UK and Northampton, MA, USA: Edward Elgar Publishing. ISBN: 978 1 78990 942 5; <http://dx.doi.org/10.4337/9781789909432>.

† We thank Andreas Diekmann, Nan Dirk de Graaf, Michelle Jackson, Werner Raub, and participants in the online meetings held in December 2020 for helpful comments on an earlier version.

sociology.¹ In this sense, the discussion of the nature of experimental sociology is more complex than the quest for the study of causality typically advocated by the other experimental social sciences.

Although definitions of what constitutes an experiment abound, every definition refers to the ideas of ‘control’, ‘manipulation’, or a combination of both. Control refers to the capacity of the researcher to keep all the extraneous factors, that could confound the investigation, constant. One could argue that control is not enough to characterize the experimental method, since the combination of observational data and careful statistical modeling also provides the researcher with a high degree of control of, at least, the observable intervening factors in the investigation. Typically, we also expect some kind of manipulation to define a design as experimental. Manipulation refers to the purposeful intervention of the researcher in the data generation process. An important case within this narrow definition is manipulation by ‘random assignment’, in which units are assigned to experimental treatments or conditions with known probability. Given that it is hard to know all the extraneous variables that could confound an investigation, experiments use random assignment to guarantee that the comparison units are indeed comparable. Therefore, an experiment is a methodological tool that allows the researcher to draw causal conclusions through control and random assignment.²

As a final remark on the history of experimental sociology, we would like to distinguish three stages in the development of this subdiscipline. In an initial stage, which goes from the early twentieth century to the 1970s, natural field experiments prevailed (Oakley 1998). The focus was on the empirical evaluation of public interventions related to topics such as income, employment support or housing allowance. In this first phase, most sociologists believed that pure (laboratory) experiments were not feasible in sociology and only natural field experiments could be used to study social problems. A second stage, starting in the 1960s and lasting for two decades approximately, experienced the emergence and rapid growth of laboratory experiments that tested the new micro-sociological theories developed in the 1950s and 1960s, particularly ‘social exchange theory’ (Blau 1964; Homans 1961).³ Finally, the third and the most recent stage (this century) in the development of experimental sociology is more eclectic. Current experimental sociology is influenced by other experimental social sciences, particularly by experimental social psychology and experimental economics, and the range of experimental designs is wider than in the previous two phases. The influence of psychology and economics has resulted in the use of a variety of methodological choices and standards in experimental sociology. For instance, some sociological experiments employ monetary incentives – particularly those testing game theoretical models – while others use

¹ Greenwood’s typology includes five: ‘pure experiment’, ‘uncontrolled experiment’, ‘ex post facto experiment’, ‘trial-and-error experiment’, and ‘controlled observational experiment’. In this chapter, we only focus on the first type for reasons discussed below.

² Note that there are important sociological types of research designs – usually also called experimental – that aim to maintain a high degree of control without the researcher’s explicit manipulation or intervention. These are the so-called ‘ex post facto’, ‘quasi-experiments’ or ‘natural experiments’. While these are important designs in sociology, our overview of the field will be based on the narrower definition of an experiment – what many sociologists would call ‘pure experiments’ – which involves both control and manipulation.

³ Another example is the seminal and highly influential study of Marwell & Ames (1979), who were the first to conduct experimental public goods games to study the provision of public goods. Their public good experiment was followed by numerous related experimental applications across social sciences.

other types of incentives or do not incentivize research participants at all. Similarly, the use of experimental designs involving ‘deception’, a common practice in psychology but unacceptable in economics, is not always contested in sociology. Some important well-known experiments deceive participants, for example those studying social conformity such as Willer et al. (2009). In contrast, those in the rational choice tradition do not typically use deception.

This chapter aims to highlight some key features of the experimental method in sociology. It does so by (1) discussing the need for experimentation in sociology, followed by (2) a discussion on an important methodological issue, namely validity, (3) a summary of various types of experimental designs, (4) an illustration of the use of experimentation in relation to some core sociological research, (5) a selection of new and interesting developments in experimental sociology, and (6) a concluding section.

2. WHY WE NEED EXPERIMENTS IN SOCIOLOGY

Most textbooks on social science experimental methods distinguish at least two types of experiments with regard to their main goal: ‘theory-driven’ and ‘empirically-driven’ experiments (Kagel & Roth 2020; Morton & Williams 2010; Willer & Walker 2007). Theory-driven experiments test hypotheses derived from theoretical models. In contrast, empirically-driven experiments search for empirical regularities and investigate phenomena that cannot be explained by existing theories. Whether an experiment is conceptualized as theory- or as empirically-driven has profound consequences for issues of causality and generalizability.

In this section, we focus on causality; generalizability is discussed below. In sociology, an ideal theory-driven experiment would start by proposing an action-based sociological explanation that can be operationalized and tested by means of a laboratory or field experiment. Take, for instance, the puzzle of why people enforce social norms even when they privately consider that these norms are false or even harmful. Using agent-based computational models, a possible solution of this puzzle was theoretically modelled by Centola et al. (2005) and then tested in a series of laboratory experiments by Willer et al. (2009). Hence, after establishing a potentially falsifiable hypothesis derived from a sociological theory, mostly about a causal effect, theory-driven experiments become a useful tool to test the direction and the strength of the purported effect. When a randomized experiment is not feasible, either because of the nature of the research question or because of logistical reasons, sociologists rely on other designs to test causal effects, including natural experiments and other types of statistical modeling with observational data (for the study of causal effects using observational data, see Breen’s chapter on causal inference).

