202

Annales Universitatis Paedagogicae Cracoviensis

Studia Sociologica VIII (2016) vol. 1

Selected Topics in Experimental Social Science

edited by
Tadeusz Sozański

Strona internetowa czasopisma/Journal website:
www.ifis.up.krakow.pl/studia_sociologica

ISSN 2081–6642

Wydawca
Wydawnictwo Naukowe Uniwersytetu Pedagogicznego
Zapraszamy na stronę internetową: http://www.wydawnictwoup.pl
druk i oprawa Zespół Poligraficzny UP, zam. 58/16

Spis treści/Contents

FOLIA 202

Annales Universitatis Paedagogicae Cracoviensis

Studia Sociologica VIII (2016), vol. 1, p. 5–41
ISSN 2081–6642

WSTĘP/EDITOR'S INTRODUCTION

Tadeusz Sozański
Pedagogical University of Cracow, Poland

Experimental Social Science

Abstract

My editorial introduction to *Selected Topics in Experimental Social Science* – the collection of papers which is the bulk of this issue of *Studia Sociologica* – grew out of the address I delivered at the opening session of the International Symposium on Experimental Research in the Social Sciences, Cracow, June 12–13, 2015. I extended my introductory presentation so much that it turned into a full-size article combining meta-theoretical reflections on theory and experiment in empirical sciences with information on laboratory experiments which were done by the Chair of Research on Group Processes from 1989 until the untimely death of Professor Jacek Szmatka (1950–2001), the founder of that research centre which no longer exists at the Jagiellonian University.

Key words: experiment, methodology of empirical sciences, three generations of sociological theories, network interaction system

Observation and experiment

In his worldwide used handbook, *The Practice of Social Research,* Earl Babbie placed the chapter on experimental method in Part Three which deals with 'modes of observation.' The chapter begins with the statement: 'At base, experiments involve (1) taking action and (2) observing the consequences of that action.' (Babbie 2014, p. 221). 'An experiment differs from other types of scientific investigation in that rather than searching for naturally occurring situations, the experimenter *creates* the conditions necessary for observation.' (Aronson et al. 1990, p. 11). In fact, while any *empirical science* rests on the *observation* of *regularities, experimental science* combines observation with a planned *intervention* in the *natural* course of events. Sometimes such an intervention is possible, sometimes it is not. One could drop balls, as did Galileo, from the Leaning Tower of Pisa, but even today no one is able to remove one planet from the Solar System to observe the System's behaviour after such an intervention.

In *social science* the range of actions you can take to observe their consequences is fairly wide, even though it is additionally limited by ethical concerns. But what do we learn more from experimenting than from observation alone, or why do we need to do experiments? (Webster, Sell 2014). Let us illustrate the problem with probably

the oldest social science experiment described by Herodotus in 5th century B.C. and commented by Antoni Sułek (1989). The starting point for the experimenter, the pharaoh Psammetichus of Egypt, was the fact familiar to everyone: almost all people learn their first (ethnic) language from their biological parents. Observation certainly confirms this *statistical regularity.* Observation would probably suffice to provide enough evidence supporting the *general law* which states that *every* new-born child is able to learn *any* human language through communication with *any* speaker of this language. We also know from observation that communication between a child and the person who uses his or her ethnic language to address the child during daily contacts is in its early phase highly asymmetric with respect to speaking-to-listening proportion and initiating sequences of utterances. Therefore, it comes as no surprise to expect that the less active interaction partner will acquire the language of the more active one. Nevertheless, a question arises to which an answer cannot be arrived at without breaking the natural order. The question is: What will happen if the caretaker does *all* what must be done to meet vital needs of a new member of the human race (including establishing social bonds with him or her on the *preverbal* level) with the only *difference* that consists in refraining from speaking to the baby but waiting instead for it to start verbal communication.

Psammetichus *assumed* that any new-born would start speaking *a* human language at some stage of its normal development. With this (theoretical?) assumption he could seek an answer to the question of *which* language the child would speak, which kind of *exploratory research* is stronger than a mere attempt to learn *what* will happen when the natural process is blocked. Actually, the pharaoh's ambition was to carry out a true experiment aimed at *testing a hypothesis*. His hypothesis, which was not derived from any *general theory*, was very *specific* as it claimed that *everyone* would speak Egyptian provided that the ability to speak that language was not overridden by forced reception of a stream of words in another language. Sułek (1989, p. 650) praised Psammetichus for acknowledging – which is by no means a rule for the rulers – the negative result of his test. The first utterance of the *experimental subject* was recognized by the *experimenter's confederate* as the name of bread in the Phrygian language.

The story on the pharaoh-experimenter illustrates the advantage of experiment over passive observation. We learned from that ancient study and its subsequent *replications* by other rulers (Sułek 1989, p. 647) that the assumption of innate ability to speak a *concrete* human language was wrong, which of course does not invalidate a general *paradigm* that tells us to look for *innate* sources or determinants of social behaviour.[1]

[1] Hamlin, Wynn, and Bloom (2007) have shown that preverbal infants can recognize and distinguish between three abstract types of *social actions* possible for an actor *A* who sees an actor *B* trying to achieve a goal (*A* can *help B*, or *hinder B*, or *stay neutral*). Moreover, the authors infer from certain nonverbal responses that 'infants prefer an individual who helps another to one who hinders another, prefer a helping individual to a neutral individual, and prefer a neutral individual to a hindering individual.' They claim that the evidence they

Definitions of experiment point to other important characteristics of this type of investigation. Every time I teach 'Methods of social research' to sociology students I quote at the very beginning of my lecture on experimental method a definition which comes from the book by Antoni Sułek (1979, p. 15). In English translation his definition reads as follows (italics mine).

> An experiment is a *repeatable* procedure that consists in a *planned change* of some factors in a situation under investigation and simultaneous *control* of other factors, a procedure that is performed in order to learn from *observation* the answer to the question of what are the *consequences* of that change.

The term *control* is often referred to the *power* over the *whole* setting or situation under study, including the ability to change values of some *variables* (*factors*) which work in the situation. The definition I quoted makes a distinction between *experimental manipulation* ('planned change') and proper *experimental control* ('control of other factors'). *Manipulation* means the ability to endow *units of analysis* (usually individuals or groups) with values of *independent variables.* The experimenter's power consists in that it is up to him which value from a specified range is assigned to any unit of analysis. Values are created by performing certain operations (*experimental treatments*) on the units or placing every unit in one of few *experimental conditions.*

In a narrower sense, *experimental control* reduces to eliminating possible effects on the dependent variable of variables other than the independent variables. This purpose can be achieved in a few ways[2] of which *randomization*, or *random assignment* of units to conditions, has the widest applicability; in addition, it enables controlling variables unknown to the experimenter.

The third component of Sułek's definition is hidden behind the words 'observation' and 'consequences.' It is the *measurement of the dependent variable*, an operation that is done *after* manipulation to see the effect of 'planned change' of values of the independent variable/s on the dependent variable. All three components were explicitly named by Jerzy Brzeziński (1996, p. 286) in his definition of 'experimental model of testing hypotheses on the dependence between the dependent variable/s and independent variable/s.' In his handbook of research methods in psychology, the 'experimental model' appears as one of three 'models of *testing hypotheses*'; the other two are *multiple regression* model and *ex post facto* model; the meaning of the latter is similar to that of *correlational study* (Aronson et al. 1990, p. 28–31), which

---

gathered supports the hypothesis that some elementary moral evaluations are innate rather than learned.

[2] Some variables can be controlled by disabling their action in an *artificial* environment. For example, the use of a computer network in an interaction setting instead of face-to-face contact eliminates many variables characterizing communication partners. You can even hide from them their gender if they are forced to communicate with each other by means of a special code instead of a natural language (many ethnic languages allow their users to recognize the gender of one's communication partner from the grammatical forms he or she uses). Other ways of experimental control are: *holding variables constant* (Aronson et al. 1990, p. 18–20) and *matching* (p. 148–150).

is in fact a kind of observation as the values of *all* independent variables are only *registered* by the researcher, whereas in any experiment at least one independent variable must be *manipulated*.

Testing hypotheses – as a more ambitious epistemic goal than a mere *description* of some regularity – is a no less important property of experimental research than the investigator's ability to *construct* a completely or partially artificial, relatively isolated setting, and to trigger, control, and measure certain processes that occur therein. Under such a broad understanding (more general than the one implicit in the aforementioned definitions) of experimental method, the diversity of *experimental designs* stems from various ways in which testable hypotheses are being formulated in the *empirical sciences*.

A hypothesis may have the form of a *prediction* that a definite *phenomenon* will occur whenever certain conditions are met.[3] Another simple experimental design consists in measuring the values of the same variable Y (e.g., body temperature or blood pressure) twice for the same set of units of analysis (e.g., a group of students) where the first measurement is taken *before* and the second *after* performing an operation on these units (e.g. having the students to read an exciting story). To rule out alternative explanations of the *change* (predicted by the hypothesis to be tested) in the average level of Y, a *control group* is often needed besides the *experimental group*. In the control group, Y is also measured twice but between two measurements no action is taken that could change the state of the experimental system. The two groups are formed at the very beginning of the experiment by dividing a *random sample* taken from a *population* (the one for which the hypothesis to be tested is expected to hold true) into two subsets by means of a chance mechanism (say, flipping a coin that guarantees that every unit is equally likely to become a member of either group). The *classic design* thus obtained can in some cases be simplified by skipping the first measurement in both groups. The difference between 'action' and 'no action' can be interpreted in turn as a difference in the values of a 2-valued variable X, which then becomes an *independent variable* in relation to the *dependent variable* Y.

Many experimental designs involve a comparison of *mean values* of the dependent variable across conditions. A *comparison of two means* is also at the core of the first social science experiment which was run in the 1880s by Maximilien Ringelmann, French agricultural engineer.[4] In the decade which saw inventing the first automobile, he embarked on an examination of pulling efficiency of horse teams and discovered that the mean force of a team (the mean is obtained by dividing the overall force of a team by the number of its members), was always lower than the mean computed from the values obtained separately for each team member.

[3] For example, when a group of persons are given an opportunity to report aloud their assessments of a stimulus which they all are exposed to, then there arises – from individual judgements via the interaction process – a group norm which subsequently affects individual perceptions (see the description of Sherif's experiment in Cartwright, Zander 1960, p. 23–25).

[4] His report was published in 1913 to the effect that the birth of experimental *social* psychology is usually traced back to Triplett's study (1898) which gave rise to the *together and apart* paradigm, the dominant paradigm in early experimental social science. See Brown 2000, Chapter 5.

Interestingly, Ringelmann employed students instead of horses. He told them to pull a rope tied to a dynamometer, first *apart* and next *together* in teams of varying size. He could *communicate* with his experimental *subjects* because they, like him, were *human beings.* Without their cooperation he would have been unable to carry out his research.

The necessity of communication between the experimenter and the people whose behaviour is to be studied may also entail some undesirable consequences that do not arise in *experimental natural science* where experimentation consists in the measurement of a number of variables in strictly controlled laboratory conditions. Research methods, such as testing hypotheses on *cause-effect relationships* between variables and designing experiments so as to compare groups with respect to the average level of dependent variable, prevail in the social and behavioural sciences. They are also commonly used in other empirical sciences along with more advanced ways of producing scientific knowledge. Physicists do not compare group means. They check if the readings from measurement instruments agree with the theoretically predicted values, which are *calculated* from the formulas expressing functional relationships between variables chosen to describe the current state of a physical system, such as a falling ball.

Such a 'hard' approach to *theory building* and *theory testing* has been recommended for use in the social sciences by Willer and Walker (2007). Drawing on an earlier book by Willer (1987), they distinguish *theory-driven* experiments from *empirically driven* experiments. According to them, these two types of experiments essentially differ on the level of the very logic of scientific investigation. Shane Thye (2014, p. 74–76) denies the radical nature of the opposition, noticing that both types of experiments face similar problems to cope with such as threats to *internal validity* from *confounding factors*. Although I share his view, it is not my intention to belittle the importance of the distinction made by Willer and Walker. It reflects the dissimilarity in a few respects of two ways of sociological theorizing described by Szmatka and Sozański (1994). I return to this topic later in this paper after delineating in the next section a broader meta-theoretical context in which the purport of experiment as a method devised for testing hypotheses in any *basic empirical science* can be properly understood.

Basic characteristics of the basic sciences[5]

Every basic science, no matter whether formal or empirical, natural or social, 'is oriented to the production and evaluation of knowledge claims' where the term knowledge claim is referred to any statement which 'can be accepted or rejected on the basis of some criterion of truth.' (Cohen 1989, p. 52–53). Methodology of the basic sciences formulates epistemic criteria for evaluating solutions to scientific

[5] In this and the following sections I use re-edited excerpts from Chapter 1 ('Structural Mathematical Sociology') of my still unfinished book (*The Mathematics of Exchange Networks*). The full text of Chapter 1 is available on my personal website (http://www.cyfronet.krakow.pl/~ussozans/chap1.pdf).

problems. While the form of the problems considered tractable in any particular science may also be subject to meta-theoretical analysis, the range of these problems and their substantive content is always determined by one or more paradigms, a *paradigm* being defined roughly as a set of guidelines, accepted by the academic community, as to what and how can be studied in a given discipline or subdiscipline.

What are the distinguishing features of *science* as a special kind of *knowledge*? The answer is that scientific knowledge already is or should be developed so as to be:

> (1) *intersubjectively communicable*; (2) *methodically* produced and validated; 3) *systematized*; (4) *consistent*; (5) *logically provable* or *empirically testable*; (6) as *certain* as possible; (7) rich in *information*; (8) *universal*; (9) *general*; (10) *precise* and *accurate*; (11) *parsimonious* and *simple*; (12) *abstract*; (13) *conditional;* (14) *cumulative*.

Most of these characteristics are also included in Markovsky's (1997) list of the criteria for evaluating scientific *theories*. *Intersubjective communicability* is achieved in each discipline through *codifying* its language, or establishing clear, workable criteria upon which meaningful statements can be distinguished from those recognized as meaningless. Codification of *scientific discourse* inevitably leads to supplanting natural language by artificial *formal languages* in which complex expressions are built from simpler ones by applying to them certain explicitly stated *rules* so that meaningful statements are recognized from their syntactic structure. Formalization of the *syntax* (relations within a system of signs) is a necessary step preceding the codification of two other aspects (distinguished by Morris in his *Foundations of the Theory of Signs*, 1938) of any language (more generally, any *semiotic system*), *semantics* (relations between language expressions and the objects in the 'world' to which they refer) and *pragmatics* (relations between a language and its users).

Scientific knowledge should be produced *methodically*, even if it ultimately grows out of unplanned discoveries of new facts or new conceptual representations of known facts. *Methods* are prescriptions on how to perform various activities at every stage of the research process, primarily at its last and most important stage when knowledge claims are validated upon 'some criteria of truth.' In the *formal sciences*, a knowledge claim is *accepted* if and only if it can be deduced from already accepted claims by means of logical *rules of inference*. The *deductive method* is also used in *empirical sciences* along with *empirical testing* (in particular, *experimental method*), a way of validating knowledge claims which is peculiar to these sciences. By requiring scientific knowledge to be produced methodically, we also mean that the *evidence* needed to test a hypothesis must be collected with the use of intersubjectively controllable *data generation procedures*.

Science also differs from common-sense knowledge in the degree of *systematization*. This requirement pertains both to terms and propositions, two basic components of any knowledge. *Terms* are names of things, properties, relations, functions, and other constructs studied in a given field. *Propositions* (sentences), as formed with the use of terms, constitute the higher level of the language. What is even more important, they are conceived of as statements which can be *true* or *false* in a given *domain* in which the terms occurring in them are *semantically interpreted*.

Collections of terms and propositions should be structured so as to form *terminologies* and *theories*.

Contradictory hypotheses may coexist in science, yet among the propositions that are accepted in a given discipline there should never be two sentences such that one of them is the negation of the other. *Consistency*, defined by this requirement, is the most fundamental condition any jointly accepted collection of knowledge claims must satisfy. In particular, every scientific theory should be consistent.

We require scientific knowledge to be *intersubjectively provable* or *testable*, but we have to acknowledge the fact that all proofs and tests are *relative*. In the *formal sciences*, a hypothesis is accepted as a *theorem* if there exists its demonstration based on explicit specific axioms whose consistency is usually justified by invoking a more fundamental theory. The *empirical sciences* use the *deductive method* too – as a way to derive consequences from already accepted theoretical propositions and as part of testing procedures.

To *test* an empirical theory, one must first identify a number of situations that meet the theory's *scope conditions* and admit of gathering *evidence* indispensable for validating theoretical predictions. The scope conditions (see Cohen 1989, p. 83; Foschi 1997) determine the range of systems to which the theory applies; they can also specify special system states or some additional circumstances in which theoretically predicted events should occur. Since empirical systems that meet all scope conditions are seldom found in nature, one cannot do without constructing fully or partially artificial systems. Created by the researcher, they are easier to study than natural systems but are no less real than the latter.

For any *empirical system* that meets a theory's scope conditions, one must state some more or less specific *hypotheses* concerning its predicted 'behaviour.' Hypotheses should be derived then from the theory, supplemented, if necessary, with auxiliary assumptions which may point out operational counterparts of *theoretical variables.* To test theory-based predictions and thus the theory itself, one must *observe* and register actual behaviour of the system under study; observation usually amounts to measuring values of some variables. If the observed behaviour of the system agrees with the predicted behaviour *within the margin of error*, then the theory is said to have been corroborated by the evidence generated to test it.

If observation of a 'natural' course of events cannot provide sufficiently rich and unambiguous evidence, one has to create an artificial setting in order to give nature an opportunity to speak in a more extensive or more articulate way. In either case, the researcher must devise a procedure to *generate evidence* interpretable in the context of his or her theory, or a procedure for translating the cues emitted by the external world into meaningful *data*. In an ideal world, such a procedure would be dictated by the theory alone. In the real world, it should be designed so as to minimize 'error' or 'noise' occurring also in experimental systems as they are made from the material taken from the real world and are never completely protected against the influence of the external environment.

Given an adequate research design and reliable measurement techniques, the outcome of a test should depend on whether the theory undergoing verification correctly depicts regularities operating within a well defined category of things or

events. An empirical theory must be supported by the evidence in a number of tests to get incorporated into the body of established knowledge in a given discipline. Theoretical propositions which have been accepted and have few other desirable properties (universality and generality being considered most important) are called *laws*. Once accepted, an empirical law can be applied outside the setting in which its predictive power has been confirmed. Although our confidence in a law grows with each successful application thereof, the *certainty* characterizing mathematical knowledge can never be attained in empirical sciences. While a mathematical theorem, once correctly demonstrated, will be accepted forever, laws in empirical sciences are vulnerable to refutation. However, an empirical law need not be automatically discredited if negative results of further tests raise doubts about its validity. If some observations depart from the predictions deduced from a well established theory, the first suspicion is that the theory has been incorrectly applied. Such an explanation is possible because scientific knowledge is necessarily *conditional* (Cohen 1980), that is, any scientific knowledge claim is applicable only if definite scope conditions are met.

The core laws of an empirical theory that are protected from hasty falsification are called *principles.* Their epistemic status is the most contentious issue in the philosophy of science. While for 'realists' principles render objective *regularities*, for 'conventionalists' – also called 'instrumentalists' – they are but *tools* invented to enable a selective, concise and coherent account of the data. Willer and Walker (2007, p. 59) ask 'What, then, does theoretic science assert about the regularity of the world?' and answer 'It claims that whether the world is regular cannot be judged independently of the theories through which the world is understood' (p. 59). Such an answer shows authors' sympathy for the instrumentalist meta-theoretical stance, however it is expressed less radically than in earlier statements (to be quoted later in this paper) by Willer himself (1987).

As Imre Lakatos noticed (1970), an empirical theory does not drop out of the corpus of accepted scientific knowledge because of being simply falsified. Once approved, a theory is abandoned only if it can be replaced by a new theory which accounts for all the facts explained by the old theory as well as for some facts that the latter cannot explain. It is the strongest meaning of the postulate that scientific knowledge should grow *cumulatively*.

Every *investigation*, scientific or judicial, theoretically or practically oriented, is aimed at reducing cognitive *uncertainty*, first of all, in any situation where hypothetical answers to a question are known, but one is not sure which of them is true. 'In a somewhat aphoristic form, science is an information-seeking process' (Szaniawski 1976, p. 297). In the light of formal *information theory*, richness of *information* and *certainty*, items (6) and (7) on our list of the goals pursued by science, turn out to be conceptually intertwined. However, their understanding must remain intuitive until an *intersubjective* practical method for measuring epistemic probability becomes available. In general, the pragmatic aspect of the language of science admits of limited codification, which opens the door for sociological interpretations of methodological rules as mere norms or conventions approved by academic communities.

Universality and *generality* are two qualities that distinguish *laws* from other accepted scientific propositions. The broader the scope of a theory, the more general the theory is regardless of the nature, abstract or historical, of entities it deals with. In logic, the term *general statement* is referred to any proposition stating that *all* things have a property *v*. The derivation of a particular conclusion from a general statement is probably the most familiar pattern of deductive reasoning ('All men are mortal, therefore I am mortal'). Generality is in fact a semantic concept because the phrase 'all things' acquires a definite meaning with pointing out a set *S* whose elements (or rather their names) are to be substituted for *s* in the proposition 'for all *s*, *v*(*s*).' *Laws* are usually construed as *strictly general* statements, which means that they should hold true in *domains* with infinitely many objects, or indefinitely many (however large is the set of all men who have ever lived or will live on earth, it is finite).

Universality should not be confused with generality. 'A universal statement is a statement whose truth is independent of time, space, or historical circumstance' (Cohen 1989, p. 78). To ascertain whether an empirical theory is universal, one must test it in at least two settings that differ with space-time or sociocultural coordinates. In the social sciences, 'the cross-national and cross-cultural replication experiment is the only method of testing a theory for universality' (Szmatka 1997, p. 95).

According to Cohen (1989, p. 178), universality and deductive systematization are both required of a collection of conceptually interrelated testable statements in order that it can be called an *empirical theory*. If universality is skipped as a too restrictive condition of theoreticity, it returns as the basis of the traditional distinction between *nomothetic* and *idiographic* (*historical*) sciences, the former being defined as those capable of producing universal theories. The scope conditions of a universal theory do not state *when and where* in the *real world* to find systems to which the theory applies. Nevertheless, one must show that such systems do exist because otherwise the theory would not be testable. An empirical theory need not claim universality. In order to be testable, it must also have a definite scope that is specified by indicating the time, place, nation or culture where the theoretically predicted regularities should occur.

Attempts to generalize a theory as much as possible and make it universal may result in disregarding other, no less important, goals of science that are usually easier to achieve under more restrictive scope conditions. Generality and universality really count only if they go together with *precision* and *accuracy*, as is the case with Newton's laws of motion, which not only apply to a broad class of mechanical systems, but yield *specific, quantitative predictions* which agree remarkably well with measurement results. 'Although a theory may generate predictions that are highly precise, the *accuracy* of those predictions – their correspondence to empirical observations – may vary' (Markovsky 1997, p. 19). There exist sociological theories which offer exact predictions of the behaviour of some social systems, yet the gap between observed and predicted results is often too wide and contingent on uncontrollable events. Hence, the social sciences on the whole cannot yet be counted among *exact sciences*, or those nomothetic empirical sciences that meet the standards of precision and accuracy to a high degree.

In all empirical sciences, the quest for precision forces the transition from concepts to variables. To transform a *concept* into a *variable*, one must first select an appropriate *unit of analysis* (the concept may admit of more than one option in this matter). Next, the *domain* of the variable should be pointed out – as the set of objects varying in the respect considered important by the researcher. Lastly, there must be invented a way of assigning *values* (usually numerical) to the elements of the domain. Constructing variables and theory building always go together. While in well-developed sciences this takes the form of *fundamental measurement* based on laws relating theoretical quantities to each other, in the social sciences the prevailing approach is *measurement by fiat*, as Torgerson (1958, p. 21–25) called taking an *operationally* defined variable to represent a *latent* theoretical variable on the basis of '*presumed* relationships between observation and the concept of interest' (p. 22).

In formal *set-theoretic* terms, a *variable*[6] is a *mapping* of a set of objects under study into a set of numbers. In the empirical sciences, variables are used to formulate theoretical hypotheses and their directly testable consequences. What can be studied for a single variable is merely the *distribution* of its values assumed in a set of objects (the whole population or a *sample* taken from it). Given two or more variables, one wants to know how their values *co-vary* over the common domain. To construct a theory whose propositions have the form of interrelated 'covariance hypotheses' (Blalock 1969), one has to select a set of variables and decide which of them are to play the role of *independent variables* in relation to the remaining variables called *dependent*; it is a matter of theory to *predict* values of the latter from the known values of the former. If there are few independent variables, they are assumed to vary independently of one another, which in an experimental setting should be guaranteed by a proper *research design*. Even though theories in empirical sciences are often constructed so as to render *causal* linkages among variables, it is the concept of *dependence* (statistical or functional) rather than causality that is given a more technical meaning in theory and research.

Patterns of theorizing and experimenting in social science

Presuming that the context will steer the reader to the proper understanding of *science* and *social*, I have not yet explained what is meant in this paper by *social science*. The singular is used to highlight *methodological* unity of social sciences, as well as to leave aside the question of where to trace out the borders between social psychology, a predominantly experimental science, sociology, and economics. *Substantive* unity of 'social science' is founded on making interaction of members of the human species the most elementary object of investigation. Any social scientific study of the processes going on between two or more persons must take into account not only natural (physical or biological) aspects of these processes but also

[6] Variables – in this meaning – should not be confused with logical variables. The latter are symbols (in formal languages) or common nouns (in natural languages) that enable us to speak of things, points, numbers, or other entities without the necessity to point out concrete elements of appropriate sets.

the very fact that people communicate with each other with the use of certain codes (systems of signs) which are part of *cultural reality* (Znaniecki) equivalent to the Popperian 'third world' (see more in the chapter mentioned in footnote 5).

'Science can be thought of as consisting of theory on the one hand and data (empirical evidence) on the other. The interplay between the two makes science a going concern' (Torgerson 1958, p. 2). The meaning of 'science' in the expression 'social *science*' is as *broad* as in the cited statement. The saying attributed to Rutherford – 'In science there is only physics or stamp collecting' – denigrates many empirical sciences, which, unlike physics, do not yet meet and some possibly will never meet all of the 14 requirements I listed at the beginning of the previous section (the last two of them I have not yet mentioned are (11) *parsimony* and *simplicity*, and (12) *abstractness*).

Contemporary mainstream *social theory* has gone far away from positivist and postpositivist (Popper, Lakatos, Toulmin) meta-theorizing on social science. What is labelled as *positivism* is criticized either for theory-free 'stamp collecting' or importing to the social sciences the patterns of doing theory that are believed to be unique to the 'natural sciences.' While some views traditionally associated with positivism, such as the idea of *theory-free sense data*, deserve rejecting outright for having little to do with real science, a few other postulates, also considered untenable by leading figures of contemporary social theory, are worthy of defence (Turner 1985). I mean, first of all, the principle of *demarcation* between *empirical* and *formal* sciences on the one hand, and *hermeneutic* or *philosophical* sciences on the other.[7] The demarcation principle does not remove from science the questions of existence. Kurt Lewin was right to claim that 'The taboo against believing in the existence of a social entity is probably most effectively broken by handling this entity experimentally.' (Cartwright, Zander 1960, p. 18).

Experimental testing empirical theories in exact sciences resembles demonstrating consistency of *formal theories* through *constructing* their *semantic models.* Similarly, the experimenter's task is to build an *empirical system* in which *observational statements* derived from the theory are true. Theoretical predictions, or *empirical consequences* of a *formalized empirical theory*, are *deduced* from the formal theory (the one which was used to formalize the empirical theory that is to be tested) and certain rules linking abstract objects and variables with their observable counterparts.

[7] The recent dispute in Poland over the prerogatives of the Constitution Court encourages non-lawyers to raise the problem of what epistemic status should be attributed to the *legal sciences*. As a sociologist and mathematician, I would like to know if the *science of law* is a formal science or an empirical science. If neither of the two is true, should *assessing consistency* of bills with the constitution be regarded as a task requiring *philosophical* competence? Do the experts in constitutional law who are making judgements in such matters resort to yet another kind of knowledge? When I found convincing, however on a purely *intuitive* basis, some arguments – presented by a few lawyers with academic degrees – in defence of the position of the government and ruling majority, I asked an eminent professor of sociology to let me know his position in the debate. He replied to my letter by sending me *solely* the list of outstanding professors of law who used their *scholarly authority* to back the parliamentary opposition and the chairman of Constitution Court. My curiosity about the nature of the legal sciences remains unsatisfied.

Physics has always been perceived as an embodiment of the *ideal type* of *exact* science. *Sociology*, first called by its father 'social physics' (Comte abandoned this name, having noticed that Quetélet used it to denote the study of *statistical* regularities) emerged from social philosophy to gradually achieve the status of a *normal* empirical science ('normal' in the sense proposed by Thomas Kuhn in *The Structure of Scientific Revolutions,*1962). The publication of *Le Suicide* (1897) played an essential role in that process. The research paradigm underlying Durkheim's landmark work does not envisage experimental testing theoretical hypotheses on the dependence between the phenomena abstractly defined in his theory. While *sociology* has remained until today an overwhelmingly non-experimental science, an experimental design is constitutive of many paradigms in *social psychology*, including the *together and apart* paradigm, which appeared in social science at more or less the same time (Triplett published his paper in 1898) as Durkheim's *sociological* theory relating the frequency of suicide acts to the level of social integration of a group. The experimental paradigm in question consists in *comparing individual* performance measured in the baseline situation in which the person is set to work alone with the performance observed in the situation where the task is being done by the person in the *presence* of another performing the same task simultaneously. Whereas the baseline situation is defined unambiguously, the 'social' ('co-action') situation, for being defined as mere presence of the other person doing the same, can be enriched with additional characteristics bearing on further theorizing inspired by *inventing* the paradigm. For instance, the experimenter may encourage the subject to compete with the co-actor as was in the case of original Triplett's experiment. Thus the paradigm leaves room for introducing into the experimental setting manipulable factors to learn the sufficient and necessary conditions for the effect of *social facilitation* (significantly better performance in the social situation) to occur. Willer and Walker (2007) point to the advantages of *theory-driven experimenting.* The idea of *experimentally driven theorizing* is no less promising and compatible with the practice of social research.

Robert Merton (1968) saw in Durkheim's suicide theory a classic example of a 'theory of the middle range.' He believed that theories of this kind would successfully challenge 'total systems of sociological theory' as he called *conceptual* images of the social world. Such a general conceptual framework may lead to formulating proper theories (a *theory* must contain interrelated propositions apart from concepts) explaining some phenomena. Szmatka and Sozański (1994, p. 225–231) called that product of old and new sociological theorizing – also known as 'grand theory' – *theories of the first generation*. These theories are *abstract* (they contain terms like 'social system') but suffer from the lack of testing procedures and explicitly stated scope conditions. Sociological theories that are free from these deficiencies form two other 'generations.' Since the latter word suggests the process of replacing old products with new ones, Szmatka and Lovaglia (1996) changed 'generation' to 'genus' to concede that all three kinds of theorizing co-exist in contemporary sociology and none of them is going to supersede others in the foreseeable future.

Theories of the second genus are expected to provide a systematic account of multidimensional differentiation that is actually observed in natural social settings and concrete populations where regularities usually occur in a blurred form due

to complex and casual ties within the multitude of variables. If the main sources of variation and specific patterns of dependence cannot be identified prior to data collection – for the lack of a 'theoretical model' – one may try to construct a 'methodological model' (Skvoretz, Fararo 1998), or try to extract regularities directly from the data by means of standard procedures of multivariate statistical analysis. The choice of variables is then subordinated to the main goal defined as explaining the largest possible share of the total variance of each dependent variable. An experimental test of a theory of the second genus takes the form of *factorial experiment* classified by Willer and Walker as *empirically driven experiment*.

Theories of the third genus unlike the theories of the second genus are *abstract* and claim *universality*. They are constructed with the aim of bringing the social sciences closer to the *exact* natural sciences. While precision and accuracy are highly desirable properties, the focus is on parsimony and simplicity. The postulate of *parsimony* states that in science 'entities must not be multiplied beyond necessity' where 'entities' may be primary terms, axioms, laws, variables, etc. The postulate of *simplicity* means, in particular, preference for the use of simple functions or formulas to describe inter-variable relationships. The power of exact sciences lies in that generality and universality, parsimony and simplicity need not be sacrificed for the sake of precision and accuracy.

Each theory of the third genus describes the behaviour of a class of abstract or *ideal* systems by means of a *small* set of *theoretical* variables. Some of them, though not necessarily all, must have *observable counterparts* in *empirical replicas* of abstract systems. In regard to *natural sciences*, Toulmin (1953, p. 44–56) used much similar criteria to contrast 'physics' with 'natural history.' 'Natural historians' want to *explain* facts they observe in the world. To do this, they invoke 'general laws' of the form '*all* As are Bs.' 'But so long as one remains within natural history there is little scope for *explaining* anything: "Chi-chi is black because Chi-chi is a raven and all ravens are black" is hardly the kind of thing a scientist calls an explanation.' (Toulmin 1953, p. 49).

Lee Freese (1980, p. 191–192) presented a similar distinction between the 'generalizing view' and the 'instrumental view' of theories and laws.

> Laws ... are not meant to be generalizations about the world of everyday experience. The regularities they describe exist in a theoretically possible world but not in the actual world. … If theories are construed as describing some idealized state of affairs in a closed system ... then they [laws] are devices for calculating changes in the system when other things are equal. Though other things are never equal outside of the closed theoretical system … laws may serve as tools for engineering some change in an open empirical system whose departures from some theoretically true state of affairs can be measured.

The *instrumental* view of laws may appear incompatible with the realist stance in the philosophy of science. Actually, a law, which in its abstract form applies *directly* to a class of 'theoretically possible' systems, applies *indirectly* to relevant real-world systems. Its successful indirect application to an 'open system' should be possible due to universality. However, even in laboratory systems the impact of extraneous

variables can be so strong that the law fails to provide accurate predictions. Szmatka and Sozański, referring to Willer statements (1987, p. 221), addressed this problem in the following passage (1994, p. 230–231).