Empirically-driven experiments, on the other hand, are useful for establishing empirical regularities (Jackson & Cox 2013; Willer & Walker 2007), which in turn can contribute to theory development. These types of experiments rely on the ‘Rubin causal model’ of causal inference (Rubin 1974). This model assumes that only the factors that can be purposely manipulated in experiments (i.e., the ‘treatments’) can be causal. The paradigmatic example of empirically-driven experiments are the so-called ‘randomized controlled trials’ (RCTs), especially those purposely designed to test a policy intervention. Take, for

instance, experimental studies of the labor market effects of ‘negative income tax’, ‘income maintenance’ or ‘housing allowance’ policies. In all these experiments, individuals or families are randomly allocated to treatment and control conditions, that is, some participate in the program and some do not. Although the interventions might be theoretically informed, the main goal of these RCTs is to assess the effects of a specific intervention. Thus, the generalizability of the results to other instances or the confirmation of a theoretical hypothesis are of secondary interest, at most.⁴

From the perspective of rigorous sociology, a thorough understanding of a causal process via experiments is only possible through the lens of an appropriate theory of human action and with a further focus on unravelling mechanism-based explanations (see Jackson’s chapter on sociology as a population science where she encourages population science sociologists to make more use of experiments for ‘identifying the micro-level mechanisms proposed as explanations of the macro-level empirical regularities’, p. 22). This is in line with the idea of ‘causation as a generative process’, the type of causality advocated by analytical or mechanism-based sociology (Blossfeld & Prein 1998; Goldthorpe 2001; Hedström 2005; Hedström & Swedberg 1998). For discussions on the importance of linking theory construction with empirical research and for searching for mechanism-based explanations, see the chapter by Raub, De Graaf & Gërxhani, as well as Manzo’s chapter on analytical sociology.

To conclude on why we need experiments in sociology, regardless of whether the goal of experimentation is testing a theory, discovering a new regularity, or studying the effectiveness of a new policy intervention, experimental designs have a number of advantages compared with observational studies. Beyond a high degree of control and hence a more precise estimate of a causal relationship, experiments are easier to replicate. The latter contributes not only to the robustness of empirical findings and, thus, to cumulative knowledge, but also to more integrity and transparency (see also the chapter by Auspurg and Brüderl on reproducibility and credibility). The main disadvantage of experiments is artificiality, related to either the setting in which they are conducted or the pool of participants whose behavior is under study. Artificiality, in turn, has implications for the external validity of the experimental findings. This is what we turn to next.

3. VALIDITY OF EXPERIMENTS IN SOCIOLOGY

When it comes to experiments, the main methodological challenge sociologists tend to think about is validity, particularly external validity or generalizability. Validity in empirical research refers to the degree in which the empirical results are robust and replicable. In sociology, most researchers use the distinction between internal and external validity introduced by Campbell and his collaborators more than 60 years ago (Campbell 1957; Campbell & Stanley 1963; Cook & Campbell 1979). In their original contributions, internal validity relates to *causality* – inferences ‘in’ or ‘within’ the experiment – and external validity to *generalizability* – inferences ‘beyond’ or ‘outside’ the experiment. However, this conceptualization of validity is far from being consensual among social

⁴ Deaton & Cartwright (2018) offer an extensive discussion of the pros and cons of RCTs in the social sciences.

science methodologists and practitioners and several debates have emerged since its introduction. For an excellent discussion, see Falk & Heckman (2009).

A prominent debate in social sciences concerns the logical relationship between internal and external validity. Central to this debate is the idea of a tension between the two, that is, there is a trade-off between control (internal validity) and generalizability (external validity) of experimental results (Jimenez-Buedo & Miller 2010). To maximize internal validity one ought to *shield* the experiment against external disturbances that could interfere with the experimental treatment. However, this shielding against extraneous factors may limit the transportability of the experimental results to other contexts and populations. Conversely, a realistic (i.e., non-artificial) experimental environment may enhance the external validity of an experimental design, but limit the capacity of the researcher to control all the potential intervening factors that can interact with the treatment. Although intuitive, the idea of the trade-off between internal and external validity has been widely contested. For many, internal and external validity cannot be in a trade-off relationship because internal validity is a pre-requisite of external validity. In other words, there is no point in worrying about the generalization of some experimental results until one is confident about the internal validity of that result.

The methodological and practical consequences of the assumed trade-off between internal and external validity has structured the debate about the validity of social science experiments for decades. The most recent instantiation of that debate frames the issue of validity in terms of the theoretical and the empirical approaches to validity. This discussion mimics the above-mentioned distinction between theory-driven and empirically-driven experiments. The empirical approach to validity proposes that there are empirical methods (e.g., randomized control trials) that are as good as other methods (e.g., laboratory experiments) in attaining internal validity, but superior in achieving external validity (Al-Ubaydli & List 2015). This approach refers to the concepts of ‘realism’, ‘similarity’, and ‘representativeness’. Intuitively, it requires research constructs to be as similar to or as representative as possible of reality. Not surprisingly, it favors methods that rely on observations or experiments performed on relevant samples and in natural conditions. Depending on the social science discipline, natural field experiments and representative survey experiments would be the ‘gold standards’ for valid experimental research. A number of social scientists have recently used empirical studies to back their claims about the superior validity of natural field experiments and to question the generalizability of laboratory results. These studies typically focus on the original preoccupation of Cook & Campbell (1979) and try to show that laboratory results cannot be generalized to other persons, settings, treatment variables and measurement variables. Their main claim is that the absence of correspondence between results in the laboratory and in the field is evidence of a lack of external validity of results obtained under laboratory conditions (Bader et al. 2019).

In contrast, the theoretical approach to validity argues that there is no gold methodological standard and that external validity can only be achieved through theory. It is a theoretical claim, and not a single empirical finding, that can be generalized. Deaton & Cartwright (2018) have recently argued that establishing causality via empirically-driven experiments does not guarantee that the causal relation will hold in general in social science experiments. Bardsley et al. (2010, p. 51) share this opinion, writing that, ‘strictly, all that happens in a particular laboratory experiment is what happens in it’. Thus, using

the conventional terminology in the Campbellian approach, experiments would lack external validity. However, the main motivation for virtually any experiment is to say something of relevance beyond the local conditions of a particular experimental trial. A common response to this tension is the argument that when experiments are theory-driven, what one generalizes is not a particular experimental result, but a 'theoretical construct' (Deaton & Cartwright 2018). Under this interpretation, 'the aim of such experiments is not to generalize to a wider population, but rather to hone and test a scientific theory that, if validated, could later be applied to a real-world context to explain human behavior' (Jackson & Cox 2013, p. 38). Therefore, an experiment tests a theoretical claim and, if the evidence is confirmatory, the theoretical approach to validity would argue that the external validity of the experimentally-supported theoretical claim has increased.

Finally, an alternative solution to the tension between internal and external validity of experimental research is to combine complementary methods with samples from various populations and across regions or countries, to cross-validate both theory- and empirically-driven experimental findings. This is elaborated further below.

4. TYPES OF EXPERIMENTAL DESIGNS IN SOCIOLOGY

Similar to related disciplines (experimental psychology, economics and political science), experimental sociology has developed different modes to address theoretically- and empirically-driven goals. Depending on the purpose and the research question asked, a researcher can conduct different types of experiments. In principle, those conducted in the laboratory provide the highest degree of control and manipulation. This makes them particularly valuable in serving the purpose of testing and appraising sociological theories, which is central to rigorous sociology as discussed in the chapter by Raub, De Graaf & Gërkhani. Given that theory is an abstraction of a complex reality, the ability of laboratory experiments to isolate the many complexities of the 'real world' makes them the best tool to test theories. For instance, if a researcher is interested in testing the role of social networks in institutional change (Lin 2002), he or she needs to consider that in reality it would be extremely hard, if not impossible, to isolate the many additional factors that could confound the causal effect of social networks on institutional change. Such factors include social preferences, social norms, power and cognitive processes such as emotions. Thus, when the purpose of an empirical study is to test theories, the possibility that laboratory experiments offer to isolate the causal processes deemed important by the theory from those that are deemed irrelevant, is crucial.

Reality is, however, complex and the richness of context can be important for understanding human behavior. An experimental design that is too abstract from the context it represents may lack relevance even if it tests a theory. While in psychology and economics context is often less relevant, the interaction between human action and different social and institutional contexts such as culture, religion, family, or workplace is central in sociology (i.e., micro-macro links; see the chapter by Raub, De Graaf & Gërkhani). Context can be included in laboratory experimentation, for instance, by running the same design across different countries (Gërkhani 2020; Herrmann et al. 2008), different

regions (Barr & Miller 2020; Bigoni et al. 2016), or across different pools of participants (Belot et al. 2015; Gërxhani & Schram 2006). This is one form of what Fisher (1935), Campbell & Stanley (1963), and Jackson & Cox (2013) refer to as ‘independent replication’, which serves the purpose of theory corroboration beyond the scope conditions of the original experiment. When, instead, the research purpose is to explore empirical regularities or to advise on public policy, field experiments and survey experiments have been argued to be more suitable tools because they study human behavior in a more ‘natural’ environment (Falk & Heckman 2009; see also Baldassarri & Abascal 2017 for a recent and excellent overview of studies based on field experiments). Depending on the research question and feasibility constraints, field experiments can take different forms, from (1) those that try to maintain control by studying the behavior of ‘real world’ people (i.e., non-student population) in a laboratory setting; to (2) lab-in-the-field experiments that try to maintain the control by bringing the experiment into the field; and (3) natural field experiments that are also conducted in the field, but without those involved being aware that they are participating in an experiment (Przepiorka & Diekmann 2013).

All these forms of field experiments can be implemented either directly in the field or through online platforms, such as Qualtrics, Mturk, Prolific, which have become increasingly popular in the last decade or so. The degree of experimental control decreases when moving from (1) to (3), because in the field the probability of participants knowing each other and communicating, sharing experiences, or using resources (like the internet) outside of the experimenter’s control increases. In turn, this introduces unwanted noise and confounding factors, which diminishes the internal validity of the research conducted. On the other hand, the more natural the setting in which the experiment is run, the higher the external validity of the research, especially when that particular setting is representative for a broader group of environments (see the showcase-chapter by Salganik, Dodds & Watts for an application).⁵

Finally, when representativeness is essential to experimental research, survey experiments, a.k.a. ‘population-based survey experiments’ (Mutz 2011) or ‘survey-based experimentation’ (Sniderman & Grob 1996), offer an attractive alternative methodological tool. By combining experimental techniques with observational studies, that is, by running a randomized experiment with a representative population sample of interest, survey experiments can provide answers to causal questions that can be generalized to the broader population of interest to the study.

... a population-based experiment uses survey sampling methods to produce a collection of experimental subjects that is representative of the target population of interest for a particular theory, whether that population is a country, a state, an ethnic group, or some other subgroup. (Mutz 2011, p. 2)

⁵ However, as argued by Falk & Heckman (2009), one should be wary of claims that field experiments *generically* have higher external validity than laboratory experiments. It is not obvious, for example, that a field experiment conducted in rural India has more relevance for human behavior in New York City than a laboratory experiment run at NYU.