> In a laboratory system, the experimenter can, to be sure, control the structural conditions of human actions but must always fill positions in the system with concrete individuals shaped in a particular sociocultural context. 'Why is it then that Galileo did not consider the colour of his shirt or the phase of the moon when he evaluated the results of his trajectory experiments?' (Willer 1987), and why do sociologists, in order to explain the behaviour of experimental subjects, sometimes need to consider such factors as personality or situation variables thought of to be 'at work' in a given setting? 'The answer does not lie in the difference between animate objects which we investigate and the inanimate objects which he investigated. Instead the answer lies in the evidently clean results of his experiments and in the fact that they could be reproduced by him or *by others* as needed.' (Willer 1987).

Why are *some* empirical sciences able to produce general and universal, precise and accurate, parsimonious and simple theories? Certainly, the ability to obtain 'evidently clean results' in repeated experiments depends to a high degree on the choice of a suitable *mathematical representation* and *research design.* According to Willer (1987, p. 220), what makes an exact science exact 'is the exact use of theory, not necessarily the exact production of clean results … the criterion should be that a better theory is one which *can* produce cleaner data, not that it would always do so.' However, a *precise* theory becomes practically useful insofar as it can provide *relatively accurate* predictions *relatively independently* of the context in which it is being applied every time. If very restrictive conditions need to be imposed in order to produce sufficiently 'clean' data, then the theory becomes useless outside the setting in which it has passed the experimental test, that is, outside the setting in which prediction accuracy has reached the level considered satisfactory in a given discipline. Hence, there is another methodological standard that *experimental exact science* must meet besides high precision and accuracy. The results of experimental tests should be *stable*, which means that a small a change of the setting in which a given regularity has been detected in its purest form should cause a relatively small decline in prediction accuracy.

Regularities in the social world

The 'criterion of truth' upon which scientific knowledge is validated is *coherence* of theory and evidence. However, once experimental evidence is *produced* by the researcher, one may be interested to know to what extent coherence, desired so much, depends on '*building* the experiment,' and to what extent it hinges upon the existence of some *regularity* or 'order' in the world out there. The passage quoted below (Willer 1987, p. 12–14) documents that Willer would like to dismiss the question but in the last resort he tends to attribute more creative power to the theorist-experimenter than to the world, thus subscribing to the viewpoint of instrumentalism.

> Within the process of scientific inference, no assumptions are made concerning the regularity or irregularity of the world. No such assumptions are needed because the relations among objects and events are first drawn in theory and only then compared point by point to bits of information from the world. … Does replication [of experiments] prove that the world is regular? No, for replication proves only that theory can so organize the world and our view of it that at least some parts of our perceptions can be made to appear regular – and that is quite another thing.

Willer is right to say that exact sciences do not start from the *assumption* that the world is regular. The *hypothesis* of regularity is arrived at through systematic *observation.* Ancient astronomers *did not assume* that celestial bodies behave regularly. They *discovered* that the position of these objects in the sky at any moment can be predicted with great accuracy. The discovery of a 'natural order' in some area of the social universe may result in formulating a theory of the second genus. For Jacek Szmatka it was no more than the first step. He believed that any scientist oriented toward 'hard science' should attempt to explain any regularity by offering a theory of the third genus, a testable, general, universal, precise theory that *abstracts from* particular occurrences of the regularity in concrete *empirical objects.* When we discussed the problem of how second and third genus theories are (or should be) related to each other, I argued that a move in the opposite direction, the transition from a given third genus theory to a second genus theory may appear necessary in some circumstances. When a theory of the third genus fails to provide accurate predictions, or, as Willer would say, when the data from a theory-driven experiment are not 'clean' enough, then one may try to 'improve' the theory – at the cost of 'spoiling' it in other dimensions (universality, parsimony) – by appending certain variables that do not fit the abstract theoretical model but make it possible to reduce unexplained variance. For example, a theory that is intended to predict outcomes of a game played by rational actors can be modified by adding actors' gender to the set of variables which are suspected to affect the decisions made by the players.

Some advocates of the *idealization strategy* claim (Wysieńska, Szmatka 2002) that testing third genus theories is conducted within the 'theory world' that transcends the concrete 'external, phenomenal reality.' Actually, *ideal* systems which serve as models of *empirical* systems are part of the mathematical world as they are *sets endowed with structures* (Bourbaki's term; see the chapter mentioned in footnote 5). *Laboratory replicas* of abstract systems do not differ in the stuff they are made of from empirical systems studied by the theories of the second genus. It is not true that 'the social laboratory, unlike the physical laboratory, may be cleanly separated from the phenomenal world outside' (Willer 1987, p. 214). Willer would be right if *live* subjects were replaced by computer programs, yet *simulating* a theory-predicted process is not equivalent to *testing* the theory. The 'theory world' can only be conceived as one of mathematical domains and set-theoretic constructs. Having entered this world, you can verify *logical consistency* of a formalized empirical theory, which, once *formulated*, has to be confronted with the data coming from the world we perceive with our senses and transform with our actions.

Theory and *evidence* should be conceived of as *two* distinct *independent* sources of information about *one* world of experience. These sources must not be attributed equal credibility as the 'voice' of the data should always count more. *Independence* does not mean that the data generation procedures must be 'theory-free.' Even a police *investigation* into a crime is not confined to gathering facts connected somehow with it. The investigation is 'driven' by the prosecutor's theory, which of course shall be modified as new facts are becoming known. In exact sciences, *relevant* experimental evidence is generated through fundamental or derived measurement. The instruments with which theoretical variables (or their empirical realizations) are measured are themselves constructed according to the prescriptions based on the theory being tested.

Physics and sociology differ in entities studied, variables chosen to describe them, paradigms, theories, and data-generation procedures. Do these sciences also differ in *general* patterns of theory testing? Let us compare a sociologist studying a task group in a laboratory with a physicist investigating the motion of a bullet. Both experimenters can trigger off some processes in empirical systems whose behaviour is going to be observed, yet the physicist cannot *tell* the bullet to move along the theoretically calculated curve, whereas the sociologist, owing to his ability to communicate with human agents, can make them familiar with his theory and induce them to behave accordingly. If we catch a sociologist talking experimental subjects into the behaviour predicted by his theory, should we blame him of violating a *methodological norm* or should we rather recognize his *communicative action* as a 'legal' way of testing a *sociological* theory?

If the only purpose of an experiment were to 'reproduce' the *form of a regularity*, then it would suffice to *simulate* theoretical behaviour in a 'virtual system' where 'virtual' does not mean 'imaginary' or 'mental.' A 'virtual system,' on the one hand entirely artificial, is 'real' as constructed within the real world with the use of technical devices. For example, a *virtual dyadic social system* can be composed of; (1) two interacting programs running on two networked computers; or (2) an individual interacting with a computer program or even a pair of persons – provided that *live* human agents, even though they act 'consciously,' have been 'programmed' by the experimenter to 'reproduce' a theoretically predicted regularity. Therefore, when we need to learn – what we cannot know in advance – if real actors actually behave as regularly as our theory claims, we have to carry out an experimental test on a system that is real rather than virtual, that is, a system whose behaviour is 'driven' by internal objective forces rather than by the theory to be tested, or, more exactly, by the experimenter armed with his theory and techniques.

The nature of *regularities* in the social world has intrigued old and new 'masters of sociological thought.' Anthony Giddens (1984, xix), a leading figure in contemporary theorizing of the first genus, equates regularities with 'generalizations,' thus agreeing in this respect with the positivist tradition that he criticizes for the neglect of human subjectivity and creativity.

> Some [generalizations] hold because actors themselves know them – in some guise – and apply them in the enactment of what they do … Other generalizations refer to circum-

> stances, or aspects of circumstances, of which agents are ignorant and which effectively 'act' on them … 'structural sociologists' tend to be interested in the generalizations in this second sense … But the first is just as fundamental to social science as the second.

Sociologists often explain regularities characterizing 'social practices' observed in certain typical situations by attributing to the people the *knowledge* of certain *rules.* Giddens believes that the knowledge of these rules prompts to the actors what they should do in these situations. He defines (1984, p. 21–22) 'rules of social life' as 'techniques or generalizable procedures applied in the enactment/reproduction of social practices.' Seen in this perspective, the soldiers' obedience to their commanders results from their knowledge of the rules that establish behavioural dependence between the occupants of inferior and superior positions in social systems of the kind called by Max Weber *Herrschaftsverband.*

Regularities of the first type consist in *enacting theories.* The 'laws' of such theories, even though they may be formulated by sociologists, do not essentially differ from 'social laws' of which the use by the actors makes the combinations of their actions predictable. However, if we use a 'theory' of which we know only that 'knowledgeable agents' accept it to explain the very fact that they 'enact' this theory, then we have to abandon even the humanistic (Weberian) conception of social *science*. Nevertheless, in many situations we must admit explanations of observed behavioural regularities in terms of 'reproducing' certain patterns (no matter whether regular behaviour was taught to the group or emerged as a result of a natural group process). It is debatable whether such an explanation can be accepted as the *only* way to account for regularities 'produced' in the lab with the use of 'theory-driven' experimental procedures.

Giddens' typology of regularities ('generalizations') has a counterpart in economics. It is the opposition between command economy and market economy. In a *market* economic system, agents freely negotiate exchange rates in transactions among one another. If the same agents are forced to act in an economic *imperatively coordinated association* (Dahrendorf's translation of Weber's *Herrschaftsverband*), they will 'reproduce' the exchange rates taken from the theory they are told to 'enact.' In a command system, the actors behave 'theoretically' for fear that they would be worse off if they did otherwise. In a market system, every actor can improve his own situation through interacting with others, which results in the formation of theoretical (equilibrium) prices. In both systems, the interaction process takes place in a structured environment. In the market case, 'freedom of choice' is institutionalized by means of definite rules concerning legal possession, production, and exchange of valued resources.

Smith preceded his paper (1982) on experimental microeconomics with the motto (from Louis Agassiz) 'Study nature, not books.' I studied both, which encouraged me to compare Giddens' meta-theorizing with the viewpoint on social regularities that grows out of the practice of experimental research. Our colleagues from the department who practiced 'social theory' or historical studies, seeing Jacek Szmatka and me doing experiments on abstract exchange systems, commented on our activities in two ways roughly (but not exactly) corresponding to Giddens' two

types of social regularities. Some, impressed by detailed instructions we read to our experimental subjects, blamed us of training them to behave theoretically and thus of misconstruing theory testing; others, who took notice of standardized conditions leaving little room for creative 'defining the situation,' criticized us for treating the subjects like rats, or as Giddens put it, as 'agents ignorant of the circumstances which effectively act on them.'

It is true that the subjects in any experimental social system are taught and induced to act in accordance with well-defined rules. Moreover, the experimenter's task is to ensure that the subjects will co-act upon a common *definition of the situation* that is given in the instructions the subjects must properly understand (and the experimenter must check if they did). On the other hand, in our experiments they were given enough freedom in decision making. Their behaviour could not be interpreted as 'enacting' a theory translated into a script.

In network exchange experiments, the enforcement of a social regularity of the first type is not an end in itself. It is needed only to set the stage for the interaction process that is expected to display a regularity of the second type. Preparing the experimental setting includes establishing definite *structural constraints* and opportunities for negotiating transactions, namely, group members are *instructed* as to with whom they are permitted to initiate and conclude transactions. A fixed set of communication channels can be easily enforced on a group with the use of a *computer network*. Since the system's structure alone cannot force 'agents' to negotiate and conclude transactions, the experimenter-theorist, in order to set the 'interaction machine' in motion, must not only induce subjects to comply with the *rules* but ensure that their actions are guided by an appropriate *motivation.* This is done by having subjects read statements, for instance, like this (Willer 1987, p. 121): 'Your goal should be to get the best score that you can for yourself through arranging the transactions most favourable to you.' Inducing the required motivation allows the experimenter to work without assuming that self-interest is a *natural* human disposition. However, it may appear unfeasible to make subjects behave selfishly as it would require that they suspend the 'natural' or learned inclinations they bring into the laboratory from the external world. Indeed, one of the first experiments on network exchange (Cook, Emerson 1978) confirmed the significance of a preference for equal division of rewards.

If there are reasons to believe that *structural* and *motivational* scope conditions of the theory being tested are met, then the observed outcomes of the group process can be compared with theoretically predicted outcomes. One may ask if the pattern that is expected to emerge from joint action will actually arise if it is known to the actors before the experiment. If the subjects come to know the predicted negotiation outcomes, they may attempt to affect the result of the experiment. How to interpret the case in which the order found to be produced by 'naive experimental subjects' does not occur when the experiment is repeated with 'knowledgeable agents'? Should we conclude that social regularities of the second type lack the 'necessity' that is attributed to the 'laws of nature'? Economists believe that the 'laws of market economy' cannot be changed by those who do not approve of some of their consequences (e.g., highly uneven income distribution). What we know from

20th century history is that people who do not like the 'natural economic order' can knock it down by destroying the structural and/or motivational scope conditions under which economic laws operate. In experiments, the most likely cause of the 'knowledge effect' is not knowledge itself but the appearance of motives that suppress or interfere with those assumed in scope conditions, e.g., the subjects behave so as to show their superiority over 'ignorant rats.'

The truth, rather obvious for sociologists (Willer 1985), about social-structural scope conditions of the laws of market economy, has long been overlooked by most economists. 'Incredibly, it is only in the 20 of these 200 years [of the history of economics] – as Vernon Smith noted (Smith 1982, p. 952) – that we have seriously awakened to the hypothesis that property right institutions might be important to the functioning of the pricing system!' Smith demonstrated that not only property right institutions do matter. His 'experimental handling' of auctions in simple markets has undermined the widespread conviction that economics, like astronomy or meteorology, has to rely on observation of real-world processes. He wrote (Smith 1999, p. 197): 'My view is that the reason economics was believed to be a nonexperimental science was simply that almost no one tried or cared.' The Nobel prize for Vernon Smith (2002, with Daniel Kahneman) gave moral support also to a group of sociologists, who were unaware of experimenting that was going on in economics, but tried and cared to do laboratory experiments on exchange, guided by sociological theories of the third genus.

Network exchange experimental paradigm. Experiments done by the Chair of Research on Group Processes at the Jagiellonian University (1989–2001)

Experimental research (unknown to economists) on exchange systems with *network structure* was initiated in *sociology* at the end of the 1970s by Richard Emerson and his collaborators (Cook, Emerson 1978) to be subsequently directed to a new path by the Elementary Theory group (Willer 1987; Markovsky, Willer, Patton 1988; Szmatka 1997). Classical *economics* has shown little interest in the study of socioeconomic systems endowed with *social constraints* (in particular *network constraints*) that *forbid* some actors from concluding some *physically possible* and mutually beneficial transactions in contrast to *free market systems* where every two *owners* of valued resources are allowed to transfer them between each other on the terms both parties *voluntarily* accept.

In any *exchange system*, a *legal* change in the allocation of control over valued resources can take place only through voluntary give-and-take actions of the actors. The *private property rule* means that each actor has *exclusive* control over some resource. The *reciprocity rule* means in turn that each party of a voluntary agreement has to give up its resource to the other party as soon as the latter has fulfilled its part of the contract. These rules constitute the fixed *institutional* ground for the functioning of any *exchange system*.

Both free market systems and *network exchange systems* can also be endowed with explicitly stated *negotiation rules*, or the rules that establish legal ways of

negotiating and concluding transactions. Smith (1982) treated these rules as a social-structural factor subject to experimental manipulation. I discovered independently the theoretical importance of negotiation rules when I carried out my replication (Sozański 1993) of an experiment done by David Willer (1987, Chapter 6). My aim was to test Willer's predictions pertaining to the behaviour of a social system in which a 'manager' (in abstract language, an actor who occupies the central position *connected* to a set of peripheral positions which are not connected between each other) negotiates with candidates for vacant jobs the financial terms of their employment. In a *hierarchical centralized network exchange system* there are as many candidates as vacancies. A *mobile hierarchy* is established by enforcing the rule that the pay guaranteed by the manager to the next applicant must be lower than that awarded to the first applicant already hired. Such a structural constraint forces peripheral actors to compete between each other for being first to reach agreement with the central actor. The competition results in accepting a pretty low pay by the winner of the *auction* to the benefit of the manager. Below I quote the conclusion from the English summary of my paper (Sozański 1993, p. 308).

> The power advantage of the 'centre' over the 'peripherals' has been observed, however, to a lesser degree than in the original experiment … the difference can be explained in terms of different modes of negotiating. The rules (imposed by the experimenter or adopted spontaneously by the subjects) which organize the negotiation process can enhance or weaken the competition among peripheral actors.

In my experiment, the 'manager' had to hear initial demands from all 'applicants' and propose himself the pay for the next person to be hired. Technically, every negotiation round began from a 'complete bidding' in which all 7 subjects (6 in peripheral positions and one in the central position) were called (by the computer program) one by one in a random order to present their proposals. Under such a *negotiation protocol* (Sozański 1993, p. 249–250), there occurred 'class solidarity' among the peripherals, counterbalancing within-class competition to some extent. While in Willer's experiment the 'applicants' went on outbidding one another, in my experiment they often demanded the same pay and accepted the uncertainty about who of them would be hired on the terms they all tried to defend.

Napoleon Bonaparte used to say: 'For war we need three things: money, money and more money.' Although many scientists repeat the same with 'war' replaced by 'research,' I always tell to my students that what we need first of all to do research are ideas, good ideas, and better ideas. In 1990–1991, when I designed and carried out my replication of Willer's experiment, the research unit founded by Jacek Szmatka (see Appendix) had just one 8-bit computer with built-in interpreter of Basic programming language. At that time Jacek was running a series of experiments designed as replications of Willer's experiments (Szmatka 1997). When I was watching my colleague creating in his lab what he called *experimental replicas* of exchange networks, I had not yet been fully acquainted even with Willer's papers (Willer 1981a,b) from which his Elementary Theory has grown. Before I began to read his book (1987), I studied the 'nature' of *empirical* microsocial systems I could

see in action in a room which was turned by Jacek into a laboratory then equipped solely with cardboard barriers needed to restrict communication between occupants of network positions.

In a *face-to-face interaction setting*, network positions do not need to have physical counterparts. The actors do not even have to realize that they 'occupy positions in a system.' For example, if we want to construct a system with *network structure* having the form of the *graph* B_1—A—B_2, and with actors s_1, s_2, s_3 occupying A, B_1, B_2, the experimental instructions may reduce to telling actors s_1 and s_3 that they are allowed to communicate only with s_2, while s_2 can communicate with s_1 and s_3. A barrier can be placed between s_1 and s_3 to make sure that they will comply with the ban on communicating with each other.

Having learned from observation the first network exchange experimental paradigm[8] that Jacek Szmatka had come to know in the University of South Carolina laboratory, I noticed that the first technical problem to be solved was recording the groupwide (however running in dyads) negotiation process going on in each round.[9] With one computer at hand, I managed to solve the problem by writing a Basic program. However, the solution involved introducing definite negotiation rules which in some respect could be considered inconvenient, namely, my program did not enable the actors to freely choose the time for making their offers to potential partners or responding to others' offers. Consequently, competition among peripheral actors for the mere opportunity to present their offers to the central actor was eliminated, which on the other hand resulted in revealing the effect from negotiation protocol.

The research grant won by Jacek Szmatka in 1994 from the Polish counterpart of US NSF recalled the truth that money does matter in doing science too. We were at long last able to equip our lab with a *local computer network* made up of 7 personal computers (the server and 6 workstations). John Skvoretz, then working with David Willer at the University of South Carolina, made available to us his program Exnet (written in Quick Basic 4.5 working under the Novell Netware operating system); he was also kind to help us install it in our laboratory, which in 1995 became ready for running technically advanced network experiments.

Our plan was to examine *all* 8 smallest non-isomorphic *exchange networks with one-exchange rule* (*one-exchange networks* for short), 2 networks with 3 positions, and 6 with 4 positions. The *one-exchange rule* means that every actor is allowed to

[8] There are many experimental paradigms and theories which have been proposed after Emerson and Cook's published their seminal paper (1978). A recent comprehensive account can be found in Molm's (2014) chapter in the 2nd edition *Laboratory Experiments in the Social Sciences.*

[9] To generate data for a single network with *n* positions you need at least one set of *n* subjects. Each set of subjects can be used in multiple rounds in which the assignment of actors to positions remains fixed. Such a sequence of rounds is called a *period*. A *session* with one group may consist of a few periods with a different actor-position assignment in each. The *rotation* technique used in the University of South Carolina laboratory allows every actor to occupy all positions in one network throughout the session. Rotation can be criticized for *systematic* use of too few out of many possible assignments of actors to positions (e.g., in a 4-point network only 4 out of 4!=24 are used). A random selection of assignments for use in one session seems to me a better method for controlling *subject variables.*

conclude no more than *one* transaction per round. This rule implies that the negotiation process in the network with the transaction opportunity graph of the form B_1—A_1—A_2—B_2 may end up with 5 outcomes: (1) no transaction; (2) a transaction between A_1 and B_1; (3) a transaction between A_2 and B_2; (4) two transactions: one between A_1 and B_1 and the other between A_2 and B_2; (5) a transaction between A_1 and A_2.

I refer to this one-exchange network and the underlying graph as the 4-Chain.[10] I will take this network, which has so far been studied most frequently of all, to describe constitutive components of the network exchange experimental paradigm used in our laboratory.

1. The first component is an operational definition of a *transaction*. In network exchange experiments, a transaction is not what the word 'exchange' suggests (a mutually agreed-on bilateral flow of valued resources). Instead, a transaction is understood as a bilateral agreement on a division of a pool of M points (usually M=24).
2. The *transaction opportunity graph* is the second component of the paradigm. The *points* of this graph are called *network positions.* Two actors *s* and *s*' who are placed in positions *P* and *P*' in a given *negotiation round* are permitted to conclude a transaction if their positions are *connected* in the network, that is, *P*—*P*' is a line of the transaction opportunity graph.
3. The third component specifies the range of transaction configurations which may occur in one round. The first assumption is that any pair of actors is allowed to conclude no more than one transaction per round. With this assumption made, an *exchange regime*[11] is defined as a collection of transaction sets. Any *transaction set* consists of lines which can be the locus of transactions within any round. A round ends if all transactions in a *maximal transaction set* have been concluded. In the 4-Chain network, one-exchange rule generates the exchange regime with 5 transactions sets of which two are maximal: {A_1—A_2} and {B_1—A_1, A_2—B_2}. Non-maximal transaction sets, such as {B_1—A_1}, may also appear in a round, which happens when the time allowed for negotiations has expired.
4. The *negotiation protocol* specifies the range of *actions* available to the actors and defines a sequence of actions that must take place to be automatically followed by a transaction. The protocol implemented in the version of Exnet we used in our laboratory admits of three types of actions. The first of them consists in sending an *offer* (a proposed division of the pool of profit points) by an actor to one of his neighbours in the transaction opportunity graph. If the recipient of the offer *accepts* it, the sender may *confirm* his offer. If he does it, the sequence of three actions initiated by him is followed by the transaction. Due to this condition all offers are tentative. The actor whose offer has been

[10] The name 4-Line is used more often in the literature but 4-Chain is a better name because it avoids confusion with the term *line* (edge, link) commonly used in *graph theory* to denote a pair of connected points (vertices, nodes). Notice that the 4-Chain graph has 3 lines.

[11] The term and the germ of the idea I elaborated in my paper (Sozański 2006, p. 398–399) comes from Friedkin (1992).

accepted may confirm it, but he may well send a new offer to the same partner or stop communication with him and start negotiations with another potential partner.[12]

5. The paradigm must also contain assumptions on the actors' *motivation* or *goal-orientation*. Each actor is supposed to negotiate so as to maximize his own profit only; 'tactical' decisions on how to pursue this goal are left to himself.
6. All actors are assumed to have *full information* about the system and its components. They are also given an opportunity to watch the course of negotiations throughout the session in which they participate.

The experimental paradigm I have just described suffices in itself to predict that only maximal transaction sets should be observed in any negotiation round. The full information assumption (6) implies that if the actors in positions A_1 and B_1 have agreed on a pool split, their transaction becomes known to the actors in positions A_2 and B_2. Their awareness of this fact motivates them to conclude a transaction on any terms. Otherwise none of them would gain any points, which is in contradiction with the assumption of individual rationality (5). As a consequence, such a round must end with the maximal transaction set $\{B_1—A_1, A_2—B_2\}$.

However, the paradigm does not by itself imply any *concrete theory* which would generate specific *predictions* as to how the pool shall be divided between the partners in each transaction within a maximal transaction set. One can only require that any plausible theory must be *structural*. Any *structural theory*, applied to the 4-Chain network, predicts the same pool split in network lines $A_1—B_1$ and $A_2—B_2$; the occupants of positions A_1 and A_2 (B_1 and B_2)[13] are expected to earn on average the same number of profit points.

If we define a *structural parameter*[14] suitable for measuring the *bargaining power* of a position in a one-exchange network, what we need in order to construct a *precise* theory predicting the negotiation process outcome in a round is a *formula* that will allow us to *calculate* the theoretical payoffs of the partners from the values of the chosen power parameter. When the first power parameter (*Graph-theoretic Power Index*, GPI) devised by Markovsky, Willer, and Patton (1988) appeared inadequate for *weak power networks*, the quest for new power parameters and new theoretical formulas began (Lovaglia et al. 1995) to continue until recently.

---

[12] I proposed (Sozański 1997, p. 314–316) an alternative protocol under which an actor addresses the same offer to all his neighbours. He may also choose one of them as the current target of his proposal. The transaction between two actors is assumed to follow automatically as soon as they choose each other and make complementary offers (agree on a split of the pool). The sequence composed of two last offers and two last partner choices can contain the four actions in any order.

[13] Positions labelled with the same letter are *automorphically equivalent*, that is, one of them is the image of the other through an *automorphism* of the transaction opportunity graph. The one-to-one mapping F of the set of 4 positions $\{A_1, A_2, B_1, B_2\}$ onto itself such that $F(A_1) = A_2$, $F(A_2) = A_1$, $F(B_1) = B_2$, $F(B_2) = B_1$ and the identity mapping are the only automorphisms of the 4-Chain graph.

[14] A *structural parameter* of a point in a graph is defined by the condition of assuming the same value for any two automorphically equivalent positions.

At this point I must stop discussing further theoretical developments and return to the story on experiments carried out in the 1990s by the crew of the Chair of Research on Group Processes at the Jagiellonian University. The series of experiments we started in 1995 was aimed at testing predictions derived from various specific theories proposed until then, their unity being based on a common *scope* and on the use of *structural variables.* Instead of focusing on the study of larger networks in which our American colleagues found certain peculiarities, we decided to concentrate our efforts on *systematic* examination of all small-size one-exchange networks in order to assess prediction *accuracy* of each *precise* theory and identify among them the one which provides best fit to the data generated for the networks for which any *general* theory of network exchange should do particularly well.

We completed our experiments by the end of 1996, yet the results, which have so far been presented only at two conferences[15], are still waiting for being published in a research paper or book chapter. Jacek found for himself another research area (Szmatka et al. 1998), *conflict networks*, also suitable for experimental treatment.[16] The cause of my long-lasting neglect was that my enthusiasm for doing *empirical* theory and research weakened a lot when I plunged again into solving *mathematical* problems (Sozański 1997, 2006). But every scientist who has once come to know the taste of experimenting will long for a return to this exciting activity. For me a return to laboratory work will no longer be possible but at least I can enjoy discussing *methodological* issues and the intricacies of the *technology* of experimenting with my colleagues who reveal equally strong commitment to experimental social science.

Conclusion and an introduction to the collection of papers that follow

Experimental social science was born more than a hundred years ago. Today it is a well-established way of doing *theory and research* in sociology and related disciplines. The first upsurge of interest in experimentation, which took place in the 1950s, yielded classical studies of *group dynamics* (Cartwright, Zander 1960). Those studies became widely known due to their coverage in social psychology handbooks (Collins, Raven 1969). A new wave of *theory-driven experimenting* and *experimentally driven theorizing* came in the 1980s as a consequence of successful attempts to construct sociological *theories of the third genus* (see earlier in this paper). This way of doing theory, which for a long time has not been recognized as a serious challenge to the first genus theorizing, is now considered legitimate as evidenced by the entries in George Ritzer's *Encyclopedia of Social Theory* (2005) devoted to the *Elementary Theory* and its authors (David Willer, Barry Markovsky).

In the 1970s, *social theory* underwent a change described by the French saying *le roi est mort, vive le roi*: Anthony Giddens, the author of *Central Problems in*

[15] The ASA Annual Meetings, San Francisco, August 1998 (Regular Session: Group Processes–Theory and Experiment on Power and Exchange) and Fourth International Conference on Theory and Research in Group Processes and Social Psychology, Cracow, June 2004. The latter conference, dedicated to the memory of Professor Jacek Szmatka, was co-organized by ISA RC #42 (Social Psychology).

[16] Experiments on conflict networks were done in our lab by Joanna Heidtman, Ph.D.

Social Theory (1979) and of another treatise (1984) I quoted earlier in this paper, became the successor of Talcott Parsons as 'king' of this *genre* of theorizing within the Anglo-Saxon world. I did not care too much about what excited at that time many sociologists in Poland and abroad. In that decade for me the most important book, the reading of which prepared me for joining Jacek Szmatka's research team in 1990 (see Appendix), was volume 2 of *Sociological Theories in Progress* (1972), a collection of articles edited by Joseph Berger, Morris Zelditch, Jr. and Bo Anderson. Another book (Berger et al. 1977), published next, was my first source of information about *Expectation States Theory* (EST)[17]. This theory was later counted by Jacek Szmatka (Szmatka, Sozański 1994, p. 229) along with the Elementary Theory (his favourite example) among the few theories epitomizing the third way of theorizing. In the 1970s – it was the time when I began my scientific activity – my interest in EST focused on *formalizing* this theory with the use of *signed graphs*, a special area within *graph theory.*[18] Later it became clear to me that the core of EST is a procedure (Berger refers to it as the 'Standardized Experimental Situation') which is used in laboratory experiments to endow a dyad with an artificial *status structure.*

Joseph Berger, Bernard P. Cohen, and Morris Zelditch, Jr. are credited with making EST a *theoretical research program* (Berger 1974) which 'consists of a set of interrelated theories, bodies of relevant research concerned with testing these theories, and bodies of research that use these theories in social applications.' (Berger 2014, p. 269). Berger, now Professor Emeritus of Sociology at Stanford University, has been the leader and currently is the senior member of the circle of scholars working under the banner of 'theory and research on *group processes*.'

In Expectation States Theory, the distinction between *low* and *high* status is defined in terms of unequal levels of *competence* (in performing a task or a special kind of tasks) that group members attribute to each other. Such a meaning given to the concept of *status* departs from the traditional Weberian understanding of status structure.[19] The same can be said about *power* and other old concepts, which

[17] To get familiar with the basics, see Martha Foschi's (2000) excellent encyclopaedic article.

[18] The English title of my Ph.D. thesis (in Polish, 1982) is: 'Structural Balance Model. Theory of Signed Graphs and It's Applications in the Social Sciences.'

[19] According to the traditional European approach to status (prestige) structure, unequal level of competence in a given area of human activities (e.g. , in doing science) need not be the main reason for unequal distribution of respect or 'status honour' (*ständische Ehre*, Weber's term). The differential *evaluation* of various kinds of tasks may also generate a hierarchy with definite consequences (e.g., interpersonal influence) for social interaction. For instance, white-collar workers enjoy a higher status than blue-collar workers because what the former do to earn a living is believed to be a 'nobler' kind of job. In the Middle Ages, knights, or those who were attributed a good command of the sword, were higher in the status hierarchy than peasants expected to be competent in operating a plough. In the academic world, both bases of inequality are found. The competence-based status structure, which exists within each discipline, in Poland has the form of a two-grade system with 'low' and high 'status' marked by 'dr' (roughly the counterpart of Ph.D.) or 'dr hab.' placed before a scholar's name. Another status hierarchy, which depends on which branch of science you deal with, is less clear but certainly expertise in *social science* is not as highly evaluated as doctorate or habilitation in *legal science*.

re-appeared in a new shape within theories rooted in experimental paradigms. Giddens is right to claim (1984, p. 283) that 'There is no more elemental concept than that of power … Power is one of several primary concepts of social science, all clustered around the relations of action and structure.' Even though Willer invokes Max Weber many times in his book (1987), again there is at best a loose connection between the Weberian understanding of power and the meaning of this concept within the theoretical research program (Willer, Markovsky 1993) that had its origin in Willer's analysis (Willer 1981a) of elementary dyadic social relations.

However weak the ties may be between *experimental* social science and sociology at large, it is not my intention to call for soft divorce with the sociological tradition. Wherever sociology is taught, its two traditional pillars remain the knowledge of major 'masters of sociological thought' and that of 'the basics of social research.' As regards this second pillar, it should be noted that experiments done in the 1980s have been noticed and appreciated outside the Group Processes circle. The experiment designed by Martha Foschi and her two younger collaborators (see Foschi, Warriner, Hart 1985) was used by Earl Babbie in the 5th (1989) and all later editions of *The Practice of Social Research* to illustrate experimentation in sociology.[20]

The Group Processes circle[21], which came into being in the 1980s, has since then remained incessantly active until today, publishing every year since 1984 a successive volume of the book series *Advances in Group Processes*.

> Early experimental research documented how the ideas and norms of one generation feed into another generation even when the members of the preceding generation are no longer present.[22] This lasting of our intellectual ancestors is clearly demonstrated in

[20] In that experiment, Martha Foschi used for the first time standards as an independent variable, a *standard* being defined as the lowest level of performance (e.g. a score in psychological test) that must be attained in order that a person's performance could be recognized as 'satisfactory' or taken as a proof of competence. Foschi later published several articles about standards, in particular, the practice of *double standards*. One of her papers (*Podwójne standardy oceny konferencji: najnowsze wyniki i nowe kierunki*, translated by Z. Karpiński) appeared in Polish in Heidtman J., Wysieńska K. (eds.). (2013). *Procesy grupowe. Perspektywa socjologiczna.* Warszawa: Wydawnictwo Naukowe Scholar. This volume edited by Heidtman and Wysieńska also contains articles (Polish translations) by other contributors to this issue of *Studia Sociologica*.