5. EXAMPLES OF CORE SOCIOLOGICAL RESEARCH IN EXPERIMENTAL SOCIOLOGY

In this section, we review several examples of sociological experiments and how they connect with core sociological research, such as social exchange theory, social network analysis and social stratification. Although these topics represent important research agendas in experimental sociology, they by no means exhaust the whole field. Other important experimental research topics include procedural and distributive justice, cooperation and pro-social behavior, social norms and organizational behavior.

5.1 Social Exchange Theory and Social Network Analysis

Humans are not atomized units that make decisions in isolation. Instead, they interact in a plethora of binary and multi-person social relationships. This embeddedness in social networks plays a central role in determining how people act (Granovetter 1985). Social exchange theory and social network analysis have been developed to better understand such human interaction in social contexts. The two are closely interrelated as social network analysis focuses on relational and structural characteristics of networks, whereas social exchange theory studies conditions and outcomes of human interactions embedded in a network structure. Under the umbrella of these two grand theories, a number of research streams has evolved that focus on a large variety of aspects of interpersonal relations, including status and power, trust, fairness, and wealth distribution. When individuals interact in a network, many of these aspects may play a role. In fact, one of the largest literatures in experimental sociology is on the effects of network embeddedness on trust in social interactions. In turn, factors such as trust between individuals or their fairness preferences may have an important influence on when networks are formed and what shape they take. Numerous game-theoretic models are applied to theoretically understand these dynamics and the resulting hypotheses have been tested experimentally (see Diekmann's chapter on rational choice sociology). Think for instance of the Trust Game, the Investment Game and variations of each (see also the chapter by Buskens, Corten & Raub on social networks). The hypotheses derived from these games have been frequently tested via laboratory experiments (Cook & Cooper 2003), but also through alternative designs such as surveys and vignette experiments (see Buskens & Raub 2013 for an extensive overview of research in experimental game theory).

Experiments, particularly laboratory experiments, in sociology have proven to be a valuable tool for testing and further developing theories on factors like those mentioned above and therefore for the further development of social exchange theory and social network analysis (Neuhofer et al. 2016). Most experiments on networks and social exchange build on the experimental setup introduced by Cook et al. (1983), which studies the relationship between (exogenous) network structure and the division of power in the network. Each participant is connected to (some) other participants via so-called links. One can only engage in exchange with someone to whom one is linked. Together, the set of links determines the network structure, which may vary from complete (everyone is linked to everyone) to empty (nobody is linked to anyone), with everything in between. In experiments on network formation, individuals themselves

determine with whom they want to link. In early experimental work, the structure in which exchange takes place is predetermined by providing an exogenous set of links. In the last decade or so, there has been an increasing interest in understanding the endogenous emergence and structure formation of social networks (e.g., Gërkhani et al. 2013; Raub et al. 2013; and also the chapter by Buskens, Corten & Raub).⁶

Social exchange theory and its numerous experimental applications relate to social network analysis because of the focus on social exchange within networks. The contribution of social exchange theory does, however, extend to a better understanding of the human nature of such interactions (i.e. the micro-level). Rational motives of exchange are considered in combination with ‘non-rational’ ones, such as emotions, cohesion and social preferences (e.g., Kuwabara 2011; Lawler et al. 2006; Molm 2007). The experimental research shows that both types of motives drive individual behavior into different forms of exchange relations. In turn, the structure of these relations affects which of these motives is activated during the exchange. However, a challenge for experimental studies on social exchange theory is to consider more realistic and thus more complex dimensions of exchange networks. Neuhofer et al. (2016) suggest the use of social media and other online networks to capture this complexity. Moreover, doing so provides an opportunity for developing further the ‘limited’ existing theories.

5.2 Social Stratification

Social stratification is another core research topic in sociology. It is about inequalities in a society based on innate or ascribed characteristics such as race, gender, wealth, education, status and power. Beyond the interest in inequalities across individuals, sociology is primarily interested in understanding systemic inequalities (i.e. the macro-level) according to one’s group membership along these characteristics. As this understanding requires more micro-macro explanations, social stratification scholars have increasingly embraced the application of experimental methods. ‘Yet social science experiments are often criticized for their lack of external validity, especially their limited generalizability to real-life settings’, which has prevented the widespread use of experiments in social stratification research (Chen & Tam 2020, p. 1; see also Breen’s chapter).

That is why the use of laboratory experiments has been rather limited (e.g., Berger & Diekmann 2015; Correll et al. 2007; Côté et al. 2015; Gërkhani et al. 2013; Nishi et al. 2015).⁷ Because different forms of field-oriented types of experiments are believed to offer a higher level of external validity, social stratification scholars have more frequently used them. They vary from (1) audit and correspondence designs, aiming to study discrimination in hiring (based on race, gender, ethnicity, class, etc.; for some excellent overviews, see Pager 2007; Quillian et al. 2019; Zschirnt & Ruedin 2016); to (2) field experiments on topics such as poverty, education and social influence (e.g., Baldassarri & Abascal 2017); and (3) factorial vignette experiments on discrimination, prejudice and judgment related

⁶ There is an even earlier and highly influential study on endogenous formation of trading relationships by Kollock (1994), which was followed up only years later.