[21] Most of the members of the Group Processes circle have been American scholars or their collaborators from other countries: Canada (Martha Foschi), Japan (Toshio Yamagishi), Turkey (Hamit Fişek), and last but not least, Poland (Jacek Szmatka and me, Jacek's former students: Joanna Heidtman, Zbigniew Karpiński, and Kinga Wysieńska-Di Carlo). Let me add in this connection that in 1994–2014 four internationally known members of this circle (Karen Cook, Guillermina Jasso, Edward Lawler, and Cecilia Ridgeway) were presidents of ISA Research Committee #42 (Social Psychology).

[22] Among those who 'are no longer present' there is one scholar whose name should be recalled here. I mean Bernard P. Cohen, the author of *Developing Sociological Knowledge* (1989), the book which heavily influenced methodological views of many scientists doing theory and research in group processes. Cohen's earlier book (*Conflict and Conformity*, 1963) about an application of *Markov chains* (a probability model) to the data from Asch's experiment was one of my first readings in *mathematical sociology*.

the Group Process meetings[23] by the continuing collective commitment to theoretical development, methodological precision, and the integrity of the scientific process.

The quoted passage closes the Editors' Preface to the 2nd revised edition of *Laboratory Experiments in the Social Sciences* (1st ed. 2007). My recent appointment as head of the Chair of Methodology of Social Research, a unit within the Institute of Philosophy and Sociology at the Pedagogical University of Cracow, my intention to enrich with an international event the celebration of the 70th anniversary of founding the school, which became my workplace 10 years ago, and lastly but not least importantly, the appearance in the same year 2014 of the 2nd edition of the aforementioned book – all that inspired me to organize the *International Symposium on Experimental Research in the Social Sciences* to be held in Cracow next year.

To attract participants I sent a call for papers to ISA Research Committees #42 and #45 (Rational Choice) of which I am a member, and I addressed specialists in methodology of experimental research. My plan appeared workable in part. Martha Foschi and Murray Webster, Jr. accepted my invitation to deliver keynote lectures. Murray Webster, the 1st editor (with Jane Sell as 2nd editor) of *Laboratory Experiments*, is a key figure in experimental social science. In 2015 he received the Cooley-Mead award.[24]

The time chosen for the symposium, 12th and 13th June, 2015 might have been inconvenient for those potential participants who were more interested in attending in June the annual Sunbelt Conference organized that year in the UK by the International Network for Social Network Analysis. As a consequence, among 'selected topics in experimental social science' you will not find the one which has always been closest to my research interests and experience, namely *experiments on network interaction systems*, or the topic which might have been treated most competently by Professor John Skvoretz, President of INSNA, 2010–2016, once a long term collaborator of Jacek Szmatka's Chair of Research on Group Processes at the Jagiellonian University.

My editor's hard work that followed the reviewing/revising phase of preparing this special issue of *Studia Sociologica* has ended with accepting 8 articles for publication. All of them except one (Szymon Czarnik's paper) are extended or re-edited versions of the papers presented at the June 2015 symposium. Since the abstracts written by the authors themselves present an overall description of their respective contributions, I will limit to a minimum my introductory comments, highlighting

[23] Since 1988 conferences on theory and research in group processes have been organized each year as an event accompanying the Annual Meetings of the American Sociological Association.

[24] The Cooley-Mead selection committee noted that 'in his distinguished career of nearly 50 years, Murray has been a leader in developing expectation states theory, identifying the processes by which status characteristics … shape and organize social interaction, and promoting rigorous, state-of-the art experimental scholarship.' The Cooley-Mead award was established in 1978 by the Social Psychology Section at the American Sociological Association. Webster joined the list of winners containing names known to every sociologist (e.g. Goffman, Homans, Bales, Merton), as well as those of several members of the Group Processes circle (Joseph Berger, Morris Zelditch, Jr., Edward Lawler, Bernard P. Cohen, Karen Cook, Cecilia L. Ridgeway, and Linda Molm).

solely some topics or questions that seem to me interesting; the readers need not of course share the commentator's view in this matter.

Martha Foschi's article (*Experimental Contributions to Sociological Immigration-Research*) reports on the results of an extensive literature search for papers in which immigration topics were investigated experimentally. She selected nine studies 'to illustrate the variety of factors and designs that have been used in this area.'[25] These studies, each interesting in itself, are compared with respect to the type of manipulated independent variables and the type of dependent variables (the latter 'consist of either written responses or actual behaviours concerning immigrants'). My observation is that most experimental studies of attitudes toward immigrants (e.g., presented as 'job applicants') are 'empirically driven,' even if manipulating independent variable consists in introducing into a vignette factorial design a variable *abstractly* defined by mere distinction between 'immigrants' and 'non-immigrants.' Some general topics in the methodology of experimentation ('artificiality' of an experimental setting, 'generalizability' of experimental findings) are examined at the end of the article. The author's conclusions clarify some matters that are often considered controversial.

The article by Murray Webster, Jr. with Jane Sell as 2nd author (*The Present Status and Future Prospects of Experiments in the Social Sciences*) begins with a presentation of the basics of the experimental method. In particular, the authors highlight the importance of what they call *strong instantiation*. Instantiation 'means creating a concrete instance of the abstract concepts in a theory or in a hypothesis, and it should be done as clearly and as powerfully as possible. Subtlety is out of place in experimental design … Weak instantiation of independent variables risks producing high variance within conditions and small overall difference across conditions.' I agree with the authors that the problem of reducing 'variance within conditions' is crucial for the success of the experimental testing of a hypothesis no matter whether the latter is derived from an *abstract* theory or comes from an analysis of a concrete experimental setting. In the second part of their paper, Webster and Sell deal with more technical matters, for instance, they 'trace developments in a standardized design that has been widely used to study status and expectation state processes' and present 'some new designs [that] are being developed to study interrelations of vocal accommodation and group position.'

In their methodological article (*Assessing Epistemic Claims by Experimental Evidence*), Robert K. Shelly and Ann C. Shelly analyse 'three ways in which epistemic claims may be advanced and assessed: triangulation, multitrait-multimethod, and meta-analysis'. 'Triangulation and multitrait-multimethod provide strong answers to the question of how do we know what we know by specifying the links between theory, data, and measures. Meta-analysis is not quite as robust on this issue…'

Two short papers that follow deal with the role of socio-cultural context in experimenting. Jane Sell and Murray Webster conclude their second contribution (*The Importance of Cross-Cultural Experiments for the Social Sciences*) – it can be regarded

[25] Quotation marks that appear here and further in this section delimit pieces of text taken from commented papers.

as a supplement to the first one – with a remark that cross-cultural replications of an experiment need to be done to demonstrate that 'general principles apply even in very different contexts and initial conditions' but also to examine how cultural specific *initial conditions* affect general laws.

In her paper (*Can Socio-Cultural Context Affect Experimental Results? The Case of the Zimbardo Prison Experiment Repeated in Poland by Artur Żmijewski*), Iza Desperak describes a repetition of the Stanford Prison Experiment[26] in Poland. The makeshift prison in Stanford had to be closed after 6 days because the 'guards' abused the power given to them by the experimenter. By contrast, in Poland an intervention of the artist playing the role of 'prison governor' was not necessary: the 'guards' and 'prisoners' resolved together to stop the performance they had been induced by him to take part in. Why? A tentative answer is given by the author in her paper.

Marcel Kotkowski is the author of the last of eight articles (*Psychophysiological Techniques for Measuring Emotion in Social Science*). His paper can serve as a useful source for any sociologist who would like to gain elementary knowledge of various techniques for measuring emotions. 'A note on each technique points out the dimension of emotion (valence or arousal) that is measured with a given technique, and informs on its previous use in sociology, as well as its major advantages and disadvantages.'

Two contributions that remain to be presented here are good examples of doing theory-informed experimental social science. In their paper (*Modelling Social Situations: Trust and Cooperation Among Strangers of Unequal Status*) Zbigniew Karpiński and Kinga Wysieńska-Di Carlo[27] report on the results of the two experiments they designed to test hypotheses that relate frequency of cooperation in certain social situations (modelled by two-person Prisoner's Dilemma game) to the configuration of partners' statuses (Low-Low, Low-High, High-Low, High-High). To derive their predictions, the authors invoke status characteristics theory as well as collective action theories, making an attempt to *integrate* 'theories originating in distinct general research programs.' The article also has didactic value: it is instructive to see how an analysis of the results of the first experiment leads the authors to design the next one.

Szymon Czarnik's article (*Reading Minds of Experimental Subjects. Insights from Pre- and Post-Experimental Surveys in a Redistribution Game Experiment*) is also instructive as it demonstrates how large can be the range of social phenomena amenable to laboratory experimentation. For the purpose of his experiment Czarnik placed each pair of subjects in a *socioeconomic system* in which: (1) The actors work and earn money proportionally to the amount of work done; (2) Their incomes are subject to taxation with the rate of linear tax depending on the actors' decisions (they are asked to reveal their preferred tax rates) and on a 'democratic' rule (the rate to be implemented in the system is computed as the average of the

[26] Willer and Walker (2007, p. 100) comment on Zimbardo's experiment in the following words: '…we are unable to identify any theory or theoretical model under test. Consequently, it is neither a method-of-difference nor a theory-driven experiment.'

[27] The co-authors declare having contributed equally to their product.

rates proposed by the actors); (3) A fixed fraction of the total tax collected from the 2-person group is lost and the rest is divided evenly between two actors to the effect that one of them benefits from the *redistribution* while the other loses some part of his income earned before taxation and redistribution; (4) At the last stage the actors are given an opportunity to make *voluntary* money transfers to each other.

Czarnik had already published a report on this experiment in a 'hard science' journal (Czarnik 2006). The paper that I received from him after the June 2015 symposium begins from recalling the redistribution mechanism and analysing the two-step game that is obtained by having the subjects make decisions in phases (2) and (4) of the system's functioning. In the current article, mathematical considerations are only a prelude to an examination of the *subjective* dimension of the collective behaviour within such as system, including the subjects' declared motives and those attributed to others ('We find experimental subjects to be predominantly negative in their assessment of intentions behind their partners' decisions …').

I conclude this last section of my introductory article by stating 5 postulates or principles guiding theory and research on group processes.[28]

1. Both *natural sciences* and *social sciences* are empirical sciences. Although they differ in the nature of objects to be studied, general methodological norms that apply to all *empirical sciences* remain valid for social science.
2. The aim of *basic social science* is to study *abstract* social systems (e.g., network interaction systems) rather than historical concrete objects (such as 'Polish society AD 2016').
3. *Theories* that are to describe regularities that characterize the functioning of these systems should be *parsimonious* in making assumptions on the nature of human actors (assumptions concerning people's motivation or their knowledge of the conditions of action). Where psychologists, whose task is to deal with human subjectivity, need to invent complex models of an 'individual in action,' social scientists should instead simplify, focusing on building more or less complicated models of 'social systems in action' in which the *form of structure* of a social action system (such as an exchange network) is taken as the central factor in explaining the system's behaviour.
4. Theory and research should begin with the study of *elementary* social phenomena or processes (power, status, influence, cooperation, etc.).
5. *Laboratory experiment* is the best method for testing theories that deal with these phenomena.

[28] My current formulation of these principles repeats the ideas already expressed in Polish in my obituary (Sozański 2001, p. 8–9) devoted to Jacek Szmatka.

Appendix

Jacek Szmatka (1950–2001)

The collection of papers *Selected Topics in Experimental Social Science* appears (as part of the current issue of *Studia Sociologica*) 15 years after Professor Jacek Szmatka passed away. This appendix is to recall to foreign readers of this paper my late colleague who was a keen advocate of using experimental method in sociology. Three days after his decease I emailed a letter to his overseas friends and professional associates. In 2004, when my presence in the Internet began, I placed that letter on my personal website (http://www.cyf-kr.edu.pl/~ussozans/) supplemented with the list of Jacek's publications in English. Both are now placed in this special issue of the journal published by Pedagogical University of Cracow to document in print an episode that the history of Polish sociology owes to Jacek Szmatka.

Dear Colleagues

The sad duty has fallen upon me to inform you that Jacek Szmatka passed away October 20, 2001, in Athens, Ohio where he was staying this semester as visiting professor.

We met in 1968 when we began studying sociology at the Jagiellonian University in Cracow. Jacek came to our city from Rzeszów (then a county town east of Cracow) where he was born in 1950. Among the members of our sociology class he was the first to receive his M.A. (1972) and Ph.D. (1975), both from the Jagiellonian University where he worked continuously from 1972.

We met for the second time as assistant professors affiliated with the Chair of Theoretical Sociology headed by Professor Piotr Sztompka. Jacek was then interested, first of all, in general methodology and social theory as documented by the titles of his Ph.D. thesis ('Theoretical Reduction in Sociology'), and that of his first book ('Individual and Society: On the Dependence of Individual Phenomena on Social Phenomena') which he published in 1980 as his 'habilitation dissertation,' a requisite in Poland in order to be appointed to the position of associate professor.

Szmatka's collaboration with American sociologists dates back to the early 1980s. He translated into Polish many classic papers on small groups as well as Jonathan Turner's *The Structure of Sociological Theory* (Polish edition, 1985). Jacek also wanted to make his native country's sociology known abroad. He was invited to the board of the International Advisory Editors of Encyclopedia of Sociology, edited by Borgatta and Borgatta (first edition, New York: Macmillan, 1990-1992) for which he wrote the entry on 'Polish sociology.' Though he conceived of theoretical sociology as a science which should deal with abstract social structures rather than historical societies, he often taught courses on the problems of Poland and Eastern Europe and co-edited (with Z. Mach and J. Mucha) a volume on these topics (*Eastern European Societies at the Threshold of Change*. New York 1993).

Jacek came to the US for the first time in 1983. Since then he was a frequent guest to America where he felt at home nearly as much as in Poland. He worked as a visiting professor at many American universities (University of Kansas, State University of New York, Stanford University, University of Washington, University of South Carolina, University of Iowa) and regularly attended Annual Meetings of the American Sociological

Association (he had been an ASA member since 1991). The circle of scholars which used to meet separately at 'group processes conferences' accompanying the ASA Meetings became his 'reference group'; they helped him reorient his scientific interests from 'grand theory' to 'hard social science.' He made many friends among the members of this group who could certainly add their own memories to this informal obituary.

University of South Carolina was the place in America that Jacek visited most frequently. There he came to know the Elementary Theory (ET) and established close ties with David Willer and his colleagues. The long-term cooperation of Dave and Jacek, which yielded several co-authored papers, began in 1989 with a common research project aimed at testing the universality of ET.

My third encounter with Jacek which gave rise to our cooperation throughout the following decade took place just at the time when Jacek got fascinated with the Elementary Theory. In Spring 1990 somewhat unexpectedly I saw my colleague, whom I had known earlier as a 'grand theorist,' doing 'cross-national experiments' in his office now turned into a laboratory.

The historic year 1989 in which the communist regime fell in our country was equally crucial in his career. Jacek had by then published his second book ('Small Social Structures: Introduction to Structural Microsociology') which established his reputation in Poland as an outstanding specialist in small group theory and research. That same year he was appointed head of the Microsociological Laboratory which he had created in the Department of Sociology at our university. Jacek's achievements had not gone unnoticed. In 1992 he received the title of professor which granted him tenure. In 1995, his research unit (renamed the Chair of Research on Group Processes in 1996) was equipped with a local computer network which, together with software received from South Carolina, enabled him and his team to actively participate in the development of Network Exchange Theory as the first lab of this kind in Europe.

Jacek owed his academic success to his bright intellect, hard work and ambition to keep pace with recent developments in his discipline. With his innovative spirit he was able to locate new research areas such as 'conflict networks' which he began to study with his collaborators a couple of years ago. As a self-made man he welcomed the new funding opportunities opened up to individual scholars when National Committee for Scientific Research (the Polish counterpart of the American NSF) began organizing research proposal competitions. He was among the few Polish sociologists who won research grants three times over the last decade. He also gained an international reputation as a conference organizer and editor of a few collective works of which the most important was the volume Status, Network, and Structure: Theory Development in Group Processes (Stanford 1997) which he co-edited with John Skvoretz and Joe Berger.

In Spring 2000, a sudden attack of strong pain made him seek relief in the hospital. When he learned how serious his disease was he did not fall into depression. He firmly believed he would win the struggle with cancer and worked as hard as he used to. He was planning to upgrade his lab so as to meet the needs of the research designed by his last Ph.D. student, Ms. Kinga Wysieńska, whom he met in Fall 1997. He invited her to join his research team which, until then, included myself and Joanna Heidtman who had been Jacek's primary collaborator in his conflict network research. When he was released from the hospital in Cracow after isotope therapy, his Polish colleagues could see him as active as usual. Soon afterwards, he took part as a session organizer during the 11th

Congress of Polish Sociology in his hometown Rzeszów in September 2000, and then travelled to the University of Iowa to teach and continue his research there as a Fulbright fellow. The therapy he received in Iowa appeared to be working so he welcomed Bob Shelly's invitation to come to Ohio the following year.

I received my last email message from him on September 23. He wrote me that he felt worse again but still believed in his recovery. Some two weeks later I learned from his family that cancer had attacked his lungs and his life was coming to an end. He died on October 20, 2001.

With Jacek's passing Polish sociology has lost an outstanding scholar whose pro--science stance had inspired many students and researchers over the years, even if the radical form in which he occasionally presented his views might have sometimes appeared irritating to some people outside the group processes circle.

He was my friend and closest collaborator with whom I had communicated daily since 1990, regardless of whether he was here in Cracow or somewhere over the ocean (from 1992 to 1998 we exchanged some 1500 email messages). I will remember him, too, as the leader of our small group, formally, the head of the Chair of Group Processes, and last but not least, as the person to whom I owe my contacts with other scholars sharing the idea of scientific sociology which Jacek had outlined in his paper (*On Four Myths about Sociology and Three Generations of Sociological Theories*) which opens the book he co-edited with me ('Structure, Exchange, and Power. Studies in Theoretical Sociology,' in Polish, Warsaw 1993).

May his name and work remain in our memory.

Tad Sozański

Jacek Szmatka's publications in English 1989–2002

Books edited

Szmatka J., Lovaglia M., Wysieńska K. (eds.) (2002). *The Growth of Social Knowledge. Theory, Simulation, and Empirical Research in Group Processes*. Westport, CT-London: Praeger, pp. 299.

Skvoretz J., Szmatka J. (eds.) (1998). *Advances in Group Processes*. vol. 15. Greenwich, CT: JAI Press, pp. 235.

Szmatka J., Skvoretz J., Berger J. (eds.) (1997). *Status, Network, and Structure. Theory Development in Group Processes*. Stanford, CA: Stanford University Press, pp. 467.

Szmatka J., Mach Z., Mucha J. (eds.) (1993). *Eastern European Societies at the Threshold of Change.* New York: The Edwin Mellen Press, pp. 313.

Articles in refereed journals or edited books

Wysieńska K., Szmatka J. (2002). *Positivism and Theory Construction in Group Processes*. In: J. Szmatka, M. Lovaglia, K. Wysieńska (eds.), *The Growth of Social Knowledge. Theory, Simulation, and Empirical Research in Group Processes*. Westport, CT-London: Praeger, p. 77–96.

Heidtman J,. Wysieńska K., Szmatka J. (2000). *Positivism and Types of Theories in Sociology*. Sociological Focus, 33, p. 1–26.

Wysieńska K., Szmatka J. (2000). *Polish and Eastern European Sociology*. In: E. Borgatta (ed.) *Encyclopedia of Sociology*, 2nd ed., vol. 3. New York: Macmillan, p. 2116–2123.

Szmatka J., Skvoretz J., Sozański T., Mazur [Heidtman] J. (1998). *Conflict in Networks*. Sociological Perspectives, 41, p. 49–66.

Szmatka J., Mazur [Heidtman] J. (1998). *Power Distribution in Conflict Networks: An Extension of Elementary Theory to Conflict Networks*. In: Skvoretz, Szmatka (eds.) 1998, p. 187–211.

Szmatka J. (1997). *Testing Elementary Theory for Universality*. In: J. Szmatka, J. Skvoretz, J. Berger (eds.), S*tatus, Network, and Structure. Theory Development in Group Processes*. Stanford, CA: Stanford University Press, p. 87–109.

Willer D., Szmatka J. (1997). *Structural Formulations and Elementary Theory*. In: J. Szmatka, J. Skvoretz, J. Berger (eds.), S*tatus, Network, and Structure. Theory Development in Group Processes*. Stanford, CA: Stanford University Press, p. 273–292.

Szmatka J., Lovaglia M. (1996). *The Significance of Method*. Sociological Perspectives, 39, p. 393–415.

Szmatka J., Mazur [Heidtman] J. (1996). *Orienting Strategies, Working Strategies, and Theoretical Research Programs in Social Exchange Theory*. Polish Sociological Review, 115 (1996/3), p. 265–288.

Willer D., Simpson B., Szmatka J., Mazur [Heidtman] J. (1996). *Social Theory and Historical Explanation*. Humboldt Journal of Social Relations, 22, p. 63–84.

Szmatka J., Willer D. (1995). *Exclusion, Inclusion and Compound Connection in Exchange Net works*. Social Psychology Quarterly, 58, p. 123–132.

Szmatka J., Sozański T. (1994). *On Four Myths of Sociology and Three Generations of Sociological Theories*. Polish Sociological Review, 107 (1994/3), p. 219–233.

Willer D., Szmatka J. (1993). *Cross-National Experimental Investigations of Elementary Theory: Implications for the Generality of the Theory and the Autonomy of Social Structure*. In: Markovsky B. (ed.). *Advances in Group Processes*. Vol. 10. Greenwich, CT: JAI Press, p. 37–81.

Szmatka J., Mach Z., Mucha J. (1993). *Introduction. In Search of the Syndrome of Threshold Situation*. In: Szmatka, Mach, Mucha (eds.) 1993, p. 1–13.

Szmatka J. (1992). *Polish Sociology*. In: Borgatta E., Borgatta M. (eds.). *The Encyclopedia of Sociology*, vol. 3. New York: Macmillan, p. 1471–1477.

Szmatka J., Warriner C.K. (1991). *The Search for a Structural Paradigm in Sociology*. Dialogue and Humanism, 1, p. 109–132.

Szmatka J. (1990). *The Relation Between Group Structure and Intra-Group Tensions and Conflict*. International Journal of Group Tensions, 20, p. 3–29.

Szmatka J. (1989). *Reduction in the Social Sciences: The Future or Utopia*? Philosophy of the Social Sciences, 19, p. 425–444.

Szmatka J. (1989). *Individualism, Holism, Reductionism*. International Sociology, 4, p. 169–186.

References

Aronson E., Ellsworth P.C., Carlsmith J.M., Gonzales M.H. (1990). *Methods of Research in Social Psychology*, 2nd ed. New York: McGraw-Hill.

Babbie E. (2014). *The Practice of Social Research*, 14th ed. Boston, MA: Cengage Learning.

Berger J. (2014). *The Standardized Experimental Situation in Expectation States Research. Notes on History, Uses, and Special Features*. In: M. Webster Jr., J. Sell (eds.), *Laboratory Experiments in the Social Sciences*, 2nd ed. San Diego, CA: Elsevier, p. 269–293.

Berger J. (1974). *Expectation States Theory: A Theoretical Research Program.* In: J. Berger, T.L. Conner, M.H. Fişek (eds.), *Expectation Status Theory: A Theoretical Research Program*. Cambridge, MA: Winthrop, p. 3–22.

Berger J., Fişek M., Norman R.Z., Zelditch M., Jr. (1977). *Status Characteristics and Social Interaction. An Expectation States Approach.* New York: Elsevier.

Berger J., Zelditch M., Jr., Anderson B. (eds.) (1972). *Sociological Theories in Progress*, Vol. 2. Boston: Houghton-Mifflin.

Blalock H.M., Jr. (1969). *Theory Construction. From Verbal to Mathematical Formulations*. Englewood Cliffs, NJ: Prentice-Hall.

Brown R. (2000). *Group Processes. Dynamics Within and Between Groups,* 2nd ed. Oxford: Blackwell.

Brzeziński J. (1996). *Metodologia badań psychologicznych* [Methodology of Psychological Research]. Warszawa: Wydawnictwo Naukowe PWN.

Cartwright D., Zander A. (eds.) (1960). *Group Dynamics. Research and Theory,* 2nd ed. New York: Harper and Row.

Cohen B.P. (1989). *Developing Sociological Knowledge: Theory and Method,* 2nd ed. Chicago, IL: Nelson-Hall.

Cohen B.P. (1980). *The Conditional Nature of Scientific Knowledge*. In: L. Freese (ed.), *Theoretical Methods in Sociology: Seven Essays.* Pittsburgh, PA: The University of Pittsburgh, p. 71–110.

Collins B.E., Raven B.H. (1969). *Group Structure: Attraction, Coalitions, Communication and Power*. In: G. Lindzey, E. Aronson (eds.), *Handbook of Social Psychology*, Vol. 4. Reading, MA: Addison-Wesley, p. 102–204.

Cook K.S., Emerson R.M. (1978). *Power, Equity and Commitment in Exchange Networks*. American Sociological Review, 43, p. 721–739.

Czarnik S. (2006). *Justice in the Shadow of Self-Interest. An Experiment on Redistributive Behavior.* Acta Physica Polonica B, 37, p. 2967–2977.

Foschi M. (2000). *Expectation States Theory*. In: E.F. Borgatta, R.J.V. Montgomery (eds.), *Encyclopedia of Sociology*, 2nd ed., vol. 2, New York: Macmillan, p. 880–886.

Foschi M. (1997). *On Scope Conditions*. Small Group Research, 28, p. 535–555.

Foschi M.,Warriner G.K., Hart S.D. (1985). *Standards, Expectations, and Interpersonal Influence.* Social Psychology Quarterly, 48, p. 108–117.

Freese L. (1980). *Formal Theorizing*. Annual Review of Sociology, 6, p. 187–212.

Friedkin N.E. (1992). *An Expected Value Model of Social Power: Predictions for Selected Exchange Networks*. Social Networks, 14, p. 213–229.

Giddens A. (1984). *The Constitution of Society: Outline of the Theory of Structuration*. Oxford: Polity Press.

Hamlin J.K., Wynn K., Bloom P. (2007). *Social Evaluation by Preverbal Infants*. Nature, 450, p. 557–559.

Lakatos I. (1970). *Falsification and the Methodology of Scientific Research Programs.* In: I. Lakatos, A. Musgrave (eds.), *Criticism and the Growth of Knowledge*. Cambridge, MA: Cambridge University Press, p. 91–195.

Lovaglia M.J., Skvoretz J., Willer D., Markovsky B. (1995). *Negotiated Exchanges in Social Networks*. Social Forces, 74, p. 123–155.

Markovsky B. (1997). *Building and Testing Multilevel Theories*. In: J. Szmatka, J. Skvoretz, J. Berger (eds.), *Status, Network, and Structure. Theory Development in Group Processes*. Stanford: Stanford University Press, p. 14–28.

Markovsky B., Willer D., Patton T. (1988). *Power Relations in Exchange Networks*. American Sociological Review, 53, p. 220–236.

Merton R.K. (1968). *On Sociological Theories of the Middle Range*. In: *Social Theory and Social Structure*. New York: The Free Press, p. 39–72.

Molm L.D. (2014). *Experiments on Exchange Relations and Exchange Networks in Sociology*. In: M. Webster Jr., J. Sell (eds.), *Laboratory Experiments in the Social Sciences*, 2nd ed. San Diego, CA: Elsevier, p. 199–224.

Skvoretz J., Fararo T.J. (1998). *Theoretical Models: Sociology's Missing Links*. In: A. Sica (ed.), *What is Social Theory? The Philosophical Debates*. Oxford, UK: Blackwell, p. 238–252.

Smith V.L. (1999). *Reflections on 'Human Action' after 50 Years*. Cato Journal, 19, p. 195–214.

Smith V.L. (1982). *Microeconomic Systems as an Experimental Science*. American Economic Review, 72, p. 923–955.

Sozański T. (2006). *On the Core of Characteristic Function Games Associated with Exchange Networks*. Social Networks, 28, p. 397–426.

Sozański T. (2001). *Jacek Szmatka,* 28 *marca* 1950 – 20 *października* 2001 [Jacek Szmatka, 28th March, 1950 – 20th October, 2001. Obituary]. Studia Socjologiczne, 163 (2001/4), p. 5–10.

Sozański T. (1997). *Toward a Formal Theory of Equilibrium in Network Exchange Systems*. In: J. Szmatka, J. Skvoretz, J. Berger (eds.), *Status, Network, and Structure. Theory Development in Group Processes*. Stanford, CA: Stanford University Press, p. 303–350.

Sozański T. (1993). *Hierarchiczne systemy wymiany. Powtórzenie eksperymentu Davida Willera* [Hierarchical Exchange Systems. A Replication of an Experiment of David Willer]. In: T. Sozański, J. Szmatka, M. Kempny (eds.), p. 233–271, p. 308–309 (summary).

Sułek A. (1989). *The Experiment of Psammetichus: Fact, Fiction, and Model to Follow*. Journal of the History of Ideas, 50, p. 645–651.

Sułek A. (1979). *Eksperyment w naukach społecznych* [Experiment in the Social Sciences]. Warszawa: PWN.

Szaniawski K. (1976). *Types of Information and Their Role in the Methodology of Science*. In: M. Przełęcki, K. Szaniawski, R. Wójcicki (eds.), *Formal Methods in the Methodology of Empirical Sciences*. Wrocław-Warszawa-Kraków-Gdańsk: Ossolineum, p. 297–308.

Szmatka J. (1997). *Testing Elementary Theory for Universality*. In: J. Szmatka, J. Skvoretz, J. Berger (eds.), *Status, Network, and Structure. Theory Development in Group Processes*. Stanford, CA: Stanford University Press, p. 87–109.

Szmatka J., Lovaglia M. (1996). *The Significance of Method*. Sociological Perspectives, 39, p. 393–415.

Szmatka J., Skvoretz J., Sozański T., Mazur [Heidtman] J. (1998). *Conflict in Networks*. Sociological Perspectives, 41, p. 49–66.

Szmatka J., Sozański T. (1994). *On Four Myths of Sociology and Three Generations of Sociological Theories*. Polish Sociological Review, 107 (1994/3), p. 219–233.

Thye S. (2014). *Logical and Philosophical Foundations of Experimental Research in the Social Sciences*. In: M. Webster Jr., J. Sell (eds.), *Laboratory Experiments in the Social Sciences*, 2nd ed. San Diego, CA: Elsevier, p. 53–82.

Torgerson W.S. (1958). *Theory and Methods of Scaling*. New York: John Wiley & Sons.

Toulmin S. (1953). *The Philosophy of Science*. London: Hutchinson.

Triplett N. (1898). *The Dynamogenic Factors in Pacemaking and Competition*. American Journal of Psychology, 9, p. 507–533.

Turner J.H. (1985). *In Defense of Positivism*. Sociological Theory, 3, p. 24–30.

Webster M., Jr., Sell J. (2014). *Why Do Experiments?* In: M. Webster Jr., J. Sell (eds.), *Laboratory Experiments in the Social Sciences*, 2nd ed. San Diego, CA: Elsevier, p. 5–21.

Willer D. (1987). *Theory and the Experimental Investigation of Social Structures*. New York: Gordon and Breach.

Willer D. (1985). *Property and Social Exchange*. In: E.J. Lawler (ed.), *Advances in Group Processes,* vol. 2. Greenwych, CT: JAI, p. 123–142.

Willer D. (1981a). *The Basic Concepts of the Elementary Theory*. In: D. Willer, B. Anderson (eds.), *Networks, Exchange and Coercion. The Elementary Theory and Its Applications.* New York: Elsevier-Greenwood, p. 25–53.

Willer D. (1981b). *Quantity and Network Structure.* In: D. Willer, B. Anderson (eds.), *Networks, Exchange and Coercion. The Elementary Theory and Its Applications*. New York: Elsevier-Greenwood, p. 109–127.

Willer D., Markovsky B. (1993). *The Theory of Elementary Relations: Its Development and Research Program*. In J. Berger, M. Zelditch Jr. (eds.), *Theoretical Research Programs: Studies in Theory Growth*. Stanford: Stanford University Press, p. 323–363.

Willer D., Walker H.A. (2007). *Building Experiments. Testing Social Theory*. Stanford: Stanford University Press.

Wysieńska K., Szmatka J. (2002). *Positivism and Theory Construction in Group Processes*. In: *The Growth of Social Knowledge. Theory, Simulation, and Empirical Research in Group Processes*. Westport, CT-London: Praeger, p. 77–96.

Eksperymentalna nauka społeczna

Redaktorski wstęp do zbioru artykułów w języku angielskim, zatytułowanego *Selected Topics in Experimental Social Science* [Eksperymentalna nauka społeczna: wybrane zagadnienia], wypełniającego prawie cały bieżący numer *Studia Sociologica*, powstał przez rozbudowanie do rozmiaru artykułu prezentacji pokazanej przeze mnie jako organizatora na sesji otwierającej międzynarodowe sympozjum („International Symposium on Experimental Research in the Social Sciences"), które odbyło się Uniwersytecie Pedagogicznym w Krakowie w dniach 12−13 czerwca 2015. Artykuł ten łączy metateoretyczne rozważania na temat teorii i eksperymentu w naukach społecznych z informacjami o eksperymentach laboratoryjnych wykonanych w Zakładzie Badania Procesów Grupowych Instytutu Socjologii UJ od powstania (1989) tej nieistniejącej już placówki badawczej do przedwczesnej śmierci (2001) jej założyciela prof. Jacka Szmatki.

Słowa kluczowe: eksperyment, metodologia nauk empirycznych, trzy generacje teorii socjologicznych, sieciowy system interakcji

FOLIA 202

Annales Universitatis Paedagogicae Cracoviensis

Studia Sociologica VIII (2016), vol. 1, p. 42–61
ISSN 2081–6642

Martha Foschi
University of British Columbia, Canada

Experimental Contributions to Sociological Immigration-Research

Abstract

Experiments have not been common in immigration research. In this article I argue that there is a need for more sociological work utilizing that methodology. I report on a review I conducted of experiments on immigration topics, discuss the advantages of those studies, and propose lines for future sociological experimental research in this area. I also describe and counter the often expressed concerns about experiments in the social sciences – namely, those about (1) artificiality of the designs and (2) limited generalizability of the findings. Full references from the review are presented.

Key words: experiments, immigration, artificiality, generalizability

To the memory of my grandparents:
Magdalena and José, Catalina and Luis, with love.

Introduction

Millions of persons around the world are currently either emigrating from their birthplaces or planning/dreaming of doing so. The reasons are many and they include: the prospect of a better future in economic, educational and/or health-related terms; an escape from political and/or religious persecution and warfare; and family reunification. This situation constitutes a major social issue and has generated a variety of policies on the part of the states involved, ranging from a welcoming accommodation to the imposition of harsh barriers by the intended immigration-countries, and from benevolent attitudes to restrictive measures by the potential emigration-lands. For refugees, travelling to their aimed destination often involves many perils, including grave risks at the hands of traffickers.

Not surprisingly, immigration now also attracts the research interest of an increasing number of sociologists. Examples – to name only a few – of the topics commonly investigated in this discipline in relation to immigrants are: employment, wages, acculturation, new-language acquisition, mother-language maintenance, educational accomplishments of their offspring, and residential segregation. These often relate to rejection by the native-born on many fronts on the basis of prejudices

and stereotypes linked to skin colour and country of origin. A wide diversity of theories are employed in the study of these topics, and the evidence comes from a variety of countries. In terms of methodology, most of the sociological research in immigration uses surveys (including census data) and ethnographies, with experiments being underutilized. On this latter point see, for example, McDermott (2006) about political science; McKenzie and Yang (2012) about immigration.

Experiments can be a powerful instrument for both discovering relationships among variables, as well as for testing theories. Here I argue for more work employing that methodology in immigration research, particularly in sociology. The following is the organization of my article I:

(a) highlight the logic and advantages of experiments,

(b) report on a review that I have carried out of published experiments on immigration topics,

(c) show the variety of experimental designs and techniques that can be used, as well as the topics that can be addressed in this area, and propose lines for future sociological experimental work within it,

(d) suggest improvements to a sample of the reviewed studies, and

(e) describe and counter the often expressed concerns about experiments in the social sciences – namely, those about (1) artificiality of the designs and (2) limited generalizability of the findings.

Definitions

An experimental design investigates the effects of one or more *independent* (antecedent) variables on one or more *dependent* (consequent) variables. An example would be to study the impact of the extent of similarity among the members of a group on the strength of the affect ties that they develop among themselves. Levels of the independent variable are the experimental (or treatment) conditions (e.g., either high or low similarity in the group members' extent of formal education, as reported to them by the experimenter); often a control (or baseline) condition consists of either a neutral level or no information in this respect.

Other factors may also be incorporated, as follows. *Scope conditions* are clauses specifying the circumstances under which a hypothesis is proposed (Berger, Zelditch 1977, p. 25–28; Cohen 1980; Foschi 1997). For instance, if the group is engaged in completing a valuable task, high motivation to do so could be added as a scope condition in the relationship between similarity and affect. Additionally, that relationship may be elaborated by the inclusion of *intervening* constructs, namely, factors that mediate the link between the independent and the dependent variables. One example of such construct would be the ease with which group members communicate with each other. In turn, depending on the particular study and its theoretical framework, these constructs may or may not be explicitly associated with observables.

Additional components of an experiment include *test limitations* and *theoretically irrelevant factors*. The former are particular characteristics of the study that are often mentioned by the researcher because of their *potential* relevance to the topic under consideration (e.g., the respondents' average age). They are not part of the

hypothesis tested. The latter are factors that could reasonably have been expected to be part of the theoretical argument but that the author explicitly lists as immaterial (e.g., if the relationship is initially assumed to hold *regardless* of task difficulty – not a trivial assumption) and are allowed to vary. It is left to empirical results to confirm or disconfirm the assumption – the author may either provide results from previous studies or suggest direction for future research in this respect. For a detailed discussion and additional examples of the various factors in an experimental design, see Foschi (2014); Jackson and Cox (2013).

In a true experiment participants are assigned at random to different levels of each independent variable – e.g., to either high or low solidarity among group members. These levels are created (manipulated) by the experimenter – for instance, this is often done through statements that emphasize either the common or the differentiating traits that the participants supposedly have.[1]

The aim is to control all other factors, so that it can be claimed that the independent variable(s) is (are) responsible for changes in the dependent variable(s). This can be achieved in three ways: (1) by keeping other factors either constant or within a narrow range, (2) by defining some factors as irrelevant (or at least irrelevant for the time being), and (3) by using random assignment to experimental and control conditions. These measures enable the researcher to identify with high certainty how a set of variables are related – an outcome that is the key theoretical advantage of experimentation (see Webster, Sell 2014; Willer, Walker 2007, Chapter 4).

There are also other types of designs in which participants can be classified into two or more levels of an antecedent variable. However, if the latter are inherent to those persons rather than created and randomly assigned by the researcher, the design is not an experiment. Because they 'belong' to the participants, such factors are often referred to as 'organismic' or 'quasi-experimental' variables. Sex-category and ethnic-group membership are two examples, and such attributes *are not truly experimental variables* if they refer to characteristics of the respondents (on this point, see McBurney 1983, p. 139; Magnusson, Marecek 2012, p. 174). On the other hand, they *are* such variables if they describe characteristics of *others*, and respondents are assigned at random to different levels of those factors – for example, if the task is to evaluate a task performer who is introduced as either a man or a woman.

[1] In this text I use the following pairs of terms as synonyms: 'participant' and 'respondent,' and 'variable' and 'factor.' When it is clear from the context, sometimes I also use 'experiments' and 'studies' interchangeably. I do, however, distinguish between 'experimental design' and 'experimental technique.' I use 'design' to refer to the variables under investigation and to the random assignment of participants to different conditions, and 'technique' to describe the particular means through which the latter are created (e.g., the participants could be asked to perform a visual perception task, or to choose between job applicants, etc.). Moreover, for simplicity, I often employ 'researcher,' 'hypothesis' and 'variable' in the singular, although in many cases it would be more accurate to use the plural form. Also for brevity, when I discuss the design of an experiment, 'assigned' means 'randomly assigned.' The term 'issue' refers to either a topic or subject matter, or to a part of a journal's volume (as in, e.g., Volume 4, Winter Issue). Finally, in line with common use in this literature, 'immigration topics, matters, etc.' involves both immigrants and refugees.

Sometimes the same variable can, in different situations, be *both* a truly experimental factor and an organismic one. For example, consider studying two groups of individuals, one consisting of immigrants and the other of native-born, with both groups being asked to complete a questionnaire on their own career preferences. 'Nativity' is not an experimental factor in that case, as no random assignment is involved. However, this variable could be truly experimental in other situations, such as having participants assigned at random to assess the suitability of two job applicants who are alike in all key social characteristics except nativity. In this article I include only studies in which either nativity or related variables (or both) are truly experimental factors, as I illustrate later in the section *Selected experiments – highlights*.

Before discussing specific studies, some additional definitions will be helpful. They are as follows.

(1) Experiments are often classified as either 'laboratory' or 'field.' That classification is often difficult to make, as there are many grey areas in between. The distinction is useful only if one remembers that, regardless of where a study is carried out, it is an experiment if and only if random assignment to conditions takes place. I exclude the so-called 'natural experiments,' or events that result in contrasting situations (e.g. a flood of a specified intensity occurring in one region of a country but not in another, even though the two regions have similar physical and environmental features) because they do not meet the condition of control by an experimenter.

(2) A 'survey experiment' (also often referred to as 'experimental survey') is one in which the main instrument for data collection is a questionnaire in which various, randomly embedded items represent different levels of the independent variable(s). The design is indeed experimental, but in this case the central interest is on making inferences about a sample (preferably representative) from a particular population defined in terms of time and place (see also section *On artificiality and generalizability* later in this text).

(3) Finally, experiments may be either 'exploratory' or 'hypothesis testing.' Although both types can be valuable to immigration research, my focus here is on the latter. That is, I am interested in how experiments can contribute to either testing/developing new hypotheses within established theories, or helping in the design of theory-based interventions, or both.

It is also important to note that I am not advocating experiments as the only or even the most important methodology for immigration research. There are many types of experimental situations that either cannot be created or it would be unethical to do so. Consider, for example, a task group consisting of two persons, A (a confederate of the experimenter) and B (a true participant). Level of A's evaluation of B's task performance is the independent variable while B's reaction to the evaluation is the dependent variable. If A's judgment is one of extremely high praise, the situation would probably become unbelievable to B. If, on the contrary, A's assessment is so low that it is meant to humiliate B, ethical considerations would indicate that the experiment should not be carried out. My purpose in this article is to point out that, with a degree of good common-sense, respect for the participants, and a creative yet realistic imagination, many sociological immigration-topics *can* be studied experimentally.

Literature search

I begin by outlining the limits of my present review. I completed an issue-by-issue online check of nine selected English-language academic periodicals for the years 2000 to 2015, and I also carried out, for the same period, an online search using 'journal article,' 'experiments,' 'immigrants,' 'refugees' and 'immigration' as key words. This time period is particularly appropriate here, given the increased academic interest in immigration matters that has occurred recently. My objective is to investigate empirically the extent to which experiments have been underutilized in this area.[2]

The choice of journals reflects my assessment of which ones would be likely to contain articles on immigration-related experiments. It also corresponds with my focus on immigration and social factors. Thus, I have excluded from my issue-by-issue search those journals that specialize in either economics or psychology outside social psychology – on the assumption that although experiments are common in them, immigration topics are not. The periodicals I reviewed were: *British Journal of Social Psychology*; *Canadian Journal of Behavioural Science* (this journal publishes work in English, as well as in French; for this review however I have considered only those in English), *European Journal of Social Psychology*; *Group Processes & Intergroup Relations*; *Journal of Experimental Social Psychology*; *Journal of International Migration and Integration*; *Political Psychology*; *Social Justice Research*; and *Social Psychology Quarterly*. My search identified 68 articles comprising a total of 106 studies on immigration matters, all truly experimental designs with respect to at least one independent variable.[3] Overall, their data were obtained in 17 countries (in addition, in two cases the location of the research is not specified but the participants are described as 'British'). Please refer to Appendix B. Inclusion in this bibliography implies that either the independent or the dependent variables of an article concern immigrants themselves (in some studies, 'immigrant' is interpreted in a general sense; in others, the term applies to people from particular

[2] Note that it is now generally the case that academic experiments have to be approved by the ethical committees of the respective institution(s) where the researcher(s) is (are) located. Here I am not referring to any so-called 'social experiments' in which immigrants and members of other vulnerable populations would participate, without their knowledge or consent, in experimentally created situations such as housing assignments.

[3] Immigrants do, of course, vary overall in such factors as nationality and ethnic background. I have excluded from this review those studies that concern either or both of those factors but do not investigate nativity and/or related variables. Similarly, I have excluded those works in which the persons to be considered by the respondents are described in terms of 'majority/minority,' 'in-group/out-group' or 'foreign/national' rather than 'native-born/immigrant (or native-born/refugee).' I also excluded a study if the out-group is likely to be composed mostly of immigrants but the authors do not identify them as such. In addition, my review does not list (a) those studies in which *participants themselves* are classified as being either native-born or immigrant (that is, those two levels do not constitute an experimental variable) and (b) random assignment to conditions is not explicitly linked to immigration topics. If participants are either native-born or immigrant, or both groups are represented, those characteristics may be considered test limitations.

geographic areas) and/or immigration topics related to them. I have excluded those studies that have physiological measures as the dependent variables.

The total number of issues I inspected in the nine journals for the 2000–2015 years is 682. Of the 68 articles that I identify in the above paragraph, 62 were published in one of those issues – the remaining 6 were from the online search. This represents 9.09% of the works (62/682) – a small figure given the extent of the search. (If one counts experimental studies rather than articles in those same journals, the number increases to 99. In that case, the proportion would have to be calculated in relation to the total number of studies representing various methodologies appearing in those issues. One should also consider that, in several instances, the various studies presented in a single article are replications of the one that appears first.)

Most of the 62 journal articles originate in the psychology-based social psychology literature[4], but there are also some from sociology, economics, and political science. Not surprisingly, the periodical that contains the most publications on immigration-related experiments (17) was the *Journal of Experimental Social Psychology*; the two that followed were *Group Processes & Intergroup Relations* (9) and *Political Psychology* (8). Although a more extensive search (e.g., including other journals and earlier years) could have been conducted, I estimate from my reading of the immigration-research literature that it would not have yielded a significantly larger number of studies, and that their subject-matters and designs would not have been considerably different.

The majority of the empirical immigration-studies appearing in these journals are surveys (including census data) and ethnographies. This finding is in line with the view I expressed earlier in the *Introduction*. It is also consistent with McDermott (2014) and McKenzie and Yang (2012) in their respective assessments that experiments on immigration topics have not been generally common in the social sciences. Indications are, however, that their number is currently growing in the published work I have examined. For reviews that reflect the increased interest in this methodology in two social science fields see, e.g., McDermott (2014) on political science, and Thau et al. (2014) on organizational behaviour.

Overall, the works I identified in my review reveal the different types and levels of difficulties that immigrants and refugees often experience in their new lands.

Selected experiments – highlights

In this section I examine in some detail a sample of seven (of the 68) articles that I have selected to highlight their diversity in research topics and techniques. The seven contain *nine* studies (see Appendix A for the full references). I chose these experiments because they represent useful illustrations of several points I wish to

[4] In some circles, it is common to distinguish between 'sociological (or structural) social psychology' and 'psychological social psychology.' It is proposed that the emphasis of the former is on the social context while the latter focuses on the individual. But the difference between the two approaches is neither always sharp nor generally agreed-upon. Thus, my view is that it is useful to think of them in terms of shades of differences.

make – my selection does not imply a comparative judgment of the quality of these studies relative to the others identified here. Those other experiments are listed in Appendix B as additional examples.

The following is a brief description of the nine experiments, with a focus on those independent variables for which there was random assignment. Although limited in size, the set is highly informative, as I discuss next in the *Overview* section of this article.

1. Caprariello, Cuddy, Fiske 2009. Participants read *vignettes* (brief accounts of situations) depicting an unfamiliar ethnic group said to be immigrating to the US in the near future. Each person was assigned to a scenario describing the group members in terms of their status (either high or low) and competitiveness (either high or low). The dependent variables assessed warmth-competence stereotypes and emotional prejudices. *Source of data*: Male and female undergraduates from a private university in the US Northeast.
2. Jackson, Esses 2000 (Study 1). Participants were divided at random to read one of two editorials about immigration (one focusing on economic competition; the other, a control condition that described vague, general immigration trends). The former condition resulted in lower levels of empowerment forms of assistance to immigrants. *Source of data*: Male and female undergraduates at the University of Western Ontario, Canada.
3. Jetten, Wohl 2012 (Studies 1 and 2). In each case, respondents were assigned to read 'a travel guide webpage' that showed either high or low continuity (the former was depicted as 'high homogeneity') between England's past and present history; respondents were also measured regarding their identification with England. Those who revealed a strong degree of such identification indicated more opposition to immigration in the low rather than in the high historical-continuity conditions. *Source of data*: Participants were undergraduates in England, identified by the authors of the article as English and either men or women.
4. Joona, Nekby 2012. New immigrants were assigned to either a treatment condition (intensive counselling and coaching by Public Employment Service caseworkers) or a control group (regular introduction-programs). The results indicate significant treatment-effects on both actual employment rates and participation in intermediate training programs. *Source of data*: Men and women, newly-arrived immigrants to Sweden.
5. Oreopoulos 2011. Using a technique often referred to as an 'audit study,' thousands of sets containing four résumés each were sent in response to online job-postings across multiple occupations. Sets were constructed to plausibly represent, in random combinations, either recent immigrants from Britain, China, India, and Pakistan, or non-immigrants – in each case either with or without 'ethnic' sounding names. Levels of other independent variables, also built at random into the sets were: place of undergraduate degree, whether job experience had been in Canada or abroad, and number of languages in which the applicant was fluent. *Ceteris paribus*, number of call-backs show employer discrimination against

applicants with 'ethnic' names and with only foreign-country work experience. *Source of data*: Male and female employment officers from various companies in the Greater Toronto Area, Canada.

6. Ramos et al. (Studies 2 and 3). In the first of these two experiments, respondents were assigned at random to read one of two versions of a brief paragraph about the relationship between Romanian immigrants in France and the French majority population. In one case, the paragraph described both high discrimination against immigrants by the French and perceived high discrimination by Romanian immigrants; in the other case, both actual and perceived group discrimination were presented as low. Respondents' minority goals (separation acculturation-strategies) were also assessed. Perceived group discrimination affected identification with other Romanians only when minority goals emphasized seeking distance from the majority. In Study 3, the paragraph was about Polish immigrants in Scotland and their relationship with Scottish people. The findings replicated those of Study 2. *Sources of data*: Romanian immigrants living in France, and Polish immigrants living in Scotland, respectively. Both samples contained men and women.
7. Vezzali et al. 2012a. Children were divided at random to participate in either a three-session intervention that involved imagining meeting an unknown immigrant-peer in various situations, or a control condition without such an imagined meeting. Those taking part in the intervention, compared to the participants in the control group, revealed more positive behavioural intentions/implicit attitudes towards immigrants. A willingness to disclose a major personal problem or secret with an immigrant child mediated the effect on intentions. *Source of data*: Fifth-grade boys and girls in Italy.

Overview

All nine selected studies show the rich variety of immigration topics that can be and have been investigated experimentally and, overall, reveal a high degree of creativity in the development of experimental situations. These studies also differ in the designs that they use and in their sample populations (e.g., in terms of age, gender, and country of residence). The theoretical backgrounds represented include, among others, stereotype content (Caprariello, Cuddy, Fiske 2009), intergroup competition and social dominance (Jackson, Esses 2000), and imagined intergroup contact (Vezzali et al. 2012a). The studies vary in the extent to which hypotheses are explicitly stated.

Except for the intervention designed by Joona and Nekby (2012) to help through counselling and coaching, the results from this set reveal the various disadvantages associated with being an immigrant. In general, as I indicate earlier, this is also the case of the present larger set of studies.

The *truly independent variables* were manipulated in several ways. Thus, there were vignettes, as well as manufactured editorials about: immigrant groups differing in their social status and competitiveness (Caprariello, Cuddy, Fiske 2009; Jackson,

Esses 2000), degrees of actual and perceived discrimination (Ramos et al. 2013), and the history of the immigration country as either high or low in the continuity of its constituent groups (Jetten, Wohl 2012). In addition, in one study (Oreopoulos 2011), job applicants were presented as either immigrants or non-immigrants and they also varied in country of origin, languages, and place of both undergraduate degree and job experience; in another (Vezzali et al. 2012a), respondents were asked either to envision meeting an unknown immigrant-peer or to participate in a control group. Joona and Nekby (2012) either did or did not provide a group of immigrants with intensive counselling and coaching by public officials.

Ramos et al. (2013, Studies 2 and 3) present an example of an additional, immigration-related *quasi-experimental variable* measured through respondents' answers, namely, support for minority goals that advocate a distanced stance vis-à-vis the majority group.

A few studies list clauses such as the following: an immigrant group whose ethnicity was unfamiliar to the respondents (Caprariello, Cuddy, Fiske. 2009), a strong identification with the participants' own country (Jetten, Wohl 2012, both studies), and an emphasis on distancing oneself from the host nation' majority (Ramos et al. 2013). Although the authors do not designate these clauses as scope conditions, they are, in fact, such theoretical limitations of their respective hypotheses. Regarding intervening variables, Vezzali et al. (2012a) explicitly incorporate willingness to self-disclose, while Jetten and Wohl (2012) add confidence in the historical group's future vitality.

Experimental limitations which are worth-noting concern the country of study and characteristics of the respondents. Here again, even the small set of nine experiments shows useful variation: as can be seen from the summaries in the previous section, participants have included undergraduates, elementary-school students, immigrants, and employment officers. The respondents' countries are Canada, England, France, Italy, Scotland, and the US. The immigrants are described as from Britain, China, India, Pakistan (Oreopoulos 2011) or Poland and Romania (Ramos et al. 2013); in the remaining cases, they are not identified by country of origin. It would be worth investigating whether or not that variation affects the findings. In addition, all nine studies have male and female participants and show results indicating that this factor, although of potential theoretical relevance, has been for the most part found to be *irrelevant* to the topic under study – an outcome that deserves attention.

The *dependent variables* consist of either written responses or actual behaviours concerning immigrants. Overall, these variables deal with the degree of opposition to them in various forms, such as lower degrees of support for empowering ways of help (Jackson, Esses 2000), hiring discrimination (Oreopoulos 2011), warmth-competence stereotypes and emotional prejudices (Caprariello, Cuddy, Fiske 2009), and low probability of both employment and participation in training programs (Joona, Nekby 2012).

I conclude this section with a note on ethical considerations. For example, the work by Oreopoulos (2011) uses a technique by which participants do not know that they are being investigated and that they are part of an experiment. In both respects the work raises questions about several ethical points, but in my reading of

the author's report I could not find a reference to how the latter have been dealt with. If they had indeed been addressed, it would have been useful to include a sentence to the effect that 'procedures were in line with the university's research policies relating to human participants.' This latter recommendation also applies more generally to the set of nine. However, since in all cases the articles have been written by university-affiliated authors, it is likely that ethical requirements have been met. Only in one of the nine experiments (namely, Joona, Nekby 2012) do the researchers explicitly indicate sensitivity to one aspect of the experimental design that they use. Their work was part of a Swedish program that provided immigrants with extra coaching in the preparation of job applications. The authors point out that random assignment helps most – but not all – of the immigrant participants and suggest steps to remedy this situation. (If one considers the entire set of 68 articles, there is a considerable number of authors who indicate that 'informed consent has been obtained,' or 'the participants were debriefed' or 'while mean responses of the overall sample were to be made public, they would remain personally anonymous.' Two examples are Barreto et al. (2003, Studies 1 and 2) and Beaupré (2003, Study 3).

Suggestions for further research

My comments in the present section refer more widely to all the studies identified in this review, rather than only to the selected nine experiments.

– Although my search yielded several experiments in a variety of immigration topics of sociological interest, most of the 68 articles (106 studies in total) originate in psychological approaches that contain individual-level variables only. Thus, in Appendix A, the only two works that concern groups are: Joona and Nekby (2012), which considers person-to-person interactions between an immigrant looking for work and a city officer providing intensive coaching about the application procedure, and Vezzali et al. (2012a), which focuses on native-born children, each of whom is involved in an imagined meeting with an unknown immigrant peer. For examples of Appendix B experiments that include groups (that is, each participant is involved in an interaction, either actual, or computer simulated, or imagined/anticipated with at least one specific other person), see Aydin et al. (2014, Study 1); Harwood et al. (2011); Siem, Lotz-Schmitt, Stürmer (2014, Studies 1, 2 and 3); Vezzali et al. (2015). In both Appendices, I have identified all such articles with an asterisk (*). Since they do not involve group interaction, I am excluding from this set those studies in which respondents are presented with vignettes about an immigrant person and asked about their reactions towards him/her (e.g., Stroessner et al. 2015) or 'editorials,' 'news stories,' 'web pages,' or 'results from surveys' on immigration issues and asked about the extent to which they agreed with those communications. It would be worthwhile to expand work in this area by adding other designs that either feature group-level variables or, at least, measure effects from the social context (note, e.g., that Caprariello, Cuddy, Fiske 2009 is the only study among the nine that incorporates

the concept of 'social structure'). A similar point about situational effects is also made in Berry (2001); Esses et al. (2008, p. 16); Tauber and van Zomeren (2013, p. 155–158); Webster and Rashotte (2009). In Foschi (2013), I illustrate how a sociological approach can be followed to investigate experimentally the practice, common in several contemporary societies, of discounting the professional credentials of immigrants.

I propose that a large part of the experimental designs that have been used in sociology and related disciplines to study biases from gender, ethnicity, and in-group/out-group classifications can be readily adapted to immigration topics (see, for example, the valuable review presented by Goar et al. (2013, p. 57–62). There is a vast literature in those areas, including interventions devised to create equal relations among task performers who differ in status. See Cohen (1994) for her significant research program on how to foster equality among different ethnic groups in task settings.

– It would also be important to develop long-terms plans that incorporate both testing the same ideas with different designs, and carrying out systematic replications. I outline both strategies below, as part of my discussion of artificiality and generalizability. (On the advantages of using different methodologies to study a given subject see, for example, Valentino, Brader, Jardina 2013, p. 152; Verkuyten 2005, p. 236–238). As well, it would be beneficial to carry out longitudinal studies – an approach that is still rare in the social sciences. Through such designs, snapshots obtained at particular times could be combined and result in a more complete picture of the topic under consideration.
– Finally, note that the hypotheses investigated in the nine studies originate in different theories and/or approaches, such as social identity, in-group/out-group, social distance, and social dominance. It would be worthwhile to attempt to find areas of convergence across some of the latter (as well as across the entire set) so that the empirical studies are systematic rather than dispersed. (For a useful discussion of theoretical and meta-theoretical integration, see Wagner 2007).

On artificiality and generalizability

Experiments have often been criticized in academic publications (including textbooks) for being artificial and not leading to general conclusions. For the most part, these criticisms have been as persistent as they have been uninformed. Some of the points I make next in response to them have been made before by others, and particularly very eloquently by Berger and Zelditch (1977); Cohen (1989, Chapter 6); Webster and Kervin (1971); Webster and Sell (2014); and Zelditch (2014, 1968). My purpose here is both to highlight those arguments and to present some of my own.

(1) I take 'artificial' to mean that participants find themselves in an unusual (out-of-the ordinary) setting, and that they are asked to make decisions about matters with which they are not fully familiar. In response to that criticism, one should consider that (a) a setting that is unusual to one person may not be so to another,

and (b) experiments are not uniform in the extent to which the setting is out-of-the-ordinary for the participants. For example, note the range represented by two experiments: at one end, one in which respondents are asked to memorize nonsensical phrases and, at the other end, one in which they are instructed to act as a team in solving a valuable task. Finally, making decisions on unfamiliar topics may in fact be a legitimate design-requirement in a study in which the author wishes to explore precisely the extent to which such a setting permits prejudices to emerge and be recorded.

Unless *either* levels of familiarity with the setting are part of the hypothesis that is being tested *or* a specific level interferes with the test (e.g., distracts the participants; precludes them from understanding the instructions), this factor is *not* a requirement of an experimental design.

(2) It is also common to read statements such as 'the experimental results do not generalize.' Sometimes this is referred to as a 'lack of external validity.' The answer to such pronouncements is not as simple as they suggest that it would be, as the matter of 'generalize to what?' is seldom addressed, and the meaning of 'external' is often vague. It is also of key importance to distinguish between (i) random *assignment* to the conditions of an experiment, and (ii) random *sampling* from a population defined in particular terms (e.g., time and space). If a person were to conduct an experiment to investigate how one independent variable affects a dependent variable *for a particular group of participants only*, it would of course make sense to take a random sample of that population.

My interest, as well as that of many other researchers doing experiments, is not in that situation. Rather, we intend to test theoretical hypotheses. Since the latter are formulated in abstract, not concrete terms, it is not possible to take a random sample from the population to which these hypotheses apply.

All research data are of course particular in one respect or another. The issue is then how data relate to abstract ideas. They do so through links that tie observables to abstract terms. For example, let us define 'social status' as a category with two or more levels, each implying different degrees of respect and, in turn, corresponding performance expectations (Berger, Zelditch 1977, p. 34–36). Sex category may then be seen as an instance of social status. Depending on their perceived applicability, status characteristics vary from specific to diffuse; sex category often has a diffuse dimension.

Let us now assume that the results from an experiment have supported a hypothesis proposing that, under specified scope conditions, sex category is a status factor and indeed has affected both assignment of competence and several behavioural, task-related responses such as interpersonal influence. The next question is how to determine the extent of support that the hypothesis has received.

There is no ready-made procedure that could make research findings more general. (Results from a representative sample apply only to the population from which the sample was drawn, unless theoretical work identifies similarities with other populations.) A wider generalization is a task that requires both theoretical thinking and empirical work, and that is achieved through the systematic, often slow, process of replication (see Cohen 1989, Chapters 13–15; Hendrick 1991; Smith 2008). That process involves the following:

(a) A hypothesis becomes more general if findings from well designed and competently conducted studies indicate that its terms can be rephrased at a higher level of abstraction. For example, if the test of a hypothesis about sex-category effects on competence expectations is replicated successfully with results from ethnicity and formal education, one can reformulate the hypothesis to refer more generally to 'status effects.'

(b) A hypothesis also becomes more general if empirical results show that some of its factors can be expanded: for example, that a scope condition can be relaxed or even deleted, or that one or more of the independent variables can assume a larger range of values.

(c) Generalization can also occur if, through replications, the empirical base indicates that some factors that had been treated as test limitations (e.g., country in which the experiments have been conducted and the level of education of the participants) are in fact irrelevant.

(d) The empirical base for a hypothesis can become more general (and stronger) if its key concepts are operationalized in different ways, either within or across studies, and/or the designs are varied. Four of the works listed in Appendices A and B exemplify some of those various means of generalizing and contribute to cumulativeness: Shinnaoui, Narchal 2010 replicated Esses et al. 2006's experiment, while Ward and Masgoret 2007's design and that of Oreopoulos 2011 are comparable in several key respects.

Summary and conclusions

In order to put my enthusiasm for experiments in a larger context, I should mention that, of course, not all such studies are created equal. A poor experiment is still a poor piece of research regardless of the type of design used. When experiments are carefully designed in the context of a theoretical program, operationalize variables successfully, data are presented in an informative way, instructions for replications are provided, the topic is not trivial, ethical guidelines are followed, and the participants are engaged with their task, this methodology is a powerful instrument.

In this article I present highlights from a review of experimental research on immigration topics. Even the relatively small number of nine selected studies from that review serves to illustrate the variety of factors and designs that have been used in this area. I also discuss and address the commonly raised criticisms about experiments in the social sciences, namely concerns about artificiality and generalizability, and make suggestions for further research. My aim is to promote the use of this methodology in the sociology of immigration. In my view, increasing the use of experimental designs in this area will foster theoretical advances and will guide fair social-interventions in immigration matters. Although only a minority of the works identified in the present review deals with groups, there are plenty of ideas in those studies to extend the experiments to more sociological topics.

Acknowledgments

I would like to thank Emma Dewit and Alina Kosel for their expert library-research assistance with the literature review.

References

Berger J., Zelditch M., Jr. (1977). *Status Characteristics and Social Interaction: The Status-Organizing Process.* In: J. Berger, M.H. Fisek, R.Z. Norman, M. Zelditch Jr. (eds.), *Status Characteristics and Social Interaction: An Expectation-States Approach.* New York, NY: Elsevier, p. 3–87.

Berry J.W. (2001). *A Psychology of Immigration.* Journal of Social Issues, 57, p. 615–631.

Cohen B.P. (1980). *The Conditional Nature of Scientific Knowledge*. In: L. Freese (ed.), *Theoretical Methods in Sociology: Seven Essays.* Pittsburgh, PA: The University of Pittsburgh, p. 71–110.

Cohen B.P. (1989). *Developing Sociological Knowledge: Theory and Method,* 2nd ed. Chicago, IL: Nelson-Hall.

Cohen E.G. (1994). *Designing Groupwork: Strategies for the Heterogeneous Classroom*, 2nd ed. New York, NY: Teachers College Press.

Foschi M. (1997). *On Scope Conditions*. Small Group Research, 28, p. 535–555.

Foschi M. (2013). *The Professional Credentials of Immigrants: A Status-and-Expectations Approach.* Advances in Group Processes: A Research Annual, 30, p. 131–155.

Foschi M. (2014). *Hypotheses, Operationalizations, and Manipulation Checks*. In: M. Webster Jr., J. Sell (eds.), *Laboratory Experiments in the Social Sciences*, 2nd ed. San Diego, CA: Elsevier, p. 247–268.

Goar C., Sell, J., Manago B., Melero C., Reidinger B. (2013). *Race and Ethnic Composition of Groups: Experimental Investigations.* Advances in Group Processes: A Research Annual, 30, p. 47–75.

Hendrick C. (1991). *Replications, Strict Replications, and Conceptual Replications: Are They Important?* In: J.W. Neuliep (ed.), *Replication Research in the Social Sciences*. Newbury Park, CA: Sage, p. 41–49.

Jackson M., Cox D.R. (2013). *The Principles of Experimental Design and Their Application in Sociology.* Annual Review of Sociology, 39, p. 27–49.

Magnusson E., Marecek J. (2012). *Gender and Culture in Psychology: Theories and Practices*. Cambridge, UK: Cambridge University Press.

McBurney D.H. (1983). *Experimental Psychology*. Belmont, CA: Wadsworth.

McDermott R. (2006). *Guest Editor's Introduction – Special Issue on Experiments in Political Psychology.* Political Psychology, 27, p. 347–358.

McDermott R. (2014). *Experimental Political Science.* In: M. Webster Jr., J. Sell (eds.), *Laboratory Experiments in the Social Sciences,* 2nd ed. San Diego, CA: Elsevier, p. 295–309.

McKenzie D., Yang, D. (2012). *Experimental Approaches in Migration Studies*. In: C. Vargas-Silva (ed.), *Handbook of Research Methods in Migration*. Cheltenham, UK: Edward Elgar, p. 249–269.

Smith R.B. (2008). *Cumulative Social Inquiry: Transforming Novelty into Innovation*. New York, NY: The Guilford Press.

Thau S., Pitesa M., Pillutla M. (2014). *Experiments in Organizational Behavior.* In: M. Webster Jr., J. Sell (eds.), *Laboratory Experiments in the Social Sciences*, 2nd ed. San Diego, CA: Elsevier, p. 433–447.

Wagner D.G. (2007). *The Limits of Theoretical Integration.* Social Justice Research, 20, 270–287.

Webster M., Jr., Kervin J.B. (1971). *Artificiality in Experimental Sociology.* Canadian Review of Sociology and Anthropology, 8, p. 263–272.

Webster, M., Jr., Rashotte, L.S. (2009). *Fixed Roles and Situated Actions.* Sex Roles, 61, p. 325–337.

Webster M., Jr., Sell J. (2014). *Why Do Experiments?* In: M. Webster Jr., J. Sell (eds.), *Laboratory Experiments in the Social Sciences*, 2nd ed. San Diego, CA: Elsevier, p. 5–21.

Willer D.,Walker H.A. (2007). *Building Experiments/Testing Social Theory.* Stanford, CA: Stanford University Press.

Zelditch M., Jr. (1968). *Can You Really Study an Army in the Laboratory?* In: A. Etzioni (ed.), *A Sociological Reader on Complex Organizations,* 2nd ed. New York, NY: Holt, Rinehart and Winston, p. 528–539.