⁷ See Chen & Tam (2020) for an overview of experimental social stratification research related to the evolution of social inequality, social inequality and mobility, and the social status effect. For an overview of laboratory experimental studies on the demand side of the labor market and the role of hiring in generating labor market (in)equalities, see Gërkhani (2017).

to hiring, selection and evaluation more generally (e.g. Petzold & Wolbring 2019; Wallander 2009). Moreover, there is a recent welcoming development of conducting comparative experimental research across countries, which is central in sociology and especially in social stratification research (Gërkhani 2020; Lancee et al. 2019).⁸

6. FUTURE DEVELOPMENTS IN EXPERIMENTAL SOCIOLOGY

Experimentation in sociology has been marginal compared with the other experimental disciplines in the social sciences, but has recently shown an increase in popularity. This expansion includes new and interesting developments, a selection of which we will highlight here.

6.1 Incorporation of Experimental Techniques in Surveys

As mentioned above, applying experimental techniques to representative population samples offers the best of both worlds. It increases a causal understanding and ‘mechanism-based’ explanations of a complex reality with a relevance beyond the group of participants under experimental investigation. Moreover, incorporation of experiments in large surveys allows for controlled variation in a larger number of explanatory variables. One of the central focuses of rigorous sociology is on micro-macro explanations, which can greatly benefit from a broader expertise on how to incorporate experimental methods into existing population surveys. Developing such expertise therefore provides great potential in making sociological research more rigorous. There are very few recent successful examples: Ermisch & Gambetta (2016) and Gereke et al. (2018) embed experimental trust games in population surveys in the UK and Germany, respectively.

6.2 Cross-validation by Combining Complementary Methods

Another way of increasing the validity of experimental results is by combining distinct experimental as well as observational methods, a.k.a. ‘cross-validation’ or ‘empirical cross-checking’ (see Barr et al. 2021). As outlined in the chapter by Raub, De Graaf & Gërkhani and also emphasized in Jackson & Cox (2013) and Buskens & Raub (2013), employing different research designs allows for complementary tests of the same hypotheses or for complementary ways of establishing empirical regularities. Doing so contributes to the robustness of empirical findings and ultimately to cumulative knowledge. Indeed, experimental cross-validation is becoming more frequent. Some studies have combined laboratory with field experimental designs (Bardsley et al. 2010; Guala 2005) or different types of field experimental designs (Sarsons et al. 2021); others have combined laboratory- with vignette- and observational-data (Chen & Tam 2020) or variations, such as laboratory with vignette (Barr et al. 2021), vignette with observational data (Eifler 2010; Petzold & Wolbring 2019), or laboratory with audit designs (Correll et al. 2007).

⁸ For a recent collection of social stratification studies applying these types of experimental designs, see a special issue on ‘Experimental methods in social stratification research’ edited by Barone & Solga (2020).

The higher the compatibility between the research designs combined in a study, the better is considered the balance between the internal and the external validity of the empirical findings of that study. Cross-validation is thus a development that deserves to be strengthened. A further possible step is to also expand the use of cross-country comparisons in experimental research.

6.3 Incorporation of New Cross-disciplinary Techniques to Increase Depth

The strength of rigorous sociology lies in recognizing that a better understanding of social phenomena requires a consideration of how human actions both shape and are shaped by social and institutional contexts (i.e. the micro-macro link). In studying this mutual interaction, the biological basis of behavior has, until recently, been disregarded. A number of experimental studies demonstrate the added value of considering humans' physiological and brain reactions to the social context in which they are embedded (see Franks & Turner 2013 and also the chapter by Hopcroft, Dippong, Liu & Kail on evolution, biology, and society). For example, Willer et al. (2013) and De Dreu et al. (2021) find that a variation in people's social environment leads to differential physiological responses and that these responses affect people's attitudinal and behavioral outcomes. Therefore, biological reactions may be activated by situational factors and affect behavior or attitudes (Willer et al. 2013). Sociological experimentation has yet to more frequently incorporate measures and techniques, such as neuro-imaging and psycho-physiological measures, that are becoming common in other social science disciplines.

6.4 Exchange with other Experimental Disciplines

Experimental sociologists interested in causal identification and mechanism-based explanations of micro-macro links have benefited from the advancement of experimental methods in behavioral economics and social psychology. Experimental economists have become increasingly interested in core sociological topics, such as trust, cooperation, and social norms. This has led to a number of insightful interdisciplinary collaborations (e.g., Barr et al. 2016; Bigoni et al. 2013, 2016; Efferson et al. 2016; Gërkhani et al. 2013; Schram et al. 2019). One of the most important insights is the recognition of the mutual interaction between the micro- and the macro-level of analysis. As extensively discussed in Gereke & Gërkhani (2020), breakthrough research can result from a closer cooperation between experimental sociologists and experimental economists: '... we believe there is much potential for generating new insights into the foundations of human decision-making by taking seriously the effects of actors' socio-environmental context on their preferences, expectations and behavior' (Gereke & Gërkhani 2020, p. 14). Similarly, collaborations with experimental political scientists can contribute to a better understanding of beliefs, attitudes and opinions, which, in turn, can be used to better understand the resulting behavioral outcomes (e.g., Winter & Zhang 2018). An innovative approach in this direction, especially when the research purpose is to advise on social policy, is the implementation of a randomized field experiment within a deliberative poll (e.g., Farrar et al. 2010). The latter has become increasingly popular in political science, because it seems to improve public opinion and decision making when those involved in a deliberative poll are given an opportunity to think and deliberate further on the (policy) issues

discussed. Thus, in line with Jackson & Cox (2013), we believe that progress lies on the synthesis of interdisciplinary research.

7. CONCLUSION

Delimiting the scope of experimental sociology is not an easy task. As with other sociological research, the range of topics covered is wider than in disciplines such as economics or political science. Moreover, there is a pluralism of theories, methodologies, and practical standards that could not possibly be summarized in just one chapter. However, we can still extract a few take-away messages from the history and current developments in experimental sociology.

First, although sociologists have been conducting experiments for more than a century, the experimental method has not become as central as in other social science disciplines. This could be related to the much debated tension between internal and external validity of experimental research. Owing to the high degree of control and randomization, there is general agreement that experiments score highly on internal validity. Regarding external validity, the most contested issue of experimental research, it is important to reflect on whether the research aims to test a theory or to establish empirical regularities. When the purpose of research is to test potentially falsifiable hypotheses derived from a sociological theory, external validity is ensured through the empirical corroboration of the theory. When the purpose of research is to search for empirical regularities and investigate phenomena that cannot be explained by existing theories, a researcher should consider experimental designs that are conducted in a relevant social context. Overall, a good balance between internal and external validity of experimental research is achieved through replication, by combining complementary methods with samples from various populations and across regions or countries.

Second, experimental sociology has a strong theoretical focus. That is, experimental endeavors to establish causal relationships and unravel mechanism-based explanations are driven by micro-macro sociological theories. This is especially the case of experiments that have contributed to testing and developing social exchange theory, social network analyses or rational choice theory. Sociological experiments in the social psychology tradition too are designed to test socio-psychological theories of identity, altruism or cooperation.

Finally, sociological experimentation is closely linked to rigorous theoretical and empirical sociology. Historically, social exchange theory and rational choice theory were important in the development of experimental sociology. More generally, when designing and conducting experiments, sociologists tend to hypothesize and test generative processes at the level of social action. In this respect, we believe experimentation is a privileged method of enquiry for rigorous sociology.

REFERENCES

- Al-Ubaydli, O. and J.A. List (2015), 'Do natural field experiments afford researchers more or less control than laboratory experiments?', *American Economic Review*, **105**, 462–466.
- Bader, F., B. Baumeister, R. Berger, and M. Keuschnigg (2019), 'On the transportability of laboratory results', *Sociological Methods and Research*, **50**, 1452–1481.

- Baldassarri, D. and M. Abascal (2017), 'Field experiments across the social sciences', *Annual Review of Sociology*, **43**, 41–73.
- Bardsley, N., R. Cubitt, G. Loomes, P. Moffat, C. Starmer, and R. Sugden (2010), *Experimental Economics: Rethinking the Rules*, Princeton, NJ: Princeton University Press.
- Barone, C. and H. Solga (eds.) (2020), 'Experimental methods in social stratification research', *Research in Social Stratification and Mobility*, **65**.
- Barr, A. and L. Miller (2020), 'The effect of education, income inequality and merit on inequality acceptance', *Journal of Economic Psychology*, **80**.
- Barr, A., L. Miller, and P. Ubeda (2016), 'Moral consequences of becoming unemployed', *Proceedings of the National Academy of Sciences*, **113**, 4676–4681.
- Barr, A., L. Miller, and P. Ubeda (2021), 'Is the acknowledgement of earned entitlement effect robust across experimental modes and populations?', *Sociological Methods and Research*.
- Barrera, D., K. Gërzhani, B. Kittel, L. Miller, and T. Wolbring. *Experimental Sociology: Outline of a Scientific Field*. Forthcoming, Cambridge University Press.
- Belot, M., R. Duch, and L. Miller (2015), 'A comprehensive comparison of students and non-students in classic experimental games', *Journal of Economic Behavior and Organization*, **113**, 26–33.
- Berger, J. and A. Diekmann (2015), 'The logic of relative frustration, Boudon's competition model and experimental evidence', *European Sociological Review*, **31**, 725–737.
- Bigoni, M., S. Bortolotti, M. Casari, and D. Gambetta (2013), 'It takes two to cheat: An experiment on derived trust', *European Economic Review*, **64**, 129–146.
- Bigoni, M., S. Bortolotti, M. Casari, D. Gambetta, and F. Pancotto (2016), 'Amoral familism, social capital, or trust? The behavioural foundations of the Italian north–south divide', *The Economic Journal*, **126**, 1318–1341.
- Blau, P. (1964), *Exchange and Power in Social Life*, New York, NY: John Wiley & Sons.
- Blossfeld, H.-P. and G. Prein (eds.) (1998), *Rational Choice Theory and Large-Scale Data Analysis*, Boulder, CO: Westview Press.
- Buskens, V.W. and W. Raub (2013), 'Rational choice social research on social dilemmas: Embeddedness effects on trust', in R. Wittek, T.A.B. Snijders, and V. Nee (eds.), *Handbook of Rational Choice Social Research*, Stanford, CA: Stanford University Press, pp. 113–150.
- Campbell, D.T. (1957), 'Factors relevant to the validity of experiments in social settings', *Psychological Bulletin*, **54**, 297–312.
- Campbell, D.T. and J.C. Stanley (1963), *Experimental and Quasi-Experimental Designs for Research*, Chicago, IL: Rand McNally & Company.
- Centola, D., R. Willer, and M.W. Macy (2005), 'The emperor's dilemma, a computational model of self-enforcing norms', *American Journal of Sociology*, **110**, 1009–1040.
- Chen, J.C. and T. Tam (2020), 'Uses of artificial and composite treatments in experimental methods: Reconsidering the problem of validity and its implications for stratification research', *Research in Social Stratification and Mobility*, **65**.
- Cook, K.S. and R.M. Cooper (2003), 'Experimental studies of cooperation, trust, and social exchange', in E. Ostrom and J. Walker (eds.), *Trust and Reciprocity: Interdisciplinary Lessons from Experimental Research*, New York, NY: Russell Sage Foundation, pp. 209–244.
- Cook, K.S., R.M. Emerson, M.R. Gillmore, and T. Yamagishi (1983), 'The distribution of power in exchange networks. Theory and experimental results', *American Journal of Sociology*, **89**, 275–305.
- Cook, T.D. and D.T. Campbell (1979), *Quasi-Experimentation, Design & Analysis Issues for Field Settings*, Boston, MA: Houghton Mifflin Company.
- Correll, J.S., S. Benard, and I. Paik (2007), 'Getting a job: Is there a motherhood penalty?', *American Journal of Sociology*, **112**, 1297–1339.
- Côté, S., J. House, and R. Willer (2015), 'High economic inequality leads higher-income individuals to be less generous', *Proceedings of the National Academy of Sciences*, **112**, 15838–15843.
- Deaton, A. and N. Cartwright (2018), 'Understanding and misunderstanding randomized controlled trials', *Social Science & Medicine*, **210**, 2–21.
- De Dreu, C., K. Gërzhani, and A. Schram (2021), 'Anticipating peer ranking causes hormonal adaptations that benefit cognitive performance', *American Behavioral Scientist*.
- Druckman, J.N., D.P. Green, J.H. Kuklinski, and A. Lupia (2011), *Cambridge Handbook of Political Science*, Cambridge, UK: Cambridge University Press.
- Efferson, C., C.P. Roca, S. Vogt, and D. Helbing (2016), 'Sustained cooperation by running away from bad behavior', *Evolution and Human Behavior*, **37**, 1–9.
- Eifler, S. (2010), 'Validity of a factorial survey approach to the analysis of criminal behavior', *Methodology: European Journal of Research Methods for the Behavioral and Social Sciences*, **6**, 139–146.
- Ermisch, J. and D. Gambetta (2016), 'Income and trustworthiness', *Sociological Science*, **3**, 710–729.
- Falk, A. and J.J. Heckman (2009), 'Lab experiments are a major source of knowledge in the social sciences', *Science*, **326**, 535–538.