Zelditch M., Jr. (2014). *Laboratory Experiments in Sociology*. In: M. Webster Jr., J. Sell (eds.), *Laboratory Experiments in the Social Sciences,* 2nd ed. San Diego, CA: Elsevier, p. 183–197.

Appendix A: Immigration experiments discussed in this text

Caprariello P.A., Cuddy A.J.C., Fiske S.T. (2009). *Social Structure Shapes Cultural Stereotypes and Emotions: A Causal Test of the Stereotype Content Model.* Group Processes & Intergroup Relations, 12, p. 147–155.

Jackson L.M., Esses V.M. (2000). *Effects of Perceived Economic Competition on People's Willingness to Help Empower Immigrants.* [Study 1]. Group Processes & Intergroup Relations, 3, p. 419–435.

Jetten J., Wohl M.J.A. (2012). *The Past as a Determinant of the Present: Historical Continuity, Collective Angst, and Opposition to Immigration.* [Studies 1 and 2]. European Journal of Social Psychology, 42, p. 442–450.

Joona P.A., Nekby L. (2012). *Intensive Coaching of New Immigrants: An Evaluation Based on Random Program Assignment.* Scandinavian Journal of Economics, 114, p. 575–600. (*)

Oreopoulos P. (2011). *Why Do Skilled Immigrants Struggle in the Labour Market? A Field Experiment with Thirteen Thousand Résumés.* American Economic Journal: Economic Policy, 3, p. 148–171.

Ramos M.R., Jetten J., Zhang A., Badea C., Iyer A., Cui L., Zhang Y. (2013). *Minority Goals for Interaction with the Majority: Seeking Distance from the Majority and the Effect of Rejection on Identification.* [Studies 2 and 3]. European Journal of Social Psychology, 43, p. 72–83.

Vezzali L., Capozza D., Giovannini D., Stathi S. (2012a). *Improving Implicit and Explicit Intergroup Attitudes Using Imagined Contact: An Experimental Intervention with Elementary School Children.* Group Processes & Intergroup Relations, 15, p. 203–212. (*)

Appendix B. Other immigration-related experiments identified in this search

Appel M. (2012). *Anti-Immigrant Propaganda by Radical Right Parties and the Intellectual Performance of Adolescents.* Political Psychology, 33, p. 483–493. Location: Austria.

Aydin N., Krueger J.I., Frey D., Kastenmüller A., Fischer P. (2014). *Social Exclusion and Xenophobia: Intolerant Attitudes toward Ethnic and Religious Minorities.* [Studies 1, 2, 3 and 4].

Group Processes & Intergroup Relations, 17, p. 371–387. Location: Germany. (*) for all four studies.

Barreto M., Spears R., Ellemers N., Shahinper K. (2003). *Who Wants to Know? The Effect of Audience on Identity Expression among Minority Group Members.* [Studies 1 and 2]. British Journal of Social Psychology, 42, p. 299–318. Location: The Netherlands.

Beaupré M.G., Hess U. (2003). *In My Mind, We All Smile: A Case of In-Group Favoritism.* [Study 3]. Journal of Experimental Social Psychology, 39, p. 371–377. Location: Canada.

Brader T., Valentino, N.A., Suhay E. (2008). *What Triggers Public Opposition to Immigration? Anxiety, Group Cues, and Immigration Threat.* [Studies 1 and 2]. American Journal of Political Science, 52, p. 959–978. Location: United States.

Brambilla M., Ravenna M., Hewstone M. (2012). *Changing Stereotype Content through Mental Imagery: Imagining Intergroup Contact Promotes Stereotype Change.* Group Processes & Intergroup Relations, 15, p. 305–315. Location: Italy. (*)

Bruneau E.G., Saxe R. (2012). *The Power of Being Heard: The Benefits of 'Perspective-Giving' in the Context of Intergroup Conflict.* [Study 1]. Journal of Experimental Social Psychology, 48, p. 855–866. Location: United States. (*)

Chatard A., Selimbegovic L., Konan P., Mugny G. (2008). *Performance Boosts in the Classroom: Stereotype Endorsement and Prejudice Moderate Stereotype Lift.* Journal of Experimental Social Psychology, 44, p. 1421–1424. Location: France.

Costello K., Hodson G. (2011). *Social Dominance-Based Threat Reactions to Immigrants in Need of Assistance.* [Studies 1 and 2]. European Journal of Social Psychology, 41, p. 220–231. Location: Canada.

Costello K., Hodson G. (2013). *Exploring the Roots of Dehumanization: The Role of Animal-Human Similarity in Promoting Immigration Humanization.* [Study 2]. Group Processes & Intergroup Relations, 13, p. 3–22. Location: Canada.

Craig M.A., Richeson J.A. (2014). *Not in My Backyard! Authoritarianism, Social Dominance Orientation, and Support for Strict Immigration Policies at Home and Abroad.* [Studies 1, 2 and 3]. Political Psychology, 35, p. 417–429. Location: United States.

Crawford J.T., Pilanski J.M. (2014). *The Differential Effects of Right-Wing Authoritarianism and Social Dominance Orientation on Political Intolerance.* [Study 2]. Political Psychology, 35, p. 557–576. Location: United States.

Erisen C., Lodge M., Taber C.S. (2014). *Affective Contagion in Effortful Political Thinking.* [Studies 1 and 2]. Political Psychology, 35, p. 187–206. Location: United States.

Esposo S.R., Hornsey M.J., Spoor J.R. (2013). *Shooting the Messenger: Outsiders Critical of Your Group Are Rejected Regardless of Argument Quality.* British Journal of Social Psychology, 52, p. 386–395. Location: Australia.

Esses V.M., Veenvliet S., Hodson G., Mihic L. (2008). *Justice, Morality, and the Dehumanization of Refugees.* [Study 3]. Social Justice Research, 21, p. 4–25. Location: Canada.

Falomir-Pichastor J.M., Frederic N.S. (2013). *The Dark Side of Heterogeneous Ingroup Identities: National Identification, Perceived Threat, and Prejudice against Immigrants.* [Studies 1 and 2]. Journal of Experimental Social Psychology, 49, p. 72–79. Location: Switzerland.

Falomir-Pichastor J.M., Gabarrot F., Mugny G. (2009). *Group Motives in Threatening Contexts: When a Loyalty Conflict Paradoxically Reduces the Influence of an Anti-Discrimination Ingroup Norm.* European Journal of Social Psychology, 39, p. 196–206. Location: Switzerland.

Falomir-Pichastor J.M., Muñoz-Rojas D., Invernizzi F., Mugny G. (2004). *Perceived In-Group Threat as a Factor Moderating the Influence of In-Group Norms on Discrimination against Foreigners.* European Journal of Social Psychology, 34, p. 135–153. [Studies 2 and 3]. Location: Switzerland.

Florack A., Bless A., Piontkowski U. (2003). *When Do People Accept Cultural Diversity? Affect as Determinant.* International Journal of Intercultural Relations, 27, p. 627–640. Location: Germany.

Gabarrot F., Falomir-Pichastor J., Mugny, G. (2009). *Being Similar versus Being Equal: Intergroup Similarity Moderates the Influence of In-Group Norms on Discrimination and Prejudice.* [Studies 1 and 2]. British Journal of Social Psychology, 48, p. 253–273. Location: Switzerland and France.

Gadarian S.K., Albertson B. (2014). *Anxiety, Immigration, and the Search for Information.* Political Psychology, 35, p. 133–164. Location: United States.

Hainmueller J., Hiscox M.J. (2010). *Attitudes towards Highly Skilled and Low-Skilled Immigration: Evidence from a Survey Experiment.* American Political Science Review, 104, p. 61–84. Location: United States.

Harwood J., Paolini S., Joyce N., Rubin M., Arroyo A. (2011). *Secondary Transfer Effects from Imagined Contact: Group Similarity Affects the Generalization Gradient.* British Journal of Social Psychology, 50, p. 180–189. Location: United States. (*)

Horverak J.G., Sandal G.M., Pallesen S., Timmerman M.E. (2013). *Hiring Rankings of Immigrant Job Applicants: Immigrants' Acculturation Strategies and Managers' Personality Trait Perception.* Journal of International Migration and Integration, 14, p. 493–510. Location: Norway.

Konan P.N., Chatard A., Selimbegovic L., Mugny G., Moraru A. (2011). *Deflecting Stereotype Threat through Downward Comparison: When Comparison with Immigrants Boosts Performance of Stigmatized Native Students.* [Studies 1 and 2]. Social Justice Research, 24, p. 191–205. Location: Switzerland. (*) for Study 1 only.

Lima-Nunes A., Pereira C.R., Correia I. (2013). *Restricting the Scope of Justice to Justify Discrimination: The Role Played by Justice Perceptions in Discrimination against Immigrants.* European Journal of Social Psychology, 43, p. 627–636. [Studies 2 and 3]. Location: Portugal.

Lucas T., Rudolf C., Zhdanova L., Barkho E., Weidner N. (2014). *Distributive Justice for Others, Collective Angst, and Support for Exclusion of Immigrants.* [Study 2]. Political Psychology, 35, p. 775–793. Location: United States.

Matera C., Stefanile C., Brown R. (2011). *The Role of Immigrant Acculturation Preferences and Generational Status in Determining Majority Intergroup Attitudes.* [Studies 1 and 2]. Journal of Experimental Social Psychology, 47, p. 776–785. Location: Italy.

Molinsky A. (2005). *Language Fluency and the Evaluation of Cultural Faux Pas: Russians Interviewing for Jobs in the United States.* Social Psychology Quarterly, 68, p. 103–120. Location: United States.

Motyl M., Hart J., Pyszczynski T., Weise D., Maxfield M., Siedel A. (2011). *Subtle Priming of Shared Human Experiences Eliminates Threat-Induced Negativity toward Arabs, Immigrants, and Peace-Making.* [Study 2]. Journal of Experimental Social Psychology, 47, p. 1179–1184. Location: United States.

Newman B.J., Hartman T.K., Taber C.S. (2012). *Foreign Language Exposure, Cultural Threat, and Opposition to Immigration.* [Studies 2 and 3]. Political Psychology, 33, p. 635–657. Location: United States. (*) for Study 2 only.

Newman B.J., Hartman T.K., Taber C.S. (2014). *Social Dominance and the Cultural Politics of Immigration.* [Study 1]. Political Psychology, 35, p. 165–186. Location: United States.

Oyamot C.M., Jr., Fisher E.L., Deason G., Borgida E. (2012). *Attitudes toward Immigrants: The Interactive Role of the Authoritarian Predisposition, Social Norms, and Humanitarian Values.* Journal of Experimental Social Psychology, 48, p. 97–105. Location: United States.

Pagotto L., Visintin E.P., De Iorio G., Voci A. (2013). *Imagined Intergroup Contact Promotes Cooperation through Outgroup Trust.* Group Processes & Intergroup Relations, 16, p. 209–216. Location: Italy. (*)

Philippot P., Yabar Y. (2005). *Stereotyping and Action Tendencies Attribution as a Function of Available Emotional Information.* European Journal of Social Psychology, 35, p. 517–536. [Studies 1 and 2]. Location: Belgium.

Prati F., Crisp R.J., Rubini M. (2015). *Counter-Stereotypes Reduce Emotional Intergroup Bias by Eliciting Surprise in the Face of Unexpected Category Combinations*. [Studies 2 and 3]. Journal of Experimental Social Psychology, 61, p. 31–43. Location: Italy.

Rothschild Z.K., Landau M.J., Molina L.E., Branscombe N.R., Sullivan D. (2013). *Displacing Blame Over the Ingroup's Harming of a Disadvantaged Group Can Fuel Moral Outrage at a Third-Party Scapegoat.* Journal of Experimental Social Psychology, 49, p. 898–906. Location: United States.

Rubin M., Paolini S., Crisp R.J. (2010). *A Processing Fluency Explanation of Bias against Immigrants.* [Studies 1 and 2]. Journal of Experimental Social Psychology, 46, p. 21–28. Location: Australia.

Scheepers D., Saguy T., Dovidio J.F., Gaertner S.L. (2014). *A Shared Dual Identity Promotes a Cardiovascular Challenge Response During Interethnic Interactions.* [Studies 1 and 2]. Group Processes & Intergroup Relations, 17, p. 324–341. Location: The Netherlands. (*) for both studies.

Shinnaoui D., Narchal R. (2010). *Brain Gain to Brain Waste: Individual Biases, Prejudice, and Discounting of Migrant Skills.* Journal of International Migration and Integration, 11, p. 423–437. Location: Australia.

Siem B., Lotz-Schmitt K., Stürmer S. (2014). *To Help or Not to Help an Outgroup Member: The Role of the Target's Individual Attributes in Resolving Potential Helpers' Motivational Conflict .* [Studies 1, 2 and 3]. European Journal of Social Psychology, 44, p. 297–312. Location: Germany. (*) for all three studies.

Smeekes A., Verkuyten M. (2014). *When National Culture Is Disrupted: Cultural Continuity and Resistance to Muslim Immigrants.* [Study 3]. Group Processes & Intergroup Relations, 17, p. 45–66. Location: The Netherlands.

Smeekes A., Verkuyten M., Poppe E. (2011). *Mobilizing Opposition towards Muslim Immigrants: National Identification and the Representation of National History.* [Studies 1, 2 and 3]. British Journal of Social Psychology, 50, p. 265–280. Location: The Netherlands.

Smeekes A., Verkuyten M., Martinovic B. (2015). *Longing for the Country' s Good Old Days: National Nostalgia, Autochtony Beliefs, and Opposition to Muslim Expressive Rights.* [Studies 3 and 4]. British Journal of Social Psychology, 54, p. 561–580. Location: The Netherlands.

Smith L.G.E., Postmes T. (2010). *The Power of Talk: Developing Discriminatory Group Norms through Discussion.* [Studies 1, 2, 3 and 4]. British Journal of Social Psychology, 50, p. 1193–215. Location: Not specified – participants were described as 'British students.' (*) – for Studies 1, 2 and 3.

Stroessner S.J., Scholer A.A., Marx D.M., Weisz B.M. (2015). *When Threat Matters: Self-Regulation, Threat Salience, and Stereotyping.* [Study 2]. Journal of Experimental Social Psychology, 59, p. 77–89. Location: United States.

Sweetman J., Spears R., Livingstone A.G., Manstead A.S.R. (2013). *Admiration Regulates Social Hierarchy: Antecedents, Dispositions, and Effects on Intergroup Behavior.* [Studies 2 and 3]. Journal of Experimental Social Psychology, 49, p. 534–542. Location: Wales.

Tauber S., van Zomeren M. (2013). *Outrage towards Whom? Threats to Moral Group Status Impede Striving to Improve via Out-Group Directed Outrage.* [Study 2]. European Journal of Social Psychology, 43, p. 149–159. Location: The Netherlands.

Thomsen L., Green E.G.T., Sidanius J. (2008). *We Will Hunt Them down: How Social Dominance Orientation and Right-Wing Authoritarianism Fuel Ethnic Persecution of Immigrants in Fundamentally Different Ways.* [Studies 1 and 2]. Journal of Experimental Social Psychology, 44, p. 1455–1464. Location: United States and Switzerland.

Valentino N.A., Brader T., Jardina A.E. (2013). *Immigration Opposition among U.S. Whites: Ethnocentrism or Media Priming of Attitudes About Latinos?* Political Psychology, 34, p. 149–166. Location: United States.

Van Tilburg W.A.P., Igou E.R. (2011). *On the Meaningfulness of Existence: When Life Salience Boosts Adherence to Worldviews.* [Study 1B]. European Journal of Social Psychology 41, p. 740–750. Location: The Netherlands.

Verkuyten M. (2004). *Emotional Reactions to and Support for Immigrant Policies: Attributed Responsibilities to Categories of Asylum Seekers.* [Study 2]. Social Justice Research, 17, p. 293–314. Location: The Netherlands.

Verkuyten M. (2005). *Immigration Discourses and their Impact on Multiculturalism: A Discursive and Experimental Study.* [Study 2]. British Journal of Social Psychology, 44, p. 223–240. Location: The Netherlands.

Vezzali L., Capozza D., Stathi S., Giovannini D. (2012b). *Increasing Outgroup Trust, Reducing Infrahumanization, and Enhancing Future Contact Intention via Imagined Intergroup Contact.* Journal of Experimental Social Psychology, 48, p. 437–440. Location: Italy. (*)

Vezzali L., Stathi S., Giovannini D., Capozza D., Visintin E.P. (2015). '*And the Best Essay Is …' Extended Contact and Cross-Group Friendships at School.* British Journal of Social Psychology, 54, p. 601–615. Location: Italy. (*)

Ward C., Masgoret A.M. (2007). *Immigrant Entry into the Work Force: A Research Note from New Zealand.* International Journal of Intercultural Relations, 31, p. 525–530. Location: New Zealand.

Wiley S., Figueroa J., Lauricella T. (2014). *When Does Dual Identity Predict Protest? The Moderating Roles of Anti-Immigrant Policies and Opinion-Based Group Identity.* European Journal of Social Psychology, 44, p. 209–215. Location: United States.

Wirtz C., Doosje B.J. (2013). *Reactions to Threatening Critical Messages from Minority Group Members with Shared or Distinct Group Identities.* [Studies 1, 2, 3 and 4]. European Journal of Social Psychology, 43, p. 50–61. Location: The Netherlands.

Zagefka H., Tip L.K., González R., Brown R., Cinnirella M. (2012). *Predictors of Majority Members'Acculturation Preferences: Experimental Evidence.* Journal of Experimental Social Psychology, 48, p. 654–659. Location: Not specified – participants self-reported to be of 'British nationality.'

Zárate M.A., Garcia B., Garza A.A., Hitlan R.T. (2004). *Cultural Threat and Perceived Realistic Group Conflict as Dual Predictors of Prejudice.* Journal of Experimental Social Psychology, 40, p. 99–105. Location: United States.

Zárate M.A., Shaw M., Marquez J.A., Biagas D., Jr. (2012). *Cultural Inertia: The Effects of Cultural Change on Intergroup Relations and the Self-Concept.* [Studies 1, 2 and 3]. Journal of Experimental Social Psychology, 48, p. 634–645. Location: United States.

Co wniosły eksperymenty do socjologicznych badań imigracji

W badaniach nad imigracją eksperymenty nie było powszechną praktyką. W artykule tym przedstawiam argumenty uzasadniające potrzebę wykorzystania tej metodologii w socjologii, relacjonuję wyniki swojego przeglądu prac poświęconych imigracji, w których metoda ta była stosowana, omawiam pożytki z tych prac

oraz wskazuję kierunki przyszłych badań eksperymentalnych w tej dziedzinie. Opisuję także i krytykuję często wyrażane obawy o zasadność stosowania eksperymentów w naukach społecznych, mianowicie przeciwstawiam się: (1) zarzutowi „sztuczności" założonej w samym planie badawczym; (2) przekonaniu o ograniczonej możliwości uogólniania wyników badań. W artykule podana została także pełna lista pozycji bibliograficznych wykorzystanych w przeglądzie.

Słowa kluczowe: eksperymenty, imigracja, sztuczność, uogólnialność

FOLIA 202

Annales Universitatis Paedagogicae Cracoviensis

Studia Sociologica VIII (2016), vol. 1, p. 62–80
ISSN 2081–6642

Murray Webster, Jr.
University of North Carolina – Charlotte, USA

Jane Sell
Texas A & M University, USA

The Present Status and Future Prospects of Experiments in the Social Sciences

Abstract

Social Science experiments appeared in psychology at the University of Leipzig in the 1880s, and natural setting and laboratory experiments appeared in sociology four or five decades later. Experiments are a particular kind of research design involving control of independent variables before measurement of dependent variables. While all designs are subject to confounding factors, random allocation to conditions is generally a satisfactory protection against them. Strongly instantiating variables and pretesting all operations are essential. Power assessments are equally helpful. We trace developments in a standardized design that has been widely used to study status and expectation state processes, including improvements in operations with video and computers, and new ways to create interaction variables. Some new designs are being developed to study interrelations of vocal accommodation and group position. Factorial vignettes are a technique for introducing experimental control outside of a laboratory, permitting rapid collection of large amounts of data. Virtual reality equipment and computer simulations similar to those used for drivers' education and flight training show promise for experimental use but they have not yet been used for this purpose. Finally, we consider some misunderstandings about experimental research that may impede more general use of this methodology, and suggest some corrections for the misunderstandings.

Key words: experimental design, misunderstandings regarding experiments, status, expectations, big data

Introduction

Compared with the physical sciences, experiments came late to the social sciences, but of course the social sciences themselves are much newer. All of them began in the last two decades of the 19th century. Psychology became a separate discipline in 1879 at the Institute for Experimental Psychology founded by Wilhelm Wundt at the University of Leipzig. A few years later G. Stanley Hall, who had studied with Wundt, founded an experimental laboratory at The Johns Hopkins University in Baltimore. The University of Chicago established the first U.S. department of sociology in 1892, but experimental research in sociology appeared only several decades later. Political Science, Economics, Communications (or Speech) departments appeared in the U.S. about the same time as sociology, and other social sciences came on the scene in succeeding decades. None of them other than Psychology used

experimental research until decades later. Back then, interest focused on large-scale problems, including social cohesion, bureaucracy, societal types, international migration, the growth of states, and types of leadership. While experiments can be used to reflect on all of those issues, most social scientists at the time thought of experiments as applicable only to individuals and small groups.

Psychologists have used experiments throughout most their history, at least as long ago as in Thorndike's (1905) and Watson's (1913) laboratories. The best known early psychological experiments are probably the studies in classical conditioning begun in the 1890s by Ivan P. Pavlov who trained in biology and medicine, where experiments were well established.[1]

In the U.S., naturalistic studies of small groups including families (Davis 1929; Bernard 1933; Burgess, Cottrell 1939; Terman 1938) and adolescent gangs (Thrasher 1927; Whyte 1955) began to appear, and by the 1930s and 1940s a few laboratory experiments were reported (Sherif 1936; Asch 1948, 1951; Schachter 1951).[2] The famous studies at the Hawthorne Works of the Western Electric Company (west of downtown Chicago) done from 1927 to 1932 (Roethlisberger, Dixon 1939) probably were the earliest social experiments to employ some experimental control.[3]

Other social sciences – political science, communications, economics, and a few branches of anthropology – also have found experimental research valuable in study of phenomena. Sometimes experiments are used to address enduring theoretical issues such as conformity (Asch 1951; Cohen 1963). Other times, experimental methods are appropriate to study important topics from new theoretical developments, such as the growth of rational choice theories in political science (Axelrod

[1] Pavlov's specialization in the digestive system may partially account for his experimental designs and research approach; he received a Nobel Prize in 1904 for his work on the digestive system. The opus on learning is Conditioned Reflexes (1927). The unconditioned stimulus in the first experiments was not a bell but a metronome. Later experiments used a buzzer, a flash of light, a rotating object, an organ tone, bubbling or crackling sounds, and different tones of a whistle (Pavlov 1927: esp. Lectures VII-XIV). A biographer reports that Pavlov disliked the field of psychology (Babkin 1949: 276-277), but his objections seem to be directed against interpreting animal behaviour as the result of cognitive processes rather than against the discipline as a whole. Watson (1913a, b) and Skinner (1953) later developed this view as Behaviourism. Pavlov and Watson seem to reject attempts to study subjective thought in animals, although not in humans; Skinner concluded that the study of subjective thought is irrelevant for humans as well.

[2] Roseborough (1953) catalogues and classifies experimental studies of small groups as far back as the 1920s, however the studies that she lists from that decade are comparisons of teaching methods in schools.

[3] Hawthorne manufactured equipment for Bell Telephone, including home telephones and many other kinds of devices, from 1905 to 1982. The management was unusually enlightened for its day, and instituted many programs to improve working conditions. In the program cited, investigators tried increasing illumination and found that productivity increased. However, they also found that decreasing the light level increased productivity. The conclusion, now known as 'the Hawthorne Effect,' is that paying attention to workers was the actual independent variable that led to increased production.

1984), or the recent interest in non-maximizing behaviour in economics (e.g., Ariely 2008, Thaler 2015).

The event that most boosted the visibility of experimental research in sociology was the establishment in 1946 of the Laboratory for Social Relations at Harvard University under the direction of Talcott Parsons, Samuel Stouffer, and Robert Freed Bales (Bales 1950, 1999; Strodtbeck 1984). Rather than studying naturally occurring groups, Bales composed *ad hoc* problem solving groups of undergraduate students. He developed techniques for observing and coding interaction – the famous 12-category system – and was the first researcher to regularly use a one-way mirror to remove observers from the interaction situation. The work was mostly observational rather than truly experimental, as Bales only rarely intervened to control independent variables as, for instance, Sherif and Asch had done. However that early laboratory research helped establish laboratory methods and experiments as legitimate research designs in sociology.

Structure of experiments[4]

Just as not every argument meets the definition of a *theory*, not every research design is an *experiment*. To clarify how experiments function to test ideas, it will be helpful to share a definition. All research designs involve *independent* and *dependent* variables, and an investigator looks for relations between them. Experiments have a unique temporal ordering, however.

> An *experiment* is a research design in which the levels of independent variables are controlled *before* measurements are collected on the dependent variable.

That definition has two parts. First is the idea of control. In an experiment, the investigator sets one or more values for independent variables. Of course many designs involve *statistical* control of independent variables, but an experimenter creates their levels through intervening in the situation. Second, the fact that independent variables are controlled before data collection – not afterwards, as in a survey, not simultaneously, as in some structured observational studies – is unique to experiments.

Half a century ago, Campbell and Stanley (1963, 1966) analysed a large variety of research designs to identify potential weaknesses in them for making inferences about relations of independent and dependent variables. They distinguished threats to *internal validity* and threats to *external validity*; they identified eight threats to internal validity and four more threats to external validity. Lack of internal validity means that independent and dependent variables actually do not covary as they appear to do in a study. Lack of external validity means that an observed covariance is unique to the groups or situation studied and would not appear elsewhere. Factors that might covary with the independent variables of interest are called 'confounding factors' or 'confounds,' and the purpose of a good experimental design is to find

[4] Issues in this section and the following are analysed in detail in Webster and Sell (eds.) (2014).

ways to eliminate their effects, or at least to measure their impact on the dependent variables, so that the true effect of the independent variables can be estimated.

To deal with the twelve confounding factors that they identified, Campbell and Stanley then described research design modifications, essentially, adding control conditions, to assess the magnitude of the confounds, and to rule them out of other conditions. Their most elaborate design has 12 conditions and requires random assignment to the conditions. Perhaps because Campbell and Stanley were writing about educational research they could envision designs with as many as 12 different groups –which could be implemented in 12 different classrooms or in 12 different schools – to assess and compensate for all of the confounding factors that they identified.

Fortunately, *all* of the threats to validity that Campbell and Stanley discussed are controlled by relatively simple experimental designs. The first one below is a two-condition experiment, and the second is a four-condition extension of it. The letters are identified as follows. R = random assignment to conditions; E = an event or the independent variable, and O = observation, or measuring the dependent variable.

Condition 1: R E O_1

Condition 2: R O_2

In words, an experimenter randomly assigns individuals to condition 1 or to condition 2. Those in condition 1 experience an event (E) – they experience an independent variable – and some dependent variable is measured (O). Those in condition 2 do not experience E, and the same dependent variable is measured for them. If $O_1 \neq O_2$, the experimenter can conclude that E and O are related.[5] If $O_1 = O_2$, of course, the experimenter fails to reject a null hypothesis that E and O are unrelated.[6]

In another design, commonly used in theoretically based experimental research, there are *k* conditions, or *k* experimental groups which correspond to *k* values (levels) of one independent variable or *k* combinations of values of two or more independent variables (factors). Then the goal is not to compare predictions to null hypotheses of 'no effect,' but rather to assess the effects of different levels of a single independent variable or the effects of particular factors alone and the effect of their interaction.

When an experimenter is working with theoretically derived predictions, particularly when the basic theory has been confirmed previously, there is no need for the control group Condition 2 of the first design.

---

[5] Empirically, if an independent variable is *sufficient* to produce a specified change in a dependent variable and also it precedes the change we say that it causes the change. Theoretically, the more modest claim of sufficiency (without causality) is appropriate.

[6] An elaboration of this design to measure levels of the dependent variable in both groups before E would, of course, assure an investigator that the two groups were initially equivalent. The elaboration might be preferred if an investigator needed to know how much effect testing had on the outcome. However, random assignment to conditions assures that the two groups are equivalent, and any effects of testing will affect the experimental and control groups equally. So long as the investigator is not particularly interested in testing but is interested in the effect of treatment (E), there is no need for the more elaborate design.

For all experimental designs, random assignment of individuals to conditions is essential. What randomization does is spread all of the unmeasured and unknown factors – the confounds – that might affect the dependent variable evenly across conditions of the experiment. This, in effect, eliminates an infinite number of confounds or alternative reasons for the outcomes. Those factors become part of the background, constant effects present in all conditions.

Confounding factors have the effect of increasing variance within conditions, making it harder to see actual differences that may exist across conditions. Because they are present in all conditions, and also because they are not of theoretical interest, it is important that confounding factors be minimized. In particular, they should not be stronger in affecting the dependent variable(s) than the intended factors. In other words, experimental designs should be *strong*.

Strong experimental designs

While an experimenter does not, by definition, know just how powerful the confounding factors will be, he or she is wise to take steps to give the theoretically important factors the best possible chances to affect the dependent variables(s). There are two general design features that help. First, an experimenter should do everything possible to remove confounds from the situation. Second, s/he should create the independent variables as strongly as possible.

To reduce the number of confounds and to minimize their effects, one should begin by understanding their nature. By definition, confounds are naturally unwanted, but, more importantly, in most cases they are unknown. When an experimenter knows for sure that certain confounding factors affect a dependent variable, those will always be eliminated from the design. Many social outcomes are affected, for instance, by friendship ties; thus, for most experiments it is wise to compose groups of unacquainted individuals.

What about unknown confounds, factors that can affect outcomes but which the experimenter has not recognized? The best approach here is to make the experimental situation as clean and simple as possible. The goal should never be to simulate a realistic situation because such situations contain multitudes of unknown confounds. Rather, the goal is to develop a highly simplified experimental situation, one containing *all* and *only* the required independent variables. Including some feature because it adds to 'realism,' is almost always a mistake because that means activating unknown factors that might well affect outcomes. An experimental group should not remind participants of their classroom, or of their co-workers. It may have certain abstract features in common with classrooms or business offices, but it should not attempt to re-create any real situation that a participant might have experienced. An experimenter simply cannot know those experiences or how they might affect behaviour, so it is wise to remove as many cues to actual situations as possible. A laboratory is a special place, part of the real world but not part of any actual situation that a participant may have experienced.

On the other side, an experimenter should attempt to make the independent variable(s) as strong as possible so that they stand out against the background of the mixed extraneous factors that randomization deals with. Three practices that help with that goal are obvious instantiation, pretesting, and post-session assessment.

Strong instantiation. 'Instantiation' means creating a concrete instance of the abstract concepts in a theory or in a hypothesis, and it should be done as clearly and as powerfully as possible. Subtlety is out of place in experimental design. Everything that participants should know for the research must be as clear and evident as possible. Weak instantiation of independent variables risks producing high variance within conditions and small overall difference across conditions.[7] The reason is that if the independent variable(s) are weak, they may not overcome effects of the background noise that random assignment has spread evenly across conditions.

A surprisingly common error in experimental design is subtlety. We repeat: subtlety, suggestions, and insinuations are out of place in experiments. For instance, if it is important for a participant to know the gender of another interactant who is unseen, the gender variable should be created strongly. Do not rely simply on giving the partner a gender-typical name. In addition, provide some gender-typical hobbies, and directly tell the participant the gender of the purported partner. Of course if a photo or a video can be used, that is helpful too, but it still needs strengthening from other information. The goal is to present a fully-realized woman or man, so that a participant can easily imagine and remember the social being that the experimenter needs to create. Of course, the experimenter needs to make sure that all receive the *same* name, image, etc. rather than different ones. This eliminates variation that could occur through slight differences in implementation of the independent variable. In this specific case of gender, it is also important to consider the participants' characteristics as well. So, for example, the relationship created by a female white partner to a male white participant will be different from that created by a female white partner to a female black participant.

Every important instruction should be repeated three times. Even if the information is presented clearly and strongly, somebody might miss it the first time. Repeat, and repeat again. Experimental participants are quite willing to hear the same information more than once, and an experimenter increases the chances that everyone will eventually get the information when it has been repeated. An experimenter cannot count on all participants being fully attentive 100% of the time, so if something is important, it should be repeated. Do participants find that repetition tedious? Perhaps some do, but that very rarely rises to the level that they resent it. Most people do not object to repetition – just as most people do not even realize that in this paragraph we have repeated this advice three times (Walker 2014).

For some studies, it is possible to give participants 'quizzes' to ensure that they understand the experimental instructions. This is routinely done in experimental economics studies, for instance, when participants must understand payoff

[7] High variance and small modal differences make it likely that statistical tests will fail to reject a null hypothesis of 'no difference' between conditions.

matrices. Participants do not continue the experiment until their answers show an adequate level of understanding.

Pretesting. Pretesting means testing all the materials such as questionnaires and instructions, and all the operations such as tasks for the participants and data collection methods. Experiments usually are simpler than natural settings but they still are complex social situations. It is usually impossible for an experimenter to be confident that everything in the design will accurately reflect what is needed for that experiment, and of course any interaction effects only increase complexity. Thus pretesting is essential (Rashotte, Webster, Whitmeyer 2005).

An experimental pretest resembles the dress rehearsal of a theatrical performance. All of the procedures of the experiment are in place, and in the best judgment of the experimenter, they fit together and will be understood as intended. A pretest is a try-out. As in a dress rehearsal, every element of an experiment is scrutinized as it occurs and afterwards. When it is feasible, an experimenter may ask other knowledgeable researchers to watch some pretest sessions and watch for misunderstandings or other unanticipated problems. After a pretest session, questionnaire and interview data should be collected from participants to learn how they perceived the situation and whether they understood and remembered the independent variables. If the pretest includes a post-session interview, as it usually should, then it is possible to enlist participants as collaborators by asking them what parts of the experiment they found confusing or difficult. Pretests also are the place to learn about unexpected interactions, such as things about the experience or data collection methods that trigger memories or emotional responses that affect behaviour.