- Farrar, C., J.S. Fishkin, D.P. Green, C. List, R.C. Luskin, and E.L. Paluck (2010), 'Disaggregating deliberation's effects: An experiment within a deliberative poll', *British Journal of Political Science*, **40**, 333–347.
- Fisher, R.A. (1935), 'The logic of inductive inference', *Journal of the Royal Statistical Society*, **98**, 39–82.
- Franks, D.D. and J.H. Turner (eds.) (2013), *Handbook of Neurosociology*, Dordrecht: Springer.
- Friedman, M. (1953), *Essays in Positive Economics*, Chicago, IL: Chicago University Press.
- Gereke, J. and K. Gërkhani (2020), 'Experimental economics and experimental sociology', in J.H. Hamilton (ed.), *Oxford Research Encyclopedia of Economics and Finance*, Oxford, UK: Oxford University Press.
- Gereke, J., M. Schaub, and D. Baldassarri (2018), 'Ethnic diversity, poverty and social trust in Germany: Evidence from a behavioral measure of trust', *PLoS ONE*, **13**.
- Gërkhani, K. (2017), 'The experimental approach to studying employers' hiring behavior', *Emerging Trends in the Social and Behavioral Sciences: An Interdisciplinary, Searchable, and Linkable Resource*, 1–14.
- Gërkhani, K. (2020), 'Status ranking and gender inequality: A cross-country experimental comparison', *Research in Social Stratification and Mobility*, **65**.
- Gërkhani, K., J. Brandts, and A. Schram (2013), 'The emergence of employer information networks in an experimental labor market', *Social Networks*, **35**, 541–560.
- Gërkhani, K. and A. Schram (2006), 'Tax evasion and income source: A comparative experimental study', *Journal of Economic Psychology*, **27**, 402–422.
- Goldthorpe, J. H. (2001), 'Causation, statistics, and sociology', *European Sociological Review*, **17**, 1–20.
- Granovetter, M.S. (1985), 'Economic action and social structure: The problem of embeddedness', *American Journal of Sociology*, **91**, 481–510.
- Greenwood, E. (1945), *Experimental Sociology: A Study in Method*, New York, NY: King's Crown Press.
- Guala, F. (2005), *The Methodology of Experimental Economics*, Cambridge, UK: Cambridge University Press.
- Hedström, P. (2005), *Dissecting the Social. On the Principles of Analytical Sociology*, Cambridge, UK: Cambridge University Press.
- Hedström, P. and R. Swedberg (1998), 'Social mechanisms: An introductory essay', in P. Hedström and R. Swedberg (eds.), *Social Mechanisms. An Analytical Approach to Social Theory*, Cambridge, UK: Cambridge University Press, pp. 1–31.
- Herrmann, B., C. Thöni and S. Gächter (2008), 'Antisocial punishment across societies', *Science*, **319**, 1362–1367.
- Homans, G.K. (1961), *Social Behavior. Its Elementary Forms*, New York, NY: Harcourt, Brace and World, Inc.
- Jackson, M. and D.R. Cox (2013), 'The principles of experimental design and their application in sociology', *Annual Review of Sociology*, **39**, 27–49.
- Jimenez-Buedo, M. and L. Miller (2010), 'Why a trade-off? The relationship between the internal and external validity of experiments', *Theoria*, **18**, 271–282.
- Kagel, J. and A. Roth (2020), *Handbook of Experimental Economics, Volume 2*, Princeton, NJ: Princeton University Press.
- Kollock, P. (1994), 'The emergence of exchange structures: An experimental study of uncertainty, commitment, and trust', *American Journal of Sociology*, **100**, 313–345.
- Kuwabara, K. (2011), 'Cohesion, cooperation, and the value of doing things together: How economic exchange creates relational bonds', *American Sociological Review*, **76**, 560–580.
- Lancee, B., G. Birkelund, M. Coenders, V. Di Stasio, M. Fernandez-Reino, A. Heath, R. Koopmans, E. Larsen, J.G. Polavieja, M. Ramos, L. Thijssen, S. Veit, R. Yemane, and D. Zwier (2019), 'The GEMM study: A cross-national harmonized field experiment on labour market discrimination: Technical report', retrieved from SSRN: <https://ssrn.com/abstract=3398191> or <http://dx.doi.org/10.2139/ssrn.3398191>
- Lawler, E. J., S. R. Thye, and J. Yoon (2006), 'Commitment in structurally enabled and induced exchange relations', *Social Psychology Quarterly*, **69**, 183–200.
- Lin, N. (2002), *Social Capital, A Theory of Social Structure and Action*, Cambridge, UK: Cambridge University Press.
- Lowell, A.L. (1910), 'The physiology of politics: Presidential address, sixth annual meeting of the American political science association', *American Political Science Review*, **4**, 1–15.
- Marwell, G. and R.E. Ames (1979), 'Experiments on the provision of public goods. I. Resources, interest, group size, and the free-rider problem', *American Journal of Sociology*, **84**, 1335–1360.
- Molm, L.D. (2007), 'Experiments on exchange relations and exchange networks in sociology', in M. Webster and J. Sell (eds.), *Laboratory Experiments in the Social Sciences*, London, UK: Elsevier, pp. 379–406.
- Morton, R. and K. Williams (2010), *Experimental Political Science and the Study of Causality*, Cambridge, UK: Cambridge University Press.
- Mutz, D.C. (2011), *Population-based Survey Experiments*, Princeton, NJ: Princeton University Press.