Problems identified during a pretest can be corrected through modification of the design and operations, at which point further pretests can be conducted. The general reason for pretesting is that nobody can anticipate in complete detail how any social situation is going to be perceived or its effects on others' behaviour. Of course if the happy outcome of pretesting is that everything works more or less as intended and no changes are needed, then pretest sessions can be treated as part of the experimental sessions and their data included.

Pretests also help to assess statistical power, or the ability of the experiment to detect effects. If there is a history of experiments with similar design and using the same dependent variables, then power for a new experiment can be estimated from existing data. However, it is frequently the case that a newly designed experiment will have new dependent variables. When this is the case, pretests are the place to get power estimates. Power estimates should then be made on the variable for which variation is expected to be the lowest, as that will provide the most conservative estimates, usually, requiring the largest N (Compton et al., 2012).

Post-session assessment. Even with careful pretesting, it is wise to remember that pretest results apply to general cases, while every individual participating in an experiment is unique. Any participant might have misunderstood or forgotten some important design feature of the experiment and thus an experimenter needs measures of success at meeting initial conditions. These may include questionnaires or individual interviewing, or, ideally, both. It is good practice to ask experimental participants to tell what they remember of the experiment's initial conditions and

independent variables. If, for instance, it is important for them to be task-focused during interaction, an experimenter could ask them whether they thought it was important to do well at the task, or ask them to describe how their interest in doing well varied during the course of the experiment.

When questionnaires and interviews both are used, good practice is to repeat some of the queries. An interviewer can review questionnaire answers before beginning the interview. If an interview question gets a different answer than the questionnaire did, the interviewer can and should ask for clarification. Participants do not always think carefully before answering questionnaires. If particular information is important to the experiment, an interviewer should persist until s/he understands just what went through a participant's mind and how he or she interpreted it.

Deception

At times, experimental designs may require deception. In this context, deception means that participants are deliberately misled about some component of the study. For example, they may receive false information about how they performed on a test or they may be told characteristics or behaviours of their study partners that are false. As much as possible, deception should be minimized, but it is still controversial. There are two points of controversy. One is the possibility that a participant might be harmed by the deception. For instance, suppose a person receives a (false) low score on a laboratory test and then feels badly about themselves. In the case of false information, the post-experimental interview should include extensive debriefing in which participants are clearly told about the false information and the reason for it.

A second potential problem of deception is raised most often by economists. This is the possibility that if experimentalists use deception in a study, participants will no longer believe what they are told in future studies (Hertwig, Ortmann 2008). From this point of view, deception by one experimenter taints participants for all other experimenters.

We disagree. In our experience, when study participants are treated with respect, they respect the studies and the experimenter. Participants become partners with the researchers in not spoiling future experience for any friends by telling them about critical features of the experience. Perhaps the strongest reason for deception is that some important theoretical questions simply cannot be answered without deception (Cook, Yamagishi 2008; Sell 2008).

Compensation (payment)

Compensation is often given to participants to encourage them to volunteer for studies. Compensation can vary significantly based upon the population from which the participants are drawn. Students might be compensated by course credit or by money. For some experiments, the money earned in the study through bargaining or

solving problems is compensation for the study. The important aspects of compensation are that participants know the payment type and amount beforehand.

However, if the incentive is so large that it makes it exceedingly difficult for a participant to refuse, it is coercive. Examples might include recruiting homeless people to participant in an experiment for which they would earn €1000.

Some developments in experiments

Expectation states designs

Joseph Berger, who had studied with Bales, developed an experimental design to study interaction sources and consequences of performance expectation states in task groups. From studying early Bales groups, Berger developed a conception of interaction having four components: action opportunities, performance outputs, unit evaluations of performances, and influence that guided his development of an experimental design to study expectation and status processes (Berger 2014). Berger's design has been used, with slight modification, up to the present for hundreds of experiments. Summaries of the research programs are available in Berger, Wagner, Webster (2014) and Webster, Walker (2016).

The design for status-expectations experiments has progressed through three phases. The initial phase, from about 1960 to 1976 allowed for pairs of individuals (occasionally 3 or 4) to receive performance information in phase 1 and to register choices and influence in phase 2. The second phase, beginning in about 1976, presented phase 1 independent variable information on video, permitting relatively easy creation of purported partners having controllable status characteristics. Scientifically, the video design permitted uniform presentation of independent variables. Operationally, it reduced fatigue errors on the part of confederates and experimenters.

The third phase, since the turn of the century, controls interaction patterns and collects data through computers. This reduces operational errors in data recording and allows for creating new patterns of independent variables. Those include controlling behaviour of participants, as well as the apparent behaviour of their partners to study effects of fairly complex interaction patterns.

Other designs for status and expectations research

The very success of Berger's experimental situation for studies of status and expectations may have impeded the search for alternative standard designs. On the one hand that may have been a virtue, facilitating the growth of cumulative knowledge as results from diverse experiments, with diverse populations, in different countries became directly comparable. On the other hand, additional designs are desirable to extend the range of applications of the theoretical ideas and to permit addressing new research questions.

Open interaction designs, including discussion groups and interaction through computers, have been used for basic and applied research (Goar, Sell 2005; Walker, Doerer, Webster 2015; Goar et al., 2013; Shelly, Shelly 2009). Many of these group interactions are also coded for other power and prestige components, including

directives, agreements and disagreements. When confederates are used, the manner of communicating can be controlled, as whether a confederate is hesitant and deferential or nondeferential (Ridgeway, Erickson 2000; Ridgeway et. al. 1998; Ridgeway, Correll 2006; Ridgeway et al., 2009).

Another new design, developed by Gregory and his colleagues (Gregory, Webster S. 1996; Kalkhoff, Gregory 2008) measures vocal frequencies in open interaction. This relies on an attribute of speech, only partially understood, that seems to reflect status, expectations and group structure. The attribute is the production of certain frequencies during speaking.

Speaking employs a range of vocal frequencies, as in the common observations that male speech uses, on average, lower frequencies than female speech, and opera singers can produce a wider range of frequencies than pop singers. Speech also produces frequencies below the range used to form words, somewhere in the spectrum below 300 Hz. While we can hear sound in that range, in speaking it does not function in word production; thus it has been called 'sub-vocal.'

Frequency variation occurs as a function of social situations, as well as across individuals. Gregory and Stephen Webster (1996) have found, using recordings of U.S. Presidential debates, that in most cases the candidate who was adjudged the winner of each debate by other criteria had adjusted his sub-vocal sounds less than the loser of the debate. Gallagher et al. (2005) successfully used this measure for studying status in simulated medical interviews. However, this technique has not yet been adapted for controlled experimental designs.

Many questions about vocal accommodation remain; these include:

• Does vocal accommodation reflect status inequality or dominance? In the Presidential debates, the winner/loser could reflect either type of inequality. Theoretically, however, status operates very differently from dominance (Ridgeway, Berger 1986). Among the more important differences, status inequality is consensual, while dominance inequality is conflictual.

• What is the best way to conceptualize sub-vocal production? Is it, for instance, a status cue, as described by Berger et al. (1986) and incorporated in explicit theory by Fisek et al. (2005)?

• Can an individual learn to control the production of sub-vocal frequencies? If so, that might complicate using it to measure status, but would offer a new intervention technique to overcome harmful effects of other status characteristics.

Factorial vignette experiments

Respondents are presented with a vignette, a short paragraph describing individuals and situations in which factors are systematically varied. For instance, a vignette used to assess perceptions of fair earnings might describe a target individual with specified gender, educational level, occupational level, and income (Jasso 2003, 2006; Jasso, Webster 1999). Each factor – gender, education, occupation, and income – could be systematically varied. Vignette information constitutes independent variables in this design, and some questionnaire response, such as perceived degree of fairness, constitutes the dependent variable. One way to analyse

data treats the independent variables in the vignettes as regressors to estimate their effects on the dependent variable. This technique was pioneered by Peter H. Rossi (1979; Rossi, Anderson 1982) and developed by Guillermina Jasso (2006).

Vignettes may be administered in classrooms, laboratories, through the postal service, or online. They gather data much faster than in a laboratory experiment, yet still control the independent variable. Given the ability to collect large numbers of responses quickly, vignettes also can study effects of a large number of independent variables.

At the same time, vignette studies have certain weaknesses. Attention paid to the vignette before responding is hard to monitor and probably varies with factors including distractions in the setting where the data are collected. Assessing whether a respondent takes the task seriously or merely provides spurious data is then harder than it is in the laboratory where individual post-session interviews are standard.

The main potential problem with vignette studies is probably that they rely on respondents' ability and willingness to reproduce imaginatively the situation described. If a vignette asks me to rate the fairness of, say, €650 per week for a welder with a high school diploma, am I able to imagine that situation? If not, then my fairness rating is probably influenced by something other than the intended independent variables.

In a vignette study Jasso and Webster (1999) found that college student respondents felt, overall, that women should be paid slightly more than comparably accomplished men. Those authors speculated that the finding might reveal a decline in the significance of gender among the young, or their lack of experience in the adult world of work. In an ingenious study with a large sample of German college and employed adult respondents, Carsten Sauer (2014) showed that a respondent's own experience strongly affected fairness judgments. Respondents from formerly socialist eastern Germany saw a small gender gap in incomes to be fair, while respondents from capitalist western Germany saw a large gap as fair. Thus the Jasso-Webster finding was probably due to respondents' experience with women students being paid the same or slightly more than men, and not to a decline in the significance of gender for their fair earnings' judgments.

Other examples of vignette style studies include Martha Foschi and her colleagues' studies of hiring decisions. In these studies, participants are randomly assigned to read different resumes of people applying for a particular job. The gender (Foschi et al. 1994; Foschi, Valenzuela 2008) or other statuses of the applicant are varied so that some of the participants read one version while others read other versions.

Vignette designs continue to improve as investigators learn which factors are easy for respondents to imagine and which are more difficult. Investigators also may wish to adopt techniques from survey research such as repeating items to check the reliability of responses, and placing a particular item early or late in a sequence to assess order effects and effects of fatigue.

Future experimental designs

Improvements in computer capacities and connectivity should increase opportunities for imaginative social scientists. We consider three such uses, two for basic research and one for either basic or applied research.

Virtual reality equipment

Laboratory experiments are concerned fundamentally with creating situations that meet the theory's scope and initial conditions, and instantiate the independent variables of interest. Virtual reality software and hardware can add to the realism and may also reduce effects of distractions. Equipment to immerse a game player in a situation is coming to market, some of which even allows a player to physically move around in actual space and in the fantastical game space. As potential participants become accustomed to using that equipment – wearing a helmet that controls visual and aural information – it can be adapted to interactional experiments. These possibilities can help create conditions that involve contextual cues otherwise difficult to implement. For instance, interfaces with timing cues or attention demanding tasks might be easier to create virtually than concretely in a laboratory. It will, almost certainly, also entail some problems that we cannot now foresee.

So-called 'Big Data' collection

The Internet potentially connects researchers with huge numbers of respondents. At present, most of the research entails collecting existing information and cross-tabulating it. Researchers can, for instance, map all phone conversations in the United Kingdom and correlate frequencies of phone contacts with levels of economic activity (Eagle, Macy, Claxton 2010). This sort of uses raises a number of interesting issues, such as developing appropriate statistics for huge samples. As those uses are not experimental, we do not discuss them further here. However, vignette designs might be administered through the Internet to large samples of respondents.

In the U.S., there are a variety of commercial sites that will conduct surveys of greatly varying quality. To date, researchers mainly have used one of two routes to accessing respondents online. One is Time Sharing Experiments for the Social Sciences, or TESS (http://www.tessexperiments.org). TESS is a competitive program supported by the National Science Foundation that requires submission of proposals for research with a nationally representative sample of adults who are paid for surveys, questionnaires, vignette studies, and the like. For more information on TESS see: http://www.tessexperiments.org/introduction.html#pays.

The second avenue is Mechanical Turk on Amazon. This is a self-selected sample of respondents who may be hired for a large variety of tasks, including responding to vignette and questionnaires. Mechanical Turk is available to anyone with a budget to pay respondents. By comparison with TESS, there are two concerns with Mechanical Turk data. First, the project will not have undergone any review, since it is open to anyone through the Amazon web site. Second, while it is possible

to request respondents having certain characteristics, this is not enforceable. Thus it would be wise to consider the respondent pool to have unknown demographic characteristics and to have come from an unspecified population.

Internet use for data collection is very new. Problems such as sampling and finding new ways to create independent variables for experimental research are important concerns, but as with most techniques, researchers are likely to come up with ways to improve the usability of these data sources and the quality of data that are generated there. It is also important to mention that in the U.S. all research studies must go through human subjects review (IRB; see below) at the researchers' institutions, regardless of where the studies are to be conducted.

Simulators for research and teaching

Computer-controlled simulators are commonly used to train aircraft pilots to deal with various flight situations. Some law enforcement agents also receive training through simulations. Simulations have the advantage of presenting rare situations and giving practice in dealing with them. Risk of harm, of course, is virtually zero, and a situation may be repeated any number of times to improve learning.

For research, simulations can compare different training methods, and they can compare relative effectiveness of several interventions. Asking a manager for a raise or dealing with a difficult co-worker in business might be learned through simulations. Because situations may be expressed in many different ways, simulations for interpersonal situations will probably have to reflect many different independent variables.

Persistent objections and new requirements

In sociology, and to varying degrees in other social sciences, experimental methods are still subject to misunderstandings and even suspicion. Yet decades of theoretical, empirical and philosophical research shows that those concerns are misplaced and based on misunderstanding.

'Experiments are artificial'

Yes they are. A laboratory is an unusual site, unlike anything that most participants have encountered before, or ever will again. We see that fact as the best argument for using this method. The artificiality objection is rooted in a misunderstanding of the purpose of laboratory experiments. The purpose is not to generalize findings directly from the laboratory to outside settings any more than one would expect husband-wife interaction in one's own family to generalize directly to someone else's marriage. Particular findings are historical facts, and historical facts are unique to the time, place, individuals, and social structures in which they appear. The concern is not with initial conditions of the laboratory but with the general principles.

The purpose of experimentation is to test predictions derived from a set of abstract general principles – that is, from a theory. If predictions are confirmed, that

increases confidence in them. What does apply outside the laboratory is the structure of general principles. It is not the findings that generalize; rather, it is the set of general principles. A useful theory can explain and predict occurrences both in a controlled experimental situation and in any natural setting where instances of the concepts of the theory may be found. Webster (2016) provides more detail and some 35 references on artificiality. Foschi (2016) discusses the artificiality objection with particular reference to cross-cultural experiments.

'Experiments are immoral'

They can be. This concern may come from fear of the unknown among people without experience of social science experiments, or it may come from conflating social science experiments with some notorious instances of unethical medical experiments.

Two social science experiments done in past decades have been disturbing – the shock experiments (Milgram 1963), and the prison experiments (Haney, Banks, Zimbardo 1973; Haney, Zimbardo 1998). Both were studies of obedience to authority. In the Milgram experiment, participants were told to administer increasingly severe electric shocks to a confederate of the experimenters who cried out in simulated pain. Results showed that over half of the participants progressed in administering shocks up to what would reasonably be considered lethal. In the prison experiment, volunteers were assigned either to be guards or prisoners. The experiment had to be terminated early because a few of the 'guards' began to treat their 'prisoners' cruelly.[8] Movies have been made about both of those experiments, and unfortunately those movies sometimes are the only introduction that students get to experiments in introductory courses.

Both of those experiments were atheoretical. They were not designed to test derivations from any general propositions about behaviour, and so there is no way of knowing conditions under which we might find comparable behaviour in other settings. While we can imagine natural settings that seem comparable to a laboratory, we do not know if they really are. Do we believe that how college students act in contrived settings under the watch of a presumably wise professor tells us about the motivations and behaviour of Nazi prison guards or medical researchers? Of course not.

IRBs, informed consent and experiments

Institutional Review Boards (IRBs) have been established for protecting the welfare of human participants in research. In the U.S., every institution receiving funding from the government must establish a committee to review and approve

[8] We hasten to add that the researchers in both studies had humane motives and interests. They did not anticipate the immoral behaviour and they were concerned to find ways to avoid such outcomes. (Whether they should have anticipated the results is another question.) Those experiments were conducted before Institutional Review Boards had been established for the protection of human subjects' welfare, and before the days of informed consent requirements for research involving humans.

all research conducted there (Hegtvedt 2014). They usually require an approved informed consent document that participants receive and sign.[9]

Establishing the IRB and informed consent certainly came from the best of motives, and we believe in them. However, we believe that they do not go far enough. A potential participant is not the best judge of how great the social psychological risks of participation will be.

A main concern is psychological stress, and most of us are very poor at judging our own tolerance for stress. If you were to describe the electric shock experiments to undergraduate college students and ask them how they would feel if they initially thought they had given large shocks to someone and then learned that they had not actually shocked anyone, most of them are likely to tell you that it would not bother them once they knew they had not really hurt the learner. But that is not what happened. Milgram reported that some participants suffered nightmares for weeks afterwards. Clearly the experience was much more upsetting than students would have guessed.

People do not know how much stress they might feel in a situation that they have never experienced, or how great their tolerance and coping skills are. It is the job of the researcher, who is a trained scientist of human behaviour, to anticipate and to minimize such stressors, whether or not members of the participant population would recognize the danger.

Summary and conclusions

Experimental research offers many advantages to a theorist. He or she can create just the kind of situation needed to test predictions, and can vary the situation to follow up on new leads. At the same time, experimental design and operations require considerable time and work, since in the simplified situation of an experiment, every detail becomes important.

Researchers have shown ingenuity in creating experimental situations. At the same time, balance is required. Too much creativity could cause everyone to design a different setting for each research question. That leads to hundreds of non-comparable findings and little cumulative understanding. On the other hand, too little creativity may mean using existing designs where they are inappropriate, or failing to develop new designs when they are needed for new research questions.

We have offered a definition of the word 'experiment' based on the well accepted terms 'independent' and 'dependent variables.' We also have recommended extensive pretesting of any design, whether entirely new or an adaptation of an existing design.

After describing some existing basic experiments, we considered some promising new designs. New is not always better; if an existing design can be used, that is always preferable for developing cumulative findings. However, new designs sometimes are needed for studying new questions or for studying recognized questions

[9] The written informed consent can be waived under some circumstances, for example, when it would be the only document linking the names with participation.

in new settings. We outlined some of the established and new designs in the study of status and expectation state processes. We also surveyed some new technologies that hold promise for general uses in future experiments. Here, as everywhere in research, imagination, good judgment, careful attention to detail, and humble recognition of a researcher's own fallibility are probably essential to increasing knowledge.

References

Ariely D. (2008). *Predictably Irrational: The Hidden Forces that Shape Our Decisions*, 2nd ed. New York: Harper-Collins.

Asch S.E. (1951). *Effects of Group Pressure on the Modification and Distortion of Judgments*. In: H. Guetzkow (ed.), *Groups, Leadership and Men*. Pittsburgh: Carnegie Press, p. 177–190.

Asch S.E. (1948). *The Doctrine of Suggestion, Prestige, and Imitation in Social Psychology*. Psychological Review, 55, p. 250–276.

Axelrod R. (1984). *The Evolution of Cooperation.* New York: Basic Books.

Babkin B.P (1949). *Pavlov: A Biography*. Toronto, Canada: The University of Chicago Press.

Bales R.F. (1999). *Social Interaction Systems: Theory and Measurement*. New Brunswick, N.J., London: Transaction.

Bales R.F. (1950). *Interaction Process Analysis: A Method for the Study of Small Groups.* Cambridge, MA: Addison-Wesley Press.

Berger J. (2014). *The Standardized Experimental Situation in Expectation States Research: Notes on History, Uses, and Special Features*. In: M. Webster Jr., J. Sell (eds.), *Laboratory Experiments in the Social Sciences*, 2nd ed. New York: Academic Press, p. 269–293.

Berger J., Wagner D.G., Webster M., Jr. (2014). *Expectation States Theory: Growth, Opportunities, and Challenges*. In: S.R Thye, E.J. Lawler (eds.), *Advances in Group Processes,* vol. 31. Bingley, UK: Emerald Publishing, p. 19–55.

Berger J. Webster M., Jr., Ridgeway C.L, Rosenholtz S.J. (1986). *Status Cues, Expectations, and Behavior*. In: E.J. Lawler (ed.), *Advances in Group Processes*, vol. 3. Greenwich, CT: JAI Press, p. 1–22.

Bernard J. (1933). *The Distribution of Success in Marriage*. American Journal of Sociology, 39, p. 194–203.

Burgess E.W., Cottrell L.S. (1936). *The Prediction of Adjustment in Marriage*. American Sociological Review, 1, p. 737–751.

Campbell D.T., Stanley J.C. (1966). *Experimental and Quasi-Experimental Designs for Research*. Boston: Houghton Mifflin.

Campbell D.T., Stanley J.C. (1963). *Experimental and Quasi-Experimental Designs for Research on Teaching*. In: N.L. Gage (ed.), *Handbook of Research on Teaching*. Chicago: Rand McNally & Company, p. 171–246.

Cohen B.P. (1963). *Conflict and Conformity: A Probability Model and its Application.* Cambridge, MA: M.I.T. Press.

Compton D.L., Love T., Sell J. (2012). *Developing and Assessing Inter-coder Reliability in Group Interaction*. Sociological Methodology, 42, p. 348–364.

Cook K., Yamagishi T. (2008). *A Defense of Deception on Scientific Grounds*. Social Psychology Quarterly, 71, p. 215–221.

Davis K.B. (1929). *Factors in the Sex Life of Twenty-Two Hundred Women.* New York: Harper.

Eagle N., Macy M., Claxton R. (2010). *Network Diversity and Economic Development*. Science, 328, p. 1029–1031.

Fisek M.H., Berger J., Norman R.Z. (2005). *Status Cues and the Formation of Expectations*. Social Science Research, 34, p. 80–102.

Foschi M. (2016). *Experimental Contributions to Sociological Immigration-Research*. Studia Sociologica, this issue.

Foschi M., Lai L., Sigerson K. (1994). *Gender and Double Standards in Assessment of Job Applicants.* Social Psychology Quarterly, 57, p. 326–339.

Foschi M., Valenzuela J. (2008). *Selecting Job Applicants: Effects from Gender, Self-Presentation, and Decision Type*. Social Science Research, 37, p. 1022–1038.

Gallagher T.J., Gregory S.W., Jr., Bianchi A.J., Hartung P.J., Harkness S. (2005). *Examining Medical Interview Asymmetry Using the Expectation States Approach***.** Social Psychology Quarterly, 68, p. 187–203.

Goar C., Sell J. (2005). *Using Task Definition to Modify Racial Inequality Within Task Groups*. Sociological Quarterly, 46 (815), p. 525–543.

Goar C., Sell J., Manago B., Melero C., Reidinger B. (2013). *Race and Ethnic Composition of Groups: Experimental Investigations*. In: S.R. Thye, E.J. Lawler (eds.), *Advances in Group Processes*, vol. 30. London: Emerald Press, p. 47–75.

Gregory S.W., Jr., Webster S. (1996). *A Nonverbal Signal in Voices of Interview Partners Effectively Predicts Communication Accommodation and Social Status Perceptions*. Journal of Personality and Social Psychology, 70, p.1231–1240.

Haney C., Banks W.C., Zimbardo P.G. (1973). *A Study of Prisoners and Guards in a Simulated Prison*. *Naval Research Review,* 30, p. 4–17.

Haney C., Zimbardo P.G. (1998). *The Past and Future of U.S. Prison Policy: Twenty-Five Years after the Stanford Prison Experiment*. American Psychologist, 53, p. 709–727.

Hegtvedt K. (2014). *Ethics and Experiments*. In: M. Webster Jr., J. Sell (eds.), *Laboratory Experiments in the Social Sciences*, 2nd ed. New York: Academic Press, p. 23–51.

Hertwig R., Ortmann A. (2008). *Deceptions in Social Psychological Experiments: Two Misconceptions and a Research Agenda*. Social Psychology Quarterly, 71, p. 222–227.

Jasso G. (2006). *Factorial Survey Methods for Studying Beliefs and Judgments*. Sociological Methods and Research, 34: p. 334–423.

Jasso G. (2003). *Factorial Survey Method (Rossi's Method)*. In: M. Lewis-Beck, A. Bryman, T. Futing Liao (eds.) *The Sage Encyclopedia of Social Science Research Methods*, vol. 1. Thousand Oaks, CA: Sage Publications, p. 374–376.

Jasso G., Webster M., Jr. (1999). *Assessing the Gender Gap in Just Earnings and Its Underlying Mechanisms*. Social Psychology Quarterly, 62, p.367–380.

Kalkhoff W., Gregory S.W. (2008). *Beyond the Issues: Nonverbal Communication, Power Rituals, and 'Rope-a-Dopes' in the 2008 Presidential Debates*. Current Research in Social Psychology, 14, p. 39–51.

Milgram S. (1963). *Behavioral Study of Obedience*. *Journal of Abnormal and Social Psychology*, 67, p. 371–378.

Pavlov I.P. (1927). *Conditioned Reflexes: An Investigation of the Physiological Activity of the Cerebral Cortex.* Translated and edited by G.V. Anrep. New York: Dover Publications, Inc.

Rashotte L. S., Webster M., Jr., Whitmeyer J.M. (2005). *Pre-Testing Designs for Experiments*. Sociological Methodology, 35, p. 151–175. Reprinted in Vogt W.P. (ed.) (2008). *Selecting Research Methods*. London: Sage.

Ridgeway C.L., Backor K., Li Y. E., Tinkler J.E., Erickson K.G. (2009). *How Easily Does a Social Difference Become a Status Distinction? Gender Matters.* American Sociological Review, 74, p. 44–62.

Ridgeway C.L., Berger J. (1986). *Expectations, Legitimation, and Dominance Behavior in Task Groups*. American Sociological Review, 51, p. 603–617.

Ridgeway C.L., Boyle E.H., Kuipers K., Robinson D.T. (1998). *How Do Status Beliefs Develop? The Role of Resources and Interactional Experience.* American Sociological Review, 63, p. 331–350.

Ridgeway C.L., Correll S.J. (2006). *Consensus and the Creation of Status Beliefs*. Social Forces, 85, p. 431–454.

Ridgeway C.L., Erickson K.G. (2000). *Creating and Spreading Status Beliefs*. American Journal of Sociology, 106, p. 579–615.

Roethlisberger F.J., Dickson W.J. (1939). *Management and the Worker: An Account of a Research Program Conducted By the Western Electric Company, Hawthorne Works*, Chicago. Cambridge, MA: Harvard University Press.

Roseborough M.E. (1953). *Experimental Studies of Small Groups*. Psychological Bulletin, 50, p. 275–303.

Rossi P.H. (1979). *Vignette Analysis: Uncovering the Normative Structure of Complex Judgments*. In: R.K. Merton, J.S. Coleman, P.H. Rossi (eds.), *Qualitative and Quantitative Social Research*. New York: Free Press, p. 176–188.

Rossi P.H., Anderson A.B. (1982). *The Factorial Survey Approach: An Introduction*. In: P.H. Rossi, S.L. Nock (eds.), *Measuring Social Judgments: The Factorial Survey Approach*. Beverly Hills, CA: Sage.

Sauer C. (2014). *A Just Gender Wage Gap? Three Factorial Survey Studies on Just Earnings for Male and Female Employees*. SFB 882, Working Paper Series, no. 29, DFG Research Centre, available from Department of Sociology, Bielefeld University, PO Box 10 01 31, 33501 Bielefeld, Germany.

Schachter S. (1951). *Deviation, Rejection, and Communication*. Journal of Abnormal and Social Psychology, 46, p. 190–208.

Sell J. (2008). *Introduction to Deception Debate*. Social Psychology Quarterly, 71, p. 213–214.

Shelly R.K., Shelly A.C. (2009). *Speech Content and the Emergence of Inequality in Task Groups*. Journal of Social Issues, 65, p. 307–333.

Sherif M. (1936). *The Psychology of Social Norms.* New York: Harper.

Skinner B.F. (1953). *Science and Human Behavior*. New York: Macmillan.

Strodtbeck F. (1984). *Presentation of the Cooley-Mead Award for 1983. Difficult Decisions behind the Originality of Symlog*. Social Psychology Quarterly, 47, p. 95–101.

Terman L.M. (assisted by Buttenwieser P., Ferguson C.W., Johnson W.B., Wilson B.P.) (1938). *Psychological Factors in Marital Happiness*. New York: McGraw-Hill.

Thaler R.H. (2015). *Misbehaving: The Making of Behavioral Economics*. New York: Norton.

Thorndike E.L. (1905). *The Elements of Psychology*. New York: A.G. Seiler.

Thrasher F.M. (1927). *The Gang: A Study of 1,313 Gangs in Chicago.* Chicago: University of Chicago Press.

Walker L.S. (2014). *Developing Your Experiment*. In: M. Webster Jr., J. Sell (eds.), *Laboratory Experiments in the Social Sciences*, 2nd ed. New York: Academic Press, p. 127–144.

Walker L.S., Doerer S.C., Webster M., Jr. (2014). *Status Participation, and Influence in Task Groups*. Sociological Perspectives, 57, p. 364–381.

Watson J.B. (1913a). *Psychology as the Behaviorist Views It*. Psychological Review, 20, p. 158–177.

Watson J.B. (1913b). *Psychology from the Standpoint of a Behaviorist*. Philadelphia: J. B. Lippincott Company.

Webster M., Jr. (2016). *Research Opportunities*. Social Psychology Quarterly, p. 5–21.

Webster M., Jr., Rashotte L.S. (2010). *Behavior, Competence, and Status*. Social Forces, 88, p. 1021–1050.

Webster M., Jr., Sell J. (eds.) (2014). *Laboratory Experiments in the Social Sciences*. 2nd ed. San Diego: Academic Press.

Webster M., Jr., Walker L.S. (2016). *The Theories of Status Characteristics and Expectation States*. In: S.B Abrutyn (ed.). *Handbook of Contemporary Sociological Theory*. New York: Springer, forthcoming.

Webster M., Jr., Whitmeyer J.M., Rashotte L.S. (2004). *Status Claims, Performance Expectations, and Inequality in Groups*. Social Science Research, 33, p. 724–745.

Whyte W.F. (1955). *Street Corner Society: The Social Structure of an Italian Slum.* Chicago: University of Chicago Press.

Stan obecny i perspektywy na przyszłość eksperymentów w naukach społecznych

Eksperymentalne badanie zjawisk społecznych należących do obszaru zainteresowania psychologii podjęto w latach osiemdziesiątych 19. wieku na uniwersytecie w Lipsku. Po 4 czy 5 dekadach eksperymenty realizowane w laboratorium i w warunkach naturalnych pojawiły się także w socjologii. Eksperyment to szczególny rodzaj planu badawczego, polegający na tym, że przed pomiarem zmiennych zależnych zmienne niezależne poddaje się kontroli. Jakkolwiek we wszystkich planach badawczych mogą wystąpić czynniki zakłócające, przed ich działaniem zazwyczaj wystarczająco chroni losowe przypisanie jednostek do warunków eksperymentalnych. Istotne jest empiryczne określenie zmiennych w taki sposób, by różnice wartości były wyraźnie zarysowane, jak również uprzednie sprawdzenie wszystkich operacji; ocena mocy testów statystycznych jest także bardzo pomocna. W artykule tym śledzimy rozwój, jakiemu podlegał standaryzowany schemat badawczy, szeroko stosowany w badaniach statusu i procesów rozważanych w teorii stanów oczekiwań. Mamy tu na myśli ulepszenia operacji z użyciem kamery wideo i komputerów i nowe sposoby tworzenia zmiennych od opisu interakcji. Obecnie rozwijane są też nowe metody badania związku między akomodacją głosu w komunikacji z partnerem a pozycją w grupie. Technika winiet czynnikowych służy do zapewnienia kontroli eksperymentalnej poza laboratorium i umożliwia szybkie zebranie dużej ilości danych. Urządzenia do wytwarzania wirtualnej rzeczywistości i symulacje komputerowe podobne do używanych w szkoleniach kierowców i pilotów rokują nadzieje na zastosowanie w eksperymentach, lecz nie były jeszcze wykorzystywane w tym celu. Na końcu rozważamy pewne nieporozumienia, które mogą utrudnić szersze stosowanie metody eksperymentu, oraz sugerujemy pewne środki zaradcze, aby usunąć owe nieporozumienia.

Słowa kluczowe: plan eksperymentalny, nieporozumienia dotyczące eksperymentów, status, oczekiwania, big data (duży zbiór danych)

FOLIA 202

Annales Universitatis Paedagogicae Cracoviensis

Studia Sociologica VIII (2016), vol. 1, p. 81–95
ISSN 2081–6642

Robert K. Shelly
Ohio University, USA

Ann Converse Shelly
Ashland University, USA

Assessing Epistemic Claims by Experimental Evidence

Abstract

Research methodology decisions require clear criteria for selection of appropriate measures and procedures. These decisions often entail rules for assessing knowledge claims. Epistemic claims assert knowledge about underlying mechanisms that produce observable phenomena. These claims specify relationships between observable attributes, concepts, and theoretical constructs. We explicate three ways in which epistemic claims may be advanced and assessed: triangulation, multitrait-multimethod, and meta-analysis. We assess each of these approaches and review research examples for each method to advance claims about the relationship of experimental evidence to theory and its validity.

Key words: **s**tatus, epistemology, experiments, meta-analysis, theory

> *If it is true that every theory must be based on observed facts, it is equally true that the facts cannot be observed without the guidance of some theory. Without such guidance, our facts would be desultory and fruitless; we could not retain them: for the most part we would not even perceive them.*
>
> Auguste Comte (quoted in Stein 2008)

Introduction

When reviewing research methodologies, it is critical that epistemological question(s) and entailed research questions be clearly delineated. Epistemology is the study of knowledge and justified belief. As the study of knowledge, epistemology is concerned with the following questions: What are the necessary and sufficient conditions of knowledge? What are its sources? What is its structure, and what are its limits? Understood more broadly, epistemology is about issues having to do with the creation and dissemination of knowledge in particular areas of inquiry. To focus the discussion more clearly, research questions need to deal with the theory of knowledge, especially the critical study of its validity, methods, and scope of knowledge claims.

Research should address five questions. What is knowledge? How is knowledge acquired? What do people know? How do we know what we know? Why do we know what we know?

In each of the analyses that follow, the nature and type of the research questions are spelled out carefully so that the chosen methodology results in answers to the questions that are epistemologically justified. Denzin's triangulation methodology focuses on the issue of *how do we know what we know.* The multitrait-multimethod process focuses on the question of *how is knowledge acquired*, as well as on the previous question. Meta-analysis has, as part of its focus, the concept of *what is knowledge*, as well as that of *what do people know*.

Three decision criteria are applied to research programs to address epistemic questions identified above. First, we assert that the nature and type of research question is spelled out carefully so that the selected methodology results in epistemologically justified answers to the research questions. Second, we assume the researcher has formulated his or her research question based on a theoretical model so that an experiment tests some aspect of a mechanism linking social conditions and attributes of the research population to expected outcomes. Finally, we assume the investigator has developed an experimental design to mitigate threats to reliability of measures and internal validity linking cause and effect.

Decision rules for each of the approaches to epistemology allow us to answer the question of *why we know what we know*. Each approach answers this question in a different way. For instance, the triangulation approach asks the researcher to assess whether information is consistent across three elements of the knowledge claim (e.g. theory, method, data) by means of a verbal comparison. The multitrait-multimethod approach attaches correlation coefficients to trait measures, methods of measurement, reliability, and validity claims but it does not explicitly specify decision criteria. The unstated implication is that coefficients are assessed with decision criteria consistent with statistical significance levels for the correlations. Meta-analysis is the statistically rigorous approach as it makes use of statistical decision rules to assess claims about what we know. There are multiple techniques for assessing effects in meta-analysis with decision criteria for each technique.

For any study, the need to refine the research questions is paramount. Many of the mechanisms employed for this purpose may be classified in one of two approaches. The first approach focuses on the effects of altering conditions to identify the effect on an outcome of changing circumstances or conditions. Often, this approach does not include a specification of the mechanisms thought to provide the link between circumstance and condition and an outcome. The second approach focuses on a theoretical model of the processes thought to link circumstances and conditions to outcomes. Frequently, the mechanism linking circumstance and condition to outcome is specified and differential outcomes posited are based on differences in circumstance and condition (Zelditch 2007).

We focus on epistemological claims that methodological processes both require and respond to in the research setting. We explore these claims with an analysis of the requirements and the logic of *triangulation*, *multitrait-multimethod technique*, and *meta-analysis*. Each analysis is followed by a research example to illustrate how experiments have contributed to answers of epistemological questions. We begin our discussion with triangulation, follow with multitrait-multimethod, and conclude with meta-analysis. Each of the research examples

employed has a robust experimental research record supporting substantive claims of a theory of human behaviour.

Relations between theory and experiment

We describe the role of theory in specifying research questions and experimental protocols. This discussion includes a review of theory as the source of research questions, the specification of experimental designs, and the role of quasi-experimental designs. Our review of these topics is brief, as many of the ideas are well known among experimentalists in social science.[1]

Experimental research can be based on two distinct formulations of a research question. In one case, the research question is based on an empirically grounded theory. The goal of a study using such a theoretical model is to determine whether or not the theory is supported by experimental results. Investigators introduce controls in the situation in which individuals interact with one another. These controls create conditions for a strong test of the theory. Logically interrelated propositions are employed to identify an appropriate research question, to test the mechanism posited in the theory, its possible domain of application, and the appropriate controls over initial conditions in the experimental setting. For instance, in studies of negotiated exchange, a theoretical model specifies the nature of the social structure that determines who may negotiate with whom, how negotiations must run in order to reach a successful outcome in each of the allowed number of rounds. The experiment thus tests for the operation of a mechanism thought to govern negotiations between social actors (Markovsky, Willer, Patton 1988).

In the second case, investigators are interested only in the effect that changes in the independent variable may have on outcomes for the dependent variable. Clinical trials in medical research often employ this approach when they attempt to assess the effectiveness and the level of toxicity of a new drug. Investigators need not posit a particular mechanism to answer these questions, they design an experiment just to test for these effects in a population.

In both instances, the investigator is concerned about threats to the internal validity of experimental procedures. These threats include: history affecting participants in the experiment, maturation of participants, selection biases in which participants are assigned to conditions of the study in ways which create biased groups, an interaction between selection and maturation, testing threats which occur when measures affect responses by biasing possible answers, regression to the group centre point, instrumentation errors, experimental mortality, and experimenter bias (Campbell, Stanley 1966, Thye 2014).

We assume that investigators have formulated research questions based on a theoretical model so the experiment tests a hypothetical mechanism linking social conditions and attributes of the research population to the expected outcomes. We also presume that investigators have developed experimental designs to mitigate the threats to internal validity. The most common means of accomplishing this

[1] For those interested in more exposition see Webster, Sell (2014) and Campbell, Stanley (1966).

second goal is to create two or more groups of participants equated by means of randomized assignment of participants to groups. These groups may or may not be assessed for equivalence on relevant variables thought to have potential influence on outcomes of the study prior to the introduction of the experimental treatment. One group is then exposed to the experimental treatment while the other is not, though equivalent experiences may be presented for the latter group. Following this treatment phase, a measurement process is implemented to determine if the treatment has had an effect on the experimental group. Outcomes of these measures are then compared to those of the unexposed group. If the experimental treatment has the expected effect, the treated group should differ from the untreated on these post exposure measures.

Our interest is how to assess evidence from studies focused on the same theoretical research question. We presume that research evidence exists from a large number of studies that may be applied to answer questions posed by a theory. We further assume that some studies satisfy the requirements for a true experiment: at least two groups, random assignment of persons to groups, exposure of one group to an experimental treatment, measures that allow the comparison of the effects of the treatment to non-treated participants, and the comparison of data to hypotheses of the theory. Some studies may fail to satisfy one of these requirements, such as random assignment, and yet satisfy other requirements of a theory test. We treat such studies as quasi-experimental tests of the theory of interest and also include them in our discussion of examples when assessing evidence for epistemic claims about the theory of interest. Campbell and Stanley (1966) address the various ways in which quasi-experiments approximate experimental designs and the potential strength of claims that may be made. Figure 1 presents a pictorial representation of our interest.

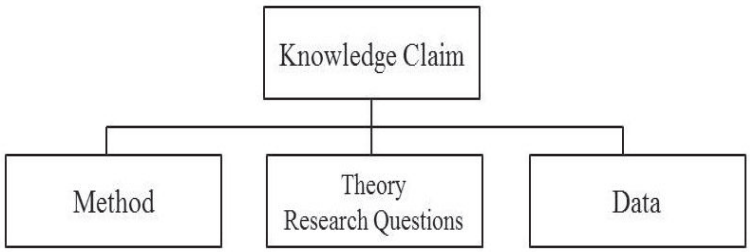


Figure 1. Assessing epistemic claims from experiments

We turn now to a discussion of three means of aggregating information from a number of studies to assess epistemic claims from experimental evidence. We begin with a review of triangulation as the least precise mechanism of comparison. We then discuss the more precise multitrait-multimethod comparison technique. We conclude with a discussion of the most precise method of aggregating research results: meta-analysis.

Triangulation as a method of assessing research claims

Triangulation as a method of assessing knowledge claims in sociological research emerged with the work of Norman Denzin (1970). The publication of his book *Research Act* invoked a methodology that is focused on comparing multiple sources of information relevant to a given knowledge claim. This effort to take into account the comparison of theory, method, and data in assessing research claims has been one of the most robust approaches employed in sociology when making claims about what we know, and how we know what we know. In our review of the main points of this approach, we include a brief discussion of its intellectual history, highlight comparisons suggested by Denzin, discuss briefly the current state of the approach, and point to the benefits and concerns arising from this approach.

Triangulation as a method of assessing knowledge claims first appeared in the social science research literature in an essay by Feigl (1931) published in one of the first volumes of the University of Minnesota studies in the philosophy of science. Subsequently, it has been cited by investigators attempting to address questions about how effectively various tests of intelligence are accomplishing their goal of measuring mental abilities. This approach relies on the idea that two or more tests could be compared to one another and to a third, more valid assessment of the ability of interest.

Denzin cites the research program of Donald Campbell and his associates in his rationale for this approach and suggests several strategies to compare knowledge claims to one another to arrive at a scientific explanation of observed behaviour. In particular, he stresses the importance of employing multiple methods emphasized by Campbell and Fiske (1959) and Webb et al. (1966). Four basic types of triangulation are identified for consideration: *data triangulation* in which time, space, and persons observed are employed to assess a knowledge claim; *investigator triangulation* in which multiple investigators observe the same activity; *theory triangulation* in which multiple theories are invoked in assessing the same activity to assess a claim; and *methodological triangulation* in which survey, experiment, and participant observation are employed to assess the same activity.

Data triangulation and investigator triangulation are relatively straightforward in their demands on the investigator. Time, space, and persons observed suggest an approach in which studies are carried out at various times, in various locations, and with a variety of participants. Results from these studies are then compared to one another to arrive at a conclusion about tested hypotheses. Investigator triangulation suggests that studies might be carried out by different investigators in various locations or across times. Hypotheses confirmed across these studies provide more confidence that the conclusions are both valid and reliable.

Theoretical triangulation presumes that an investigator is able to identify several theories that may be employed to explain the same phenomenon. The investigator is able to identify information to test various hypotheses and collect this information from the participants in a study. The resulting data allows the investigator to conclude that one or more of the theories is not supported by the data. If this is the case, the conclusion is that a critical experiment has provided evidence for the differential strength

of the theory or theories used to analyse the data. If all of the theories are supported by the data, unified theory may be developed to explain the results of the study.

If all of the hypotheses are not supported by the data, the investigator is left to conclude that his/her theories do not adequately explain the phenomenon of interest. Limited guidance is provided for further work in this last instance. Reformulation of one or more of the theories tested in this situation is the sole option for the investigator.

Methodological triangulation presumes that the investigator is able to formulate a research question so that it may be tested with multiple methods. For instance, if this approach is pursued, a research question testable in an experiment may be testable with a survey and/or an observation study. Results of each study would then be expected to be consistent with one another and with the theory from which the research question is formulated. For instance, a correlation between status and influence observed in an experimental setting would be expected to appear in survey data, as well as in an observation study of interaction in a naturalistic setting.

Our interest in applying Denzin's approach focuses on one aspect of triangulation. We presume that one theory is tested in an experimental setting as the beginning point for comparison involving data and methodological triangulation. We begin with a theory specified so that hypotheses may be tested in an experiment in which the investigator is able to approximate a random assignment of participants to different initial situations. These participants are then differentially exposed to experimental treatments and a behavioural measure of the effectiveness of the treatment is compared to the theoretical prediction. The theory is reinforced if the results are consistent with the hypotheses and becomes suspect if the results do not support the hypotheses. This logic is consistent with the position advocated by Popper (1959) that theoretical ideas must be refutable if they are to be valuable as guides to scientific inquiry.

An example of this triangulation approach is available to us in the *expectation states research program*. Various experimental studies have shown that diffuse status characteristics affect formation and enactment of expectations for future behaviour. These expectations and their translation to behaviour were tested by Berger, Cohen B., Zelditch (1972). U.S. Air Force enlisted personnel served as subjects in the study. When subjects believed their co-participant was higher ranking, they were likely to defer to them in a decision making task. When subjects believed they outranked their co-participant, they were less likely to defer to them. This result has been replicated in a number of studies with diffuse status characteristics as varied as age, year in school, academic success, and appearance. Our interest is in whether or not the result is observed in settings outside the laboratory.

Two studies are compared to this result in this example of triangulation. Each study was carried out in a field setting, in one case with survey methodology and in the other case a series of studies employed quasi-experimental designs. The survey methodology was used in a study of research and development teams in industrial firms in the United States. Members of the teams were asked to respond to a series of questions which asked them to indicate who in the team had the best ideas, whose ideas were most useful in solving research and development tasks, and who ranked higher than whom in task ability. Results, with some exceptions for individuals in management

positions, were consistent with the findings of the experimental study. Individuals with high diffuse status received higher scores on all of the measures (Cohen B., Zhou 1991).

Quasi-experimental studies were carried out in school classrooms by Elizabeth G. Cohen (Cohen E., Roper 1972, Cohen E. 1982). Students in the classrooms were assigned to small task groups so that experimental groups contained members of both diffuse status attributes of majority and minority ethnic/racial groups. Control groups were homogeneous with respect to the diffuse status attributes. Observations of task solving activity were then carried out and students were asked to identify those with the best ideas for task solutions. High diffuse status students were more influential in solving the task and frequently talked more than low diffuse status students in heterogeneous groups. In homogeneous groups, this difference in behaviour did not occur.

The result of this triangulation comparison shows that diffuse status attributes affect the formation of performance expectations and their translation into behaviour as predicted by expectation states theory. The result is observed in adults in the Air Force who believe they are interacting with higher/lower rank others, members of research and development teams in industrial settings, and school populations of children. Our conclusion is based on a verbal comparison of the results of the three studies. Possible effects of setting and time on task are not considered in this comparison. Tasks in both experimental study and quasi-experimental study are relatively brief in duration, while they rely for their definition of status on cultural knowledge of long duration. Task activity in the research and development teams is of long duration and allows for many opportunities for participants to observe behaviours that may reinforce, or contradict the diffuse status hierarchy of the group.[2]

Multitrait-multimethod assessment of knowledge claims

The multitrait-multimethod technique of assessing knowledge claims was first formally presented as a way of assessing epistemic claims by Campbell and Fiske (1959). Earlier attempts to employ a similar logic were developed by Feigl (1931). The approach relies on correlation analysis of multiple measures of multiple traits observed on individuals. An abstract example of this technique with two traits and two methods of assessing each trait is displayed in Table 1. Traits may be thought of as theoretical constructs such as numeracy and literacy with two tests for each trait. Assessment of knowledge claims for each trait and their validity rest on the strength of these relationships.

Three measures of association ab11, ab12, and ab22 are available for each pair of traits and methods. The first measure to consider is how strongly the methods of measurement are related to one another for each trait. Assessment of these correlations is accomplished by the usual rubric for determining that an association is

[2] Still to come are experiments embedding a representation of status as a continuum rather than a dichotomy and the status function as increasing at an increasing rate with status characteristics which generate status. These ideas have been discussed for over fifty years (Bales 1950, Stephan, Mishler 1952, Goode 1978, Sørenson 1978, Shelly 1998, Jasso 2001).

significantly different from zero. If these criteria are reached, each trait and method are thought to be related to one another, though a strong theoretical rationale for this expected association may not be fully developed by the investigator.

Table 1. A hypothetical multitrait-multimethod matrix

Methods	Method 1			Method 2		
	Traits	A(1)	B(1)		A(2)	B(2)
Method 1	A(1)	Rel				
	B(1)	ab11	rel			
Method 2	A(2)	rel (val)			rel	
	B(2)	ab12	rel(val)		ab22	rel

Key: rel = reliability of the measure; ab(ij) = correlation of the two measures with one another; val = validity of the measures with respect to the concept

The measures of reliability are in the cells of the table normally assigned to the unit association of a measure with itself. These may vary substantially. Again, the application of a rationale for concluding that a trait is reliably measured relies on one of two assessments. One may conclude that a measure is reliable if a normative threshold, say .60, is reached. An alternative is to specify a value if the measure is statistically different from zero.

Finally, the validity measure, val, associates the measure of one trait with the measure of this first trait as measured independently. The criterion for reaching this conclusion is again based on employing a decision rule based on statistical significance. Identifying how the measure of validity is assigned values is unclear for most experimental studies based on theoretical models.

Application of the multitrait-multimethod epistemic model has had limited success in experimental studies for several reasons. First, laboratory studies often do not include tests of the external population validity of theoretical interest in the measurement process. This fact is often cited incorrectly as a reason to question the power of the results of such studies, even though they may provide very robust and consistent results across many replications (Webster, Kervin 1971, Zelditch 2007, Thye 2014). Philosophy of science analysis of this point emphasizes generalization to a class of phenomena rather than the population of observation units (Korner 1966). Second, measurement processes in many experiments often do not include multiple measures of multiple traits. Studies frequently involve a behavioural measure or measures which serve as tests of the hypotheses. Measures of the extent to which participants meet criteria specified by the scope and initial conditions are often collected in an experiment. However, reports of how these assessments are related to behavioural measures used to test the theory are often absent in research reports. We describe an example in which this technique may be employed to assess epistemic claims. This study does not meet the criteria for experimental designs (Campbell, Stanley 1966). It is instructive as an attempt to employ multiple methods to measure multiple traits and address theoretical questions.

Shelly and Shelly (2009) report on a set of data assembled in two different contexts. In one context, university students in classes were asked to discuss a class project problem and arrive at a decision about how to proceed. In the second context, the participants were asked to brainstorm the creation of a new task for future discussion groups to solve. This data set included a quasi-experimental design in which the gender composition of the groups varied. Data analysed in this study consisted of transcripts of the group interaction.

Three traits or constructs were measured in the study. The first construct included three indicators of how often members of the groups offered contributions. The second construct, also with three indicators, concerned how often the members of a group tried to provide organizational suggestions to the group. The third construct included three indicators of the complexity of verbal expressions initiated by participants. All three concepts are linked to behaviour interchanges patterns (BIP) (Fisek, Berger, Norman 1991). Forty two of the forty five correlations between the individual measures of the concepts are statistically significant with thirty seven of these at the .01 level of significance. The study includes measures of reliability for the three constructs, all of which are quite high (.80 plus range), but does not include a direct measure of validity. If we treat the association with the BIP construct as a measure of validity, then the study satisfies the multitrait-multimethod approach to making knowledge claims from experiments.

Multitrait-multimethod assessment of knowledge claims has value in highlighting the extent to which several constructs may be compared with and contrasted to one another. Such studies often answer the *what do we know* questions, and to some extent the *how do we know what we know* questions. Other epistemic questions are less likely to find answers with this technique. This is in part due to the ambiguous nature of statistical decision criteria and the fact that social science experiments often do not have multiple measures of multiple traits realized in their design. We turn now to meta-analysis as a tool to provide answers to epistemic questions.

Meta-analysis in assessing knowledge claims

Meta-analysis has emerged as a method of assessment for knowledge claims which have been addressed in a large volume of experimental investigations, conducted by a large number of investigators in a large number of settings. This method is conceived so as to allow investigators to answer fundamental questions pertaining to epistemological claims about what is known and the conditions under which this knowledge is acquired. Often, investigations produce results that seem to contradict one another, apply in some settings and not others, and create confusion for the users trying to assess the state of knowledge related to a particular scope of investigation or a theory of behaviour.

This method of assessing knowledge claims employs one of at least four different approaches. The simplest from a technical point of view is also the least informative. One may count studies as one would count votes in an election with yeas and nays counted up and a statistical decision made by determining whether one side of

the result is statistically more probable than the other (Bushman, Wang 2009). This approach allows an assessment of what is known, but with limited precision about how one knows what is known, and how we know what is known.

A second approach makes use of effect sizes from each of the studies and takes into account the assembly of these effects in comparing the statistical distribution of expected results with the hypothesis of no effect. Effect sizes are appropriate when an index is used to quantify the relationship between any two variables or the difference between two measures of a variable in two groups. Four properties are desirable for effect sizes to be valuable in a meta-analysis. Effect sizes must measure the same thing, be substantively interpretable, computable from information in the research report, and have good technical properties (Borenstein 2009). There is a clear increase in precision about what is known and to some extent about how we know what is known.

Effect sizes may be based on raw scores, standardized scores, correlation coefficients, proportions, or odds ratios. It is possible to convert one of these measures to another. To do so from raw scores to standard scores is a simple statistical manipulation, but to convert correlation coefficients, proportions, or odds ratios to standard scores requires technical skill beyond our interest (see Borenstein 2009 for details on these procedures). Interpretation of effect sizes may be based on comparison to other effect sizes of well-known results. If multiple experimental conditions can be operationalized as a covariate, it is possible to employ analysis of variance or regression techniques to assess the relative value of different experimental designs in producing results (Shadish, Haddock 2009). Dividing effects in this way increases our knowledge of what we know by specifying conditions wherein our knowledge is supported by experiments. We also increase the how we know what we know with this analysis.

Methods for assessing effect sizes for data based on proportions may be based on the difference between two probabilities, the ratio of two probabilities, the phi coefficient, and odds ratios. Each of these means of assessing the effects observed in a dichotomous variable may be assessed for the effects of covariates on the effect measure. Standard regression techniques apply in this situation with the appropriate technique dependent on the technical property of the effect measure (e.g. logit). Techniques may include regression, adjustment, and matching to refine the analysis (Fleiss, Berlin 2009). For constructs that may be assessed as proportions, this approach creates substantial increases in precision about the relationship of context to outcome.

In our example, three theories to explain a well-known empirical link are tested with a meta-analysis of a large number of studies linking appearance of a target to assessment of the target's abilities by participants in the studies.

Appearance of a target individual leads judges to an assessment of intellectual competence. Persons who are judged to be more attractive are also judged to be more intelligent in these studies. This meta-analysis explores three possible explanations for this result and uses effect scores to test the hypotheses derived from the theories. The three theories were: implicit personality theory, expectancy theory, and status generalization theory. A total of 36 studies reported in 30 articles were analysed to test the theories in use to explain the association. The studies included

adults and children as targets, male and female judges, and perceived intellectual competence and actual intellectual competence for some of the studies.

Status generalization theory (Webster, Driskell 1983) results in the formulation of five predictions for the analysis. The first prediction is that physically attractive people should be perceived as more intellectually competent than unattractive people. The second prediction is that attractiveness effects should be stronger for males than for females. The third is that attractiveness effects should be stronger when explicit information about competence is absent than when it is present. The fourth prediction is that actual competence should be greater for more attractive people than for less attractive people. The fifth prediction is that attractiveness should have stronger effects when indirect measures rather than direct measures of competence are used.

Implicit personality theory has been employed in an earlier meta-analysis of this association (Eagly et al. 1991). Only one prediction from implicit personality theory was identified for this analysis. Attractive people should be perceived as more intellectually competent than less attractive people. This is the first prediction of status generalization theory.

Expectancy theory has also served as the basis of a meta-analysis of the association between appearance and intellectual competence (Jackson, Hunter, Hodge 1995). Expectancy theory presumes a link between a perceiver's expectancies and the behaviour of a target. The self-fulfilling prophecy thus created should account for the association between appearance and intellectual competency. This is also the fourth prediction of status generalization theory.

The results of the analysis support the first prediction of status generalization theory, and hence implicit personality theory. The second prediction of status generalization theory was supported for adults, but the test for children could not be conducted as too few studies reported relevant data. The third prediction of status generalization theory was supported for both adults and children, though the children's data was available in only two studies. There was modest support for the fourth prediction. The results were modest for children and mixed for adults, thus making expectancy theory a weak explanation for the association. Finally, for the fifth prediction, the perceived competence of adult targets was in the right direction, but did not reach statistical significance. This prediction was supported for adult judges with respondents relating appearance more strongly to indirect measures of competence than direct measures.

The conclusion of this meta-analysis is that status generalization theory is a much more robust explanation for the link between physical appearance and perceptions of intellectual competence than either implicit personality theory or expectancy theory. We summarize these results in Table 2. The + sign in the table indicates a positive result from the meta analysis. In some instances, results are weak for one subpopulation such as males when no intellectual data is available.

Table 2. Summary of theoretical predictions for the link between appearance and intellect

Predictions	Target: Adults		Target: Children
1. Appearance	+		+
2. Gender	+		No data
3. No Intellectual Data	+		+ (Weak Overall, Low Males)
4. Actual Competence	Mixed		Weak
5. Perceived Competence (Target)	+	Judge+	No data

Meta-analysis offers the best alternative for answering our five questions about how epistemic claims are assessed. It relies on being able to provide quantitative measures for dependent variables and provides the strongest assessment opportunities when the independent variables specified by a theory are available in quantity sufficient to allow conditional statements based on the theory. These conditional statements allow us to answer the questions about what is known, how the knowledge is acquired, the extent to which it may be shared across domains of inquiry, and provides conditional answers to why we know what we know. It relies on the ability to quantise (restrict to discrete values) dependent variables and provide measures of central tendency and dispersion for its application to be robust.

Concluding remarks

Our initial questions include five epistemic concerns. Knowledge, as we use the term, is based on empirical testing of theoretical ideas. What is knowledge? How is knowledge acquired? What is known? How do we know what we know? Why do we know what we know? The discussion we have developed attempts to provide answers to these questions in the domain of experimental research in social psychology. We are able to comment about each of the methods of aggregation to answer our questions. Our examples provide illustrations of the success of each approach.

First, our beginning question about what is knowledge is answered by each of the methods for aggregating what we know from research. Each provides guidance in leading to conclusions about what to include in our scientific claims about the empirical world. Each provides rules for assembling information, evaluating it, and concluding that we know (or do not know) something about some phenomena. Each example we present makes knowledge claims that are distinguishable from competing claims.

Knowledge is acquired by experiments based on a theory. This assertion is illustrated by our examples. For triangulation, the basic result is identified in an experiment and tested in field settings employing features of experimental designs. The study discussed in the multi-trait-multimethod technique has features of experimental method, but both lacks random assignment to different conditions and is at best a quasi-experimental design. The example provides important information about how groups organize themselves and solve problems. The study illustrating the meta-analysis approach makes use of experimental data to reach aggregated

conclusions about how status processes affect behaviour. They have provided valuable insights about experimental process and theoretical explanation and allow us to refine our investigative approach.

All three techniques provide information about what we know about the effects of social status on behaviour. Diffuse status characteristics such as age, gender, race, and appearance affect social behaviour so that those with advantages enjoy more opportunities to talk.

Finally, answers to the question of why do we know what we know are most successfully addressed by triangulation and meta analysis. In both instances, mechanisms are frequently specified by the theory under test in an experiment. Aggregation of information is accomplished when data, theory, and method, or what Berger (2014) refers to the 'holy trinity' of investigation in social psychology, are consistently applied to the same research question. We prefer a slightly different concluding message. Epistemology and method result in meaning for research, they provide clear decision rules for weighing evidence, and lead to sound conclusions about theoretical claims. In this sense, what is known about the processes by which social status emerges in social interaction and its consequences once established is substantiated by each of the approaches to aggregation.

Triangulation and multitrait-multimethod provide strong answers to the question of how do we know what we know by specifying the links between theory, data, and measures. Meta-analysis is not quite as robust on this issue as it is most successful at identifying how we know what we know when a theoretical mechanism has been specified, as we saw with the analysis of the link between appearance and perception of intellectual ability.

Finally, answers to the question of why do we know what we know are most successfully addressed by triangulation and meta analysis. In both instances, mechanisms are frequently specified by the theory under test in an experiment.

The application of any of these techniques depends upon a specific set of principles in the formulation and execution of an empirical research investigation. First, we presume the research question has been spelled out carefully so that the method selected results in epistemically justified answers. Second, a strong test of a theoretically based research question is best carried out within an experimental design that tests a hypothetical mechanism linking social conditions and attributes of the research population to the expected outcomes. Finally, we presume the experimental design will mitigate threats to internal validity.

References

Bales R.F. (1950). *Interaction Process Analysis*. Cambridge, MA. Addison Wesley.

Berger J. (2014). *The Standardized Experimental Situation in Expectation States Research: Notes on History, Uses, and Special Features*. In: M. Webster Jr., J. Sell (eds.), *Laboratory Experiments in the Social Sciences*, Boston: Academic Press, p. 269–293.

Berger J., Cohen B.P., Zelditch M., Jr. (1972). *Status Characteristics and Social Interaction*. American Sociological Review, 37, p. 241–255.

Borenstein M. (2009). *Effect Sizes in Meta-Analysis*. In: H. Cooper, L.V. Hedges, J.C. Valentine (eds.), *The Handbook of Research Synthesis and Meta-Analysis*. 2nd edition. New York: Russell Sage Foundation, p. 261–236.

Bushman B.J., Wang M.C. (2009). *Vote Counting Procedures in Meta-Analysis*. In: H. Cooper, L.V. Hedges, J.C. Valentine (eds.) *The Handbook of Research Synthesis and Meta-Analysis*. 2nd ed. New York: Russell Sage Foundation, p. 207–220.

Campbell D.T., Fiske D.W. (1959). *Convergent and Discriminant Validation by the Multitrait-Multimethod Matrix*. Psychological Bulletin, 56, p. 81–105.

Campbell D. T., Stanley J.C. (1966). *Experimental and Quasi-Experimental Designs for Research*. Chicago: Rand McNally.

Cohen B.P., Zhou X. (1991). *Status Processes in Enduring Work Groups*. American Sociological Review, 56, p. 179–188.

Cohen E.G. (1982). *Expectation States and Interracial Interaction in School Settings*. Annual Review of Sociology, 8, p. 209–235.

Cohen E.G., Roper S.S. (1972). *Modification of Interracial Interaction Disability: An Application of Status Characteristics Theory*. American Sociological Review, 37, p. 643–657.

Denzin N.K. (1970). *The Research Act: A Theoretical Introduction to Sociological Methods*. Chicago: Aldine.

Eagly A.H., Ashmore R.D., Makhijani M.G., Longo L.C. (1991). *What Is Beautiful is Good. A Meta-Analytic Review of Research on the Physical Attractiveness Stereotype*. Psychological Bulletin, 110, p. 109–128.

Feigl H. (1931). *The Mental and the Physical*. In: H. Feigl, M. Scriven, G. Maxwell (eds.) *Minnesota Studies in the Philosophy of Science: Concepts, Theories, and the Mind-Body Problem.* Minneapolis: University of Minnesota Press, p. 370–497.

Fisek M.H., Berger J., Norman R.Z. (1991). *Participation in Heterogeneous and Homogeneous Groups: A Theoretical Integration*. American Journal of Sociology, 97, p. 114–142.

Fleiss J.L., Berlin J.A. (2009). *Effect Sizes for Dichotomous Data*. In: H. Cooper, L.V. Hedges, J.C. Valentine (eds.), *The Handbook of Research Synthesis and Meta-Analysis*. 2nd ed. New York: Russell Sage Foundation, p. 237–253.

Goode W.J. (1978). *The Celebration of Heroes: Prestige as a Control System*. Berkeley, CA. University of California Press.

Jackson L.A., Hunter J.E., Hodge C.N. (1995). *Physical Attractiveness and Intellectual Competence: A Meta-Analytic Review*. Social Psychology Quarterly, 58, p. 108–122.

Jasso G. (2001). *Studying Status: An Integrated Framework*. American Sociological Review, 66, p. 96–124.

Korner S. (1966*). Experience and Theory: An Essay in the Philosophy of Science*. New York: Humanities Press.

Markovsky B., Willer D., Patton T. (1988). *Power Relations in Exchange Networks*. American Sociological Review, 53, p. 220–236.

Popper K.R. (1959). *The Logic of Scientific Discovery*. New York. Basic Books.

Shadish W.R., Haddock C.K. (2009). *Combining Estimates of Effect Sizes*. In: H. Cooper, L.V. Hedges, J.C. Valentine (eds.), *The Handbook of Research Synthesis and Meta-Analysis*. 2nd ed. New York: Russell Sage Foundation.

Shelly R.K. (1998). *Some Developments in Expectation States Theory: Graduated Expectations*. In: E.J. Lawler, J. Skvoretz, J. Szmatka (eds.), *Advances in Group Processes*, vol. 15. Stamford, CT: Jai Press, Inc.

Shelly R.K., Shelly A.C. (2009). *Speech Content and the Emergence of Inequality in Task Groups*. Journal of Social Issues, 65, p. 307–333.

Sørenson A.H. (1979). *A Model and a Metric for the Analysis of the Intragenerational Status Assignment Process.* American Journal of Sociology, 85, p. 361–384.

Stein J.D. (2008). *How Math Explains the World.* New York: Harper Collins.

Stephan F.F., Mishler E.G. (1952). *The Distribution of Participation in Small Groups: An Exponential Approximation.* American Sociological Review, 17, p. 598–608.

Thye S.R. (2014). *Logical and Philosophical Foundations of Experimental Research in the Social Sciences.* In: M. Webster Jr., J. Sell (eds.), *Laboratory Experiments in the Social Sciences*. 2nnd ed. New York: Elsevier, p. 53–82.

Webb E.J., Campbell D.T., Schwartz R.D., Sechrest L. (1966). *Unobtrusive Measures: Nonreactive Research in the Social Sciences*. Chicago: Rand McNally.

Webster, M., Jr., Driskell J.E. (1983). *Beauty as Status*. American Journal of Sociology, 89, p. 140–165.

Webster M., Jr., Kervin J. (1971). *Artificiality in Experimental Sociology*. Canadian Review of Sociology and Anthropology, 8, p. 263–272.

Webster M., Jr., Sell J. (2014). *Why Do Experiments?* In: M. Webster Jr., J. Sell (eds.) *Laboratory Experiments in the Social Sciences*. 2nd ed. New York: Elsevier, p. 5–21.

Zelditch M., Jr. (2007). *The External Validity of Experiments That Test Theories.* In: M. Webster Jr., J. Sell (eds.), *Laboratory Experiments in the Social Sciences*. 2nd ed. New York, Elsevier, p. 87–112.

Ocenianie tez epistemicznych w oparciu o dane eksperymentalne

Metodologia badań empirycznych zakłada konieczność decydowania o tym, jakie reguły mają być stosowane przy ocenie tez epistemicznych (*epistemic claims*, *knowledge claims*), które mają wyrażać pewną wiedzę o mechanizmach leżących u podstaw obserwowanych zjawisk, dokładniej, o powiązaniach między cechami obserwowalnymi a pojęciami i konstruktami teoretycznymi. Reguły takie wymagają określenia jasnych kryteriów wyboru odpowiednich miar i procedur. W artykule tym wyjaśniamy na czym polegają trzy sposoby, umożliwiające wysuwanie i ocenę tez epistemicznych: triangulację, łączenie wielu cech – wielu metod (*multitrait-multimethod*) i meta-analizę. Oceniamy każde z tych podejść i podajemy przykłady ich zastosowania w badaniach, by dojść na końcu do pewnych twierdzeń o związku, jaki zachodzi między wynikami eksperymentu a teorią i jej trafnością.