- Neuhöfer, S., B. Kittel, and I. Reindl (2016), 'Social exchange networks: A review of experimental studies', *Connections*, **36**, 34–51.
- Nishi, A., H. Shirado, D.G. Rand, and N.A. Christakis (2015), 'Inequality and visibility of wealth in experimental social networks', *Nature*, **526**, 426–429.
- Oakley, A. (1998), 'Experiments and social interventions: A forgotten but important history', *BMJ*, **317**, 1239–1242.
- Pager, D. (2007), 'The use of field experiments for studies of employment discrimination: Contributions, critiques, and directions for the future', *Annals of the American Academy of Political and Social Science*, **609**, 104–133.
- Petzold, K. and T. Wolbring (2019), 'What can we learn from factorial surveys about human behavior? A validation study comparing field and survey experiments on discrimination', *Methodology*, **15**, 19–30.
- Przepiorka, W. and A. Diekmann (2013), 'Temporal embeddedness and signals of trustworthiness', *European Sociological Review*, **29**, 1010–1023.
- Quillian, L., A. Heath, D. Pager, A. Midtbøen, F. Fleischman, and O. Hexel (2019), 'Do some countries discriminate more than others? Evidence from 97 field experiments of racial discrimination in hiring', *Sociological Science*, **6**, 467–496.
- Raub, W., V. Buskens, and V. Frey (2013), 'The rationality of social structure, cooperation in social dilemmas through investments in and returns on social capital', *Social Networks*, **35**, 720–732.
- Roth, A.E. (1995), 'Introduction to experimental economics', in J. Kagel and A.E. Roth (eds.), *The Handbook of Experimental Economics*, Princeton, NJ, Princeton University, pp. 3–110.
- Rubin, D.B. (1974), 'Estimating causal effects of treatments in randomized and nonrandomized studies', *Journal of Educational Psychology*, **66**, 688–701.
- Sarsons, H., K. Gërxhani, E. Reuben, and A. Schram (2021), 'Gender differences in recognition for group work', *Journal of Political Economy*, **129**, 101–147.
- Schram, A., J. Brandts, and K. Gërxhani (2019), 'Social-status ranking: A hidden channel to gender inequality under competition', *Experimental Economics*, **22**, 396–418.
- Sniderman, P.M. and D.B. Grob (1996), 'Innovations in experimental design in attitude surveys', *Annual Review of Sociology*, **22**, 377–399.
- Wallander, L. (2009), '25 years of factorial surveys in sociology: A review', *Social Science Research*, **38**, 505–520.
- Willer, D. and H.A. Walker (2007), *Building Experiments. Testing Social Theory*, Stanford, CA: Stanford University Press.
- Willer, R., K. Kuwabara, and M.W. Macy (2009), 'The false enforcement of unpopular norms', *American Journal of Sociology*, **115**, 451–490.
- Willer, R., C. Rogalin, B. Conlon, and M.T. Wojnowicz (2013), 'Overdoing gender: A test of the masculine overcompensation thesis', *American Journal of Sociology*, **118**, 980–1022.
- Winter, F. and N. Zhang (2018), 'Social norm enforcement in ethnically diverse communities', *Proceedings of the National Academy of Sciences*, **115**, 2722–2727.
- Zschirnt, E. and D. Ruedin (2016), 'Ethnic discrimination in hiring decisions: A meta-analysis of correspondence tests 1990–2015', *Journal of Ethnic and Migration Studies*, **42**, 1115–1134.