Słowa kluczowe: status, epistemologia, eksperymenty, metaanaliza, teoria

FOLIA 202

Annales Universitatis Paedagogicae Cracoviensis

Studia Sociologica VIII (2016), vol. 1, p. 96–106
ISSN 2081–6642

Jane Sell
Texas A & M University, USA

Murray Webster, Jr.
University of North Carolina – Charlotte, USA

The Importance of Cross-Cultural Experiments for the Social Sciences

Abstract

We detail how experiments differ from other types of research. In particular, laboratory experiments involve random assignment and control. They create artificial environments that are designed to test particular kinds of theories – those composed of concepts not bound to space and time. The aim of cross-cultural experiments is to ensure that theories are not tested in one cultural setting. Such experiments can help recognize how different cultures uniquely identify different initial conditions for theory development.

Key words: laboratory experiments, cross-cultural research, random assignment, initial conditions

Experiments and their characteristics

Cross-cultural research helps us find out whether and how culture affects basic principles of human behaviour. In this paper we examine some of the fundamental purposes and properties of experiments in order to determine when experimental cross-cultural research is most useful, when it is less useful, and how it can inform both theory and description.

There are many different methods of social science investigation: documentary-historical, survey, participant observation, and experiments. All of these methods can be used to test theories but they vary in the way in which the theory testing is accomplished. *Experiments* differ from other methods in several ways. First, the investigator by means of experimental 'manipulation' can endow the units of analysis (individuals or groups) with varying levels of the independent variable(s) to study the effects that inter-level differences are predicted to have on the dependent variable(s). Other methods do not create changes in independent variable(s), relying instead on observation or measurement in some settings. In the case of laboratory experiments, manipulation occurs within the artificial setting of a laboratory. In the case of field experiments, it can occur in different natural contexts such as schools, hospitals, and organizations.

Second, experiments utilize *random assignment* of treatments or conditions to the units of analysis, usually, people or groups. Randomization is an extremely important aspect of any kind of method as it serves to eliminate bias in the data. It often

takes the form of *random sampling* of respondents for surveys from some population or random sampling of articles for content analysis. In random assignment, there is nothing random about the participants; in fact, experimental subjects are often recruited for a given study because of having a particular characteristic. There is nothing random about cancer patients being recruited for a new cancer drug study, or university students being recruited for a study about learning techniques. The random assignment of treatments to individuals or groups is the random component of experimental method and its distinguishing mark. For example, the assignment of a placebo vs. a new experimental drug to cancer patients or the assignment of a visual vs. audio learning technique to students is done so that each patient or student has an equal (or known) probability of receiving each of the two treatments.

Because laboratory experiments are conducted in controlled conditions in which random assignment is feasible, they are *artificial*. Such artificiality also enables control or holding certain conditions constant. Control and randomization are powerful ways to eliminate alternative interpretations for the results of the experiment. Basically, this means that if a researcher has carefully designed the experiment, he or she can be certain that the outcome (a change in values of the dependent variable) occurs due to the independent variable.

While randomization eliminates an infinite number of alternatives for the results (such as irrelevant characteristics of individuals and what people are wearing), control eliminates alternatives associated with factors researchers know to make a difference. As an example, in experimental studies of group interaction, we know that the history of a group and of its members makes a difference. Consequently, we control on that history – often by creating groups that did not have prior interactions. As another example, we know that the information that the group members have about one another can dramatically affect the interaction process. When a person knows that the people with whom he or she is interacting share certain status characteristics, the interaction can be very different from that in the situation where the group members share none of the same status characteristics.

To eliminate alternative interpretations of the results, researchers must have a well-formed theory that pertains to the phenomenon/a under study. If the theory under investigation concerns the development of a lasting group dynamic that was set in motion by the very first interaction, the research design should enable defining the initial state of the process, as well as a series of interactions to allow the measurement of stability. At a very elementary level, this illustrates how a theoretical question must be carefully articulated before any design is considered as a way to answer the question.

Experiments can test particular types of relationships, which are expressed by *theoretical principles* of a particular kind, namely, those independent of space and time. This means that the concepts or terms used in these principles and thus the principles themselves are 'exact class.' The term *exact class* is taken from the logical distinction between concepts that are defined in such a way that their meaning never changes, and those that are 'ordinary' or changeable and malleable. This distinction has been articulated by Stephen Korner (1966) as a means to distinguish different approaches toward knowledge. Exact class concepts obey the law

of excluded middle because they are defined precisely so that it is possible to determine whether an event or object is or is not a member of a class of events. It either is or is not. There is no 'middle,' or 'almost.' Additionally, the meaning of such concepts never changes. Mathematical concepts such as 'triangle' provide a good example. A triangle is defined precisely and one can always determine if something is or is not a triangle. A square is not a triangle, in part, because it has more than three sides. The concept of triangle is the same in 2016 as it was in 1942 or 1900. No particular time or location is necessary for determining if an object is or is not a triangle. Hypotheses or propositions that incorporate triangles can be replicated precisely because they are not defined by a specific time.

On the other hand, concepts such as World War II, gender roles in 2016, and the Paris Climate Accord are ordinary terms. Their definition is absolutely conditioned by certain points in time and they are only understood in historical context. While we can, most certainly, study these concepts, we cannot replicate them because they are uniquely situated in history. Propositions or questions that utilize ordinary concepts are meant to capture history or historical change, and so are descriptive. Questions such as, 'what are Americans' attitudes toward Donald Trump in July of 2016,' are descriptive because they incorporate ordinary concepts.

While descriptive questions are critical for developing understanding, they are not appropriate for experimental tests. Experiments are artificial and as such, are poor instruments for studying particular events or even change. We would not bring large groups of Americans into an artificial setting to ask about their attitudes toward different political figures – there would be no need. This is a descriptive question. Random assignment or experimental control would be inappropriate. However, if we were interested in a snapshot of how Americans felt about political figures, such as Donald Trump, at a particular point in time, we would be concerned about *random sampling* to ensure that we could generalize to the existing population at this particular point in time. Experiments, on the other hand, can test principles utilizing concepts that do not lose their meaning in a certain context or time: concepts such as status, evaluations, public goods, task success, and behavioural constraints. These concepts, if defined exactly, can be used to examine events in the past or even in the future.

Scope conditions and initial conditions

The principles or propositions composed of theoretical or exact class concepts must be accompanied by *scope conditions*, or the parameters that delineate when propositions or hypotheses are expected to obtain. All theories have limited scope. No theory applies all of the time, and the scope conditions delineate, abstractly, when the theory being investigated is thought to apply. Newton's famous laws of acceleration of falling bodies apply only in the absence of factors such as air resistance or magnetic deflection; in other words, they take as scope conditions the absence of those factors. Expectation states theories, a prominent social psychological set of theories, describe how people in task groups organize interactions. Many of the formulations

delineate scope conditions requiring that task-oriented groups in which people meet have no history of interaction. Scope conditions are part of the theoretical specification of the theory or sets of propositions (See Cohen 2003; Walker, Cohen 1985; Foschi 1997; Sell, Martin 1983; Webster, Sell 2014). They too, are composed of concepts that are abstract because they are part of the theoretical formulation or foundation. They 'travel' with the derivations. So, if we were discussing expectation states propositions that involved scope conditions of task-oriented groups with no prior history of interaction, these conditions would follow through all the propositions and derivations of the theory.

It can be noticed that the theoretical principles mentioned are vastly different from descriptive principles that specify times and places. Because of this, these theoretical principles are artificial in the sense that they are not of our experience. Because these concepts and principles are artificial, the artificiality of the laboratory setting makes it ideal for testing. In this way, artificiality of experiments is not a disadvantage as it is sometimes mistakenly thought to be, but rather a distinct advantage, at least for dealing with theoretical principles. Other methods are not as artificial because they deal with complex everyday settings, such as actual organizations, institutions and groups in which variables cannot be easily disentangled and, of course, cannot be manipulated. Theory testing is certainly possible with the use of other methods, but causal mechanisms, in particular, are more difficult to disentangle than with the use of experimental methods.

At the same time, there must be a translation of the theoretical concepts to concepts that can be used in testable hypotheses. This process, termed *instantiation*, means that the theoretical concepts must be defined in ways that are measureable in a specific setting, in a specific place. This requires the experimenters (or other researchers as well) to develop what are usually termed 'operational' measures of abstract concepts.

Instantiations of scope conditions are termed *initial conditions*. While theoretical principles and scope conditions do not describe specific settings like a group of college students in Cracow, they certainly can apply to such settings. The key to such application is the use of initial conditions (Cohen, 1989, 2003; Foschi 1980, 1997; Webster, Sell 2014). They are specific to the setting and supply a 'starting place' for the testing of the principle. Because this is the case, the researcher must understand the cultural setting. For example, if we want to investigate how information about diffuse status affects performance within a group, we need to know what a diffuse status characteristic is in the particular culture (and the particular time) we are going to study. Or, alternatively, we could create a diffuse status characteristic, a topic which has already been investigated in a number of recent theoretical explorations (Webster, Hysom 1998; Ridgeway, Correll 2006; Ridgeway et al. 1998; Ridgeway et al. 2009; Ridgeway, Erickson 2000). A *diffuse status characteristic*, as defined within status characteristics theory, is a characteristic of an individual with at least two states differentially evaluated. With each state there are associated specific performance expectations, as well as general performance expectations. That is, those who possess the higher state of this diffuse status characteristic not only are societally defined as better at task performance than those who possess

the lower state of the characteristic, but those individuals are perceived as 'generally better' than those who have the lower state of the diffuse status characteristic. This term is exact class because it does not refer to specific time and place and is precisely defined so that it is possible to determine if a particular characteristic is or is not a diffuse status characteristic.

The instantiation or instance of a diffuse status characteristic can be very different in different contexts. While skin colour is a diffuse status characteristic in many cultures, including the United States, it is not always a diffuse status characteristic. So, for example in South Sudan in 2016, skin colour is not a relevant status characteristic but tribal membership is.

The point is that the world changes rapidly and we can explore the changes by asking descriptive questions that provide answers about the world in a specific time. However, such description may not help us in the future, because, by definition, ordinary concepts featured in description are tied to a particular time. Theoretical principles, on the other hand, because they are not tied to a particular time, are designed to generalize and apply to the future.

Cross-cultural exploration

If experiments are designed to test theoretical principles, how are tests across cultures important? John Darley, in his Presidential Column in the newsletter of the Association for Psychological Science (2001), noted the importance of careful experimental design to enable causal inference. He stated that a basic psychological approach is to discover 'truths of human functioning that transcend culture and context.' But, 'Unfortunately, a nasty thing happened on our way to our universal generalizations: culture and context turned out to have a much more fundamental effect on our generalizations than we expected.' (Darley 2001. p. 3) Darley's statement should not be taken as an argument against the development of generalizable, theoretical principles. It can be read as an argument for more cross-cultural investigations.

If experiments are only tested in one context, for example, in a college context in predominantly White institutions in the United States, then results run the risk of support only in that particular context. At least their applicability to other settings is undemonstrated. And, indeed, psychology has been called to task by Arnett (2008) who investigated six of the APA journals and found that 'research in major APA journals is concentrated on a narrow range of the world's human population, principally Americans' (Arnett 2008, p. 609).

Replication is always important; indeed, it is one of the most important defining characteristics of science. But replication in very different contexts is especially of high value because it gives the researcher more confidence that the general principles apply even given quite different initial conditions. So cross-cultural replication is *particularly* valuable because it is a low probability event by chance alone.

Examples of such experiments include those that do replicate and do not replicate: both are important. For further discussion of this point, see Foschi 1980.

Gachter, Herrmann, and Thoni (2010) demonstrate an example of cross-cultural experiments that illustrated problems with an accepted paradigm. Basically, economic models (*Homo economicus*) suggested that cultural background should not matter because, in those models, selfishness is universal. If that is the case, then results from studies in which there is a clear prediction based upon self-interest should apply all the time, in all cultures. But, recent evidence from public goods experiments shows that is not the case.

Public goods games are a staple of economics literature; they are one of four canonical games or experimental investigations in economics (Eckel 2014). They are important to many disciplines including sociology, psychology and political science. Public goods create an individual level dilemma because individual incentives conflict with overall group incentives. While it is an actor's self-interest to *not* contribute to a good, if no one contributes, the public good will not be provided and all will be worse off. An example might be developing a public park. All could benefit and enjoy the park, independently of whether each person had contributed. But, according to the basic logic of rational choice, individuals would recognize that their own contributions are unnecessary. It is rational to not contribute, that is, to free ride. Because that is the case, traditional economic models suggested that intervention, in the sense of a central government or institution, was necessary to create and maintain public goods.

A particular *experimental* paradigm for the investigation of public goods has been developed over the years. Group members are given tokens that they can either give to the public good or keep in an individual fund. At every point in time, the value of a token kept in their fund is worth more than a token put into the public good fund. Additionally, any tokens contributed to the group fund are distributed to all members, regardless of whether or not they contributed. So, basically, the actor's best strategy is to never contribute and hope that everyone else does contribute. Of course, if everybody is thinking the same way, nobody will contribute and the public good will not be provided.

As mentioned, if the traditional economic model adequately predicted behaviour, no group member would ever contribute. Consequently, economists conjectured that there should be little if any difference across cultures because while people might make errors in judgment or perhaps be of different 'types' or personality differences tied to altruism, culture should not impact their decisions.

Herrmann, Thoni, and Gachter (2008) conducted public goods experiments in 16 different subject pools and six distinct cultural areas around the world. They did find variation however:

> Our main findings are that cooperation within cultures is largely similar while there exist highly significant differences between cultures. This is true in public good experiments with and without punishment and also holds for punishment behaviour. This dual observation of within-culture similarity and cross-cultural heterogeneity is the main support for the claim that there are cultural influences on cooperation. (Gachter, Herrmann, Thoni 2010, p. 2653).

In other words, the claim cannot be made that people respond in the same self-interested way across cultures (see also the discussion in Henrich et al. 2001).

Cross-cultural description?

As we discussed above, the strength of experiments is the testing of theoretical principles, which employ exact class concepts. Experiments are not suitable for describing a culture or the individuals who participate in experiments. Experimental data are the dependent variables of the experiment, which, as noted earlier, will be defined independently of time and place. People who participate in an experiment are not selected to represent any natural group. Rather, they are selected because it is possible to instantiate certain theoretical concepts in them. For instance, a group of young people differing on age might be selected because, for them, age meets the theoretical definition of a status characteristic. The theoretical principles under test describe how status affects behaviour. If age is a status characteristic for a particular population, then the principles can be tested by observing their behaviour. But in another time and place, age might not be a status characteristic, and then the theory would have little to say about how an age-differentiated group might behave. Experiments create artificial settings to more precisely test principles. Because they use control and random assignment, rather than trying to duplicate elements of the setting, or using random sampling, they are not useful for description.

As mentioned above, Hermann, Thoni and Gachter conducted multiple replications of public goods studies that demonstrated large amounts of cultural variation. This called into question the universality of the economic principles of self-interest in public goods settings. In 2008 and then in 2010, researchers took a different strategy and tried to interpret findings in terms of culture. They divided up the cultures into different kinds of classifications to find what factors might lead to the differences in experimental results. So, the authors suggested that 'punishment may be related to social norms of cooperation,' (Herrmann, Thoni, Gachter 2008, p. 1365). To determine this, they constructed two variables. One, norms of civic cooperation, was developed from the World Values Survey and was based upon how people feel about tax evasion, benefit fraud, and avoiding paying for public transport. A second measure was a 'rule of law' indicator based on the degree to which people abide by and believe in the rules of society including the police and the courts. Researchers reasoned that if these indicators are views of the average citizen, they also typify the participants in different cultural contexts or countries. Then they ran different analyses and found that the classification of norms of civic cooperation was related to punishment in the public goods experiments but not to the rule of law differences across cultures.

The researchers believe that the experimental results combined with the classifications tell us about the norms in different cultural settings. But this is really not clear. The questionnaire measures, if they involve random sampling, are most likely to be measures of norms for different societies. But how exactly those norms relate to behaviour in the artificial, laboratory settings is not developed. It is very unlikely that the researchers are interested in a *description* of how people in an experimental

laboratory, interacting over a computer, contribute and then levy costs (punish) others that they cannot see. Are both the questionnaire information and the experimental information tapping the same norms? If they are, then what basic principles are being tested?

In other words, experiments can be designed to see how different kinds of contexts affect the dependent variables. They can do this by utilizing control so that differences among or between conditions can be attributed to the independent variable. But when the independent variable is country, there are too many factors operating to be sure what is causing what. Experiments cannot be used to characterize a population. Populations can be characterized by methods specifically designed to measure characteristics of interest that usually employ relatively large random samples.

But how then can country or culture be incorporated into experiments? Replication is one important way. Replication would mean that the same principles are being tested through experimental designs in different cultures, not to describe the cultures, but rather to see if the same theoretical principles are upheld. To test of how framing of a social dilemma changed the degree of cooperation among group members even *when the payoffs were exactly the same*, Sell et al. (2002) conducted experiments in both the United States and China. In both countries, participants were randomly assigned to two different conditions (a 'give some' public good or a 'refrain from taking' public good) and payoffs were calibrated for each country so that they were approximately equal in meaning. Such a design is equivalent to blocking on variables for which there are good reasons to expect differences. Such blocking designs are common when investigating gender, or race/ethnicity, for example. It is obvious that we cannot assign participants an ethnicity or a country, but we can block or control on the variable. In the case of the Sell et al. (2002) experiment, participants from both countries were affected in the same way by the framing and were more cooperative in the 'refrain from taking scenario' vs. the 'give some' scenario.

Another way that cross cultural experiments can allow comparison is by ensuring that there is a type of baseline condition that measures the 'initial condition' of one culture versus another. It would not tap a genuine 'norm' in society but, importantly, it would be a measure that would allow estimation for how different theoretical changes would affect the behaviour. That is, it would serve as the beginning place for the particular experimental design. In the study of cooperation in public goods and the effects of group membership on cooperation for example, a baseline of no information about group members, would function as a 'cultural calibration' measure. This measure would then be used to determine whether and how different kinds of group membership changed initial levels of cooperation. In this way, the no information condition measures behaviour for the specific experiment, not the society as a whole. Societal norms are driven by context, and the experimental context is peculiar – it is artificial.

Oh (2013) provides an example of a cross-cultural experiment that used both initial conditions to gain an estimate of cultural differences within the experimental context, as well as a replication of theoretical (exact class) conformity principles. In this study, Oh investigated whether participants from a collectivist culture of India would conform to groups similarly to the participants from the individualist culture

of the United States. To test his conjectures, Oh designed three experiments. The first two of them were preliminary experiments designed to function as initial conditions. In these studies, Oh first tapped each of the groups' individual judgments about different choice dilemmas and different opinion items. He then used these items to generate arguments that might be posed by groups in different cultures, as well as to assess the 'cultural starting place' for this particular experimental context. As an additional baseline, he conducted a second study to determine how individual responses might change from one point in time to another for participants in each culture. The third experiment tested how participants might respond to 'mere exposure' of the opinions of groups who share their identity, or how participants respond to persuasive arguments presented by group members. Oh finds that the effects of group argumentation are the same across cultures while 'mere exposure' effects are somewhat stronger in India.

Oh's study on conformity is an excellent example of the potential strengths of cross-cultural experiments. He uses the culture as 'block variable' and he measures the initial conditions for each of the cultures, to estimate the effects of the manipulation. The manipulation is a test of the concept of conformity, a concept that is not ordinary because it does not relate to any particular context. Instead, it refers only to the change from an individuals' initial choice produced through exposure to the group.

The studies discussed illustrate that experiments are not effective at tapping descriptive properties of a particular culture; that is, they cannot adequately capture cultural norms or common characteristics of entire populations. Experiments *are* effective at testing general principles and this can be done by carefully creating baseline conditions to assess cultural 'starting places' or by replicating studies through blocking (usually by country) and random assignment.

Conclusion

Experiments are powerful methods for testing theoretical principles. They enable the manipulation of independent variables to determine the effect upon the dependent variables. They do this by creating artificial settings that enable control and randomization. Cross-cultural experiments can be important for two purposes. First, they provide important replications. Replication across cultures is especially valuable because it demonstrates that general principles apply even in very different contexts and initial conditions. Secondly, cross-cultural experiments can explore how general principles are affected by the culturally specific initial conditions.

What cross-cultural experiments cannot do effectively is describe the characteristics of cultures. Because experiments are artificial, they are not adequate methods for describing what exists in settings that are time and space specific. Experiments cannot describe the norms for littering in Poland in 2016, or the attitudes of voters in Texas in 2016. Other methods, however, can be fruitfully employed for such investigations.

References

Arnett J.J. (2008). *The Neglected 95%: Why American Psychology Needs to Become Less American*. American Psychologist, 63, p. 602–614.

Cohen B.P. (2003). *Creating, Testing, and Applying Social Psychological Theories.* Social Psychology Quarterly, 66, p. 5–16.

Cohen B.P. (1989). *Developing Sociological Knowledge: Theory and Method*, 2nd ed., Nelson-Hall, Chicago, IL.

Darley J. (2001). *We Fail to Contribute to Policy Debates*. APS Observer, 14, p. 3–40.

Eckel C. (2014). *Economic Games for Social Scientists*, In: M. Webster Jr., J. Sell (eds.), *Laboratory Experiments in the Social Sciences*, 2nd ed. London: Elsevier, p. 335–356.

Foschi M. (1997). *On Scope Conditions*. Small Group Research, 28, p. 535–555.

Foschi M. (1980). *Theory, Experimentation, and Cross-Cultural Comparisons in Social Psychology*. Canadian Journal of Sociology, 5, p. 91–102.

Gachter S., Herrmann B., Thoni C. (2010). *Culture and Cooperation*. Philosophical Transactions of the Royal Society of London B: Biological Sciences, 365, p. 2651–2661.

Henrich J., Boyd R.,. Bowles S., Camerer C., Fehr E., Gintis H., McElreath R. (2001). *In Search of Homo Economicus: Behavioral Experiments in 15 Small-Scale Societies*. American Economic Review, 91, p. 73–78.

Herrmann B., Thoni C., Gachter S. (2008). *Antisocial Punishment Across Societies*. Science, 319, p. 1362–1367.

Korner S. (1966). *Experience and Theory*: *An Essay in the Philosophy of Science*. London: Routledge and Kegan Paul.

Oh S.E. (2013). *Do Collectivists Conform More than Individualists? Cross-Cultural Differences in Compliance and Internalization.* Social Behavior and Personality, 41, p. 981–994.

Ridgeway C.L., Backor K., Li Y.E., Tinkler J.E., Erickson K.G. (2009). *How Easily Does a Social Difference Become a Status Distinction? Gender Matters*. American Sociological Review, 74, p. 44–62.

Ridgeway C.L., Boyle E.H., Kuipers K., Robinson D.T. (1998). *How Do Status Beliefs Develop? The Role of Resources and Interactional Experience.* American Sociological Review, 63, p. 331–350.

Ridgeway C.L., Correll S.J. (2006). *Consensus and the Creation of Status Beliefs*. Social Forces, 85, p. 431–454.

Ridgeway C.L., Erickson K.G. (2000*). Creating and Spreading Status Beliefs*. American Journal of Sociology, 106, p. 579–615.

Sell J., Chen Z.Y., Hunter-Holmes P., Johansson A. (2002). *Cross-cultural Comparison of Resource Goods and Public Goods*. Social Psychology Quarterly, 65, p. 285–297.

Sell J., Martin M.W. (1983). *An Acultural Perspective on Experimental Social Psychology*. Personality and Social Psychology Bulletin, 9, p. 345–350.

Walker H.A., Cohen B.P. (1985). *Scope Statements: Imperatives for Evaluating Theory.* American Sociological Review, 50, p. 288–301.

Webster M., Jr., Hysom S.J. (1998). *Creating Status Characteristics.* American Sociological Review, 63, p. 351–378.

Webster M. Jr., Sell J. (2014). *Why do experiments?* In: M. Webster Jr., J. Sell (eds.), *Laboratory Experiments in the Social Sciences*, 2nd ed. New York: Elsevier, p. 5–22.

Znaczenie eksperymentów międzykulturowych dla nauk społecznych

Wyszczególniając czym eksperyment różni się od innych typów badań, wskazujemy na znaczenie kontroli i randomizacji w eksperymentach laboratoryjnych oraz potrzebę tworzenia sztucznych środowisk, w których możliwe staje się sprawdzanie szczególnego rodzaju teorii – teorii zbudowanych przy użyciu pojęć nie ograniczonych czasem i przestrzenią. Celem eksperymentów międzykulturowych jest zapewnienie, by testowanie teorii nie było przeprowadzane tylko w jednym układzie kulturowym. Takie eksperymenty mogą także ułatwić poznanie swoistych dla różnych kultur warunków początkowych (szczególnych realizacji warunków zakresowych), co służy dalszemu rozwijaniu teorii.

Słowa kluczowe: eksperymenty laboratoryjne, badania międzykulturowe, randomizacja, warunki początkowe

FOLIA 202

Annales Universitatis Paedagogicae Cracoviensis

Studia Sociologica VIII (2016), vol. 1, p. 107–116
ISSN 2081–6642

Iza Desperak
University of Lodz, Poland

Can Socio-Cultural Context Affect Experimental Results? The Case of the Zimbardo Prison Experiment Repeated in Poland by Artur Żmijewski

Abstract

The famous Phillip Zimbardo's study (1971), known as the Stanford Prison Experiment, was repeated some ten years ago in Poland by Artur Żmijewski, a video artist. His findings deserve the attention of social psychologists doing experimental research. The video released by Żmijewski in 2005 shows that his action ended with a completely different outcome than that of the original experiment: the participants themselves decided to stop it. The difference may have been a consequence of the artist's unconcern about full conformity with the methodological rules the academics consider necessary to follow in conducting experiments. However, the impact of a particular socio-cultural context provides another plausible explanation of why Żmijewski's results depart from those obtained in the original study. Polish culture differs in a number of dimensions from the culture that was taken for granted by Zimbardo in creating the Stanford experimental setting and informed his interpretation of the results obtained therein. This note offers an account of Żmijewski's project, as well as an explanation of its results in terms of cross-national comparisons. The author's aim is to provoke more discussion on the role of socio-cultural context in experimental research.

Key words: Phillip Zimbardo, Stanford Prison Experiment, Artur Żmijewski, socio-cultural context, dimension of culture

Introductory remarks

The Stanford Prison Experiment is a classical example of using experimental research in order to learn more about human nature. In his book *The Lucifer Effect: Understanding How Good People Turn Evil* (2007), Zimbardo described 'more than 30 years of research on factors which can create a "perfect storm" that leads good people to engage in evil actions' (Zimbardo 2015). He concluded that *all* humans can undergo such a transformation to which he gave the name of the *Lucipher effect*.

His conclusion, based on the results of his 1971 experiment, can be questioned, however, as it is inconsistent with the results of the repetition of Zimbardo's experiment by Artur Żmijewski in Poland. His project has not yet become known to academic scientists. Żmijewski is not a researcher but a video artist. His aim was not to carry out a strict *replication* (in the meaning this term has in the methodology of experimental research) of the classical study. His results, however, should enter academic discourse, even though the author did not publish any report in any

scientific journal. Instead, a piece of video art he produced from the recording of the experiment was shown at the 51th Biennale in Venice, giving the author an opportunity to share his product with a wider public. Żmijewski also informed Zimbardo on his repetition and its results and Zimbardo's answer was made public. It is high time for researchers to get acquainted with that unusual repetition and its results, and reflect on factors which could have brought about such a surprising effect.

The fact that a performer and video artist tried to repeat Zimbardo's experiment could be seen as no more than a curiosity. Actually, the artist designed his performance with the intention to strictly repeat the conditions of the Stanford Prison Experiment. It is worth mentioning that Zimbardo himself complimented the artist on that. Even though Żmijewski did not plan a proper replication of the experiment, what he obtained deserves more discussion and analysis because of his results are so distinct from the original.

The outcome observed by Zimbardo led him to define the Lucipher effect. Żmijewski's experiment shows that this effect can be overridden by the *solidarity effect,* or the effect of cooperation emerging between two groups of participants playing quite different roles in the experimental social system. That is, the effect observed by Zimbardo is not as universal as it is generally believed to be, but it can occur in certain circumstances due to some factors whose nature has yet to be disclosed.

Artur Żmijewski and his activities

Artur Żmijewski, a video artist and performer, is a representative figure of Polish *critical art*. According to Izabela Kowalczyk (2002), this important movement has been one of the first discourses critical of the *transformation* that Poland experienced after 1989. Some artistic activities served as a strong tool of social criticism, or even formed a sort of political declaration directed against some practices. In her book devoted to Polish critical art, Kowalczyk included a full chapter (Kowalczyk 2002, p. 275–298) about Żmijewski and his art before *Repetition.* His early works focused on human body, the theme of the Other, and social traumas. After *Repetition* the artist realized various works and organized several politically and socially engaged projects. He has been involved in the activities of the leftist think tank *Political Critique* (Krytyka Polityczna). Some of his videos may be watched on the website (http://artmuseum.pl/en/kolekcja/artysci/artur-zmijewski) of the Museum of Contemporary Art (Muzeum Sztuki Nowoczesnej) in Warsaw.

Repetition, not replication

In 2005 Żmijewski represented Poland at the 51st Art Biennale in Venice. He decided to show a video documenting his repetition of the experiment conducted in 1971 by Zimbardo. Zimbardo's simulated prison, with its architecture and rules, was recreated in Polish reality. The persons who were assigned the roles of guards and inmates were recruited by the artist from the unemployed men. He selected seven prisoners and nine guards by means of a procedure that involved psychological

tests and examinations to eliminate mentally unstable candidates. The rooms were equipped with Venetian windows (one-way mirrors) to enable filming of the course of the experiment with the use of five manually operated cameras and several night-vision industrial TV cameras. Other participants of the experimental setting included: psychologists acting as experts with the right to stop the experiment if things turned dangerous, a former prison inmate, and a sociologist who had been involved in reforming the Polish prison system.

The experiment lasted for seven days. While making the *Repetition*, Żmijewski, not for the first time, turned to the aesthetics of violence and segregation. He was aware of the contemporary context of his artistic actions. The Polish repetition took place just after the torturing of prisoners in Abu Ghraib came to light but before the perpetrators were sentenced.

In his experiment, two opposite groups, the guards and the prisoners at some moment quite unexpectedly joined hands and rebelled against the institution of prison, the artist, and the very situation he had put them in. Żmijewski commented on this in the followings words:

> Repetition suggests that what people most ardently strive for is a compromise. People don't keep torturing one another until the conflict is solved. They search, rather, for a safe status quo, negotiate, and act opportunistically (http://culture.pl/en/artist/artur-zmijewski)

Żmijewski does not use the term *replication* nor does he pretend that he is a scholar. He does not quote scientific publications nor does he use psychological jargon albeit he must have studied some literature on the Stanford Prison Experiment. His work consists in a creative, free use of scientific procedure to produce a piece of video art to be shown to the public. His aim was not to submit a research report for publication in a regular scholarly journal. A critical analysis of his product seems counterproductive. Even Zimbardo's answer to Żmijewski's note ignores its potential academic relevance. Nevertheless, the results obtained by the artist playing with the experimental method call for attention and reaction of the academia.

Joanna Tokarska-Bakir (her reflections are quoted here after the text available at http://culture.pl/en/artist/artur-zmijewski), a prominent Polish cultural anthropologist who has been active in hot public debates, suggests that Żmijewski's work should be interpreted metaphorically, not as a repetition of an old experience, but rather as a new opening of the *space of social evil*. She placed Żmijewski's work in this context in accordance with a later reinterpretation of the original experiment by Zimbardo himself. She also noticed that the volunteers who agreed to take part in the Polish experiment, including those who became the guards and even their head, did not trust authority nor did they display full identification with their roles.

This is the first reason to ask the question of how different the two cultures in which the experiment was originally conducted and repeated were. The level of social trust in Poland is one of the lowest in Europe and certainly lower than in the American society. The same can be said on the attitude toward authority. That is why one should take into account the socio-cultural context in the analysis of Żmijewski's repetition.

Video content (summary)

As the video is not available online at the moment, I am going to give a short summary of it. It is not a detailed transcript, but an overview of its content. The focus is on the last stages and the very moment when the remaining participants decided to stop the experiment.

Żmijewski's video, which is 1 hour 15 minutes long, is for now the only documentation of his experiment released by the author. The video presents the course of the experiment as a sequence of selected happenings which took place on particular days or hours. The sequence starts from the scene when one of the participants is led to the place of his destination, with his eyes covered. Another scene shows how the participants who will be playing the roles of prisoners are changing into prison clothes: long shirts (like those in Abu Ghraib?) with numbers printed on them. Next, the prisoners' photos are being taken with the numbers displayed in front of them. Such an opening of the video is followed by informing the viewers that:

> We recreated the conditions of Professor Zimbardo's experiment in order to investigate how an oppressive situation influences people's behaviour. Do they turn into torturers and victims? Or are they able to resist the temptation of ruthless exercise of power? Both the guards and prisoners are paid the equivalent of $40 a day.

The number given at the end of this statement can be interpreted in the context of the economic conditions of living in Poland at the time when Żmijewski carried out his *Repetition*. In 2005 the minimum subsistence level for a single person was $145 monthly, or around $5 daily, which is 8 times less than the amount that a participant would earn for each day in the experiment. The average monthly salary at that time reached $612, or some $20 per day, twice less than the reward for a day spent in prison. The prospect of earning a significant amount of money in a relatively short time might have appealed to the volunteers, as they were recruited from the unemployed for a long time. Some participants declared that they joined the project because of curiosity or interest in its cognitive aspect. One of them – asked if he would be ready to do everything to earn some money – refused to participate. Nevertheless, the financial aspect of participation cannot be considered entirely unimportant.

As the story unfolds before the eyes of the viewers, they are watching what is going on in the prison as it will be slowly developing its daily routine. They can observe the prisoners and the guards, as well as the committee of organizers debating about the rules and communicating their decisions to the guards. When the guards prepare and announce first regulations, one of the prisoners comes up with a question, and the performance starts out.

Once the prisoners have been assigned numbers, they are told to call one another with these numbers and to use them while reporting to the guards. The prisoner who has asked the question got the number 810; he will become one of the main characters of the story. He talks a lot, appeals for solidarity, but his behaviour results only in all of the group members' being punished.

The story goes on. When the inmates misbehave, the guards discuss among themselves how to control them, trying to contrive new punishments. Some punishments, like a limitation of walks or visits, play only a symbolic role, but later the guards will take pains to invent more severe measures to discipline the inmates, like using the truncheon, limiting sleep or access to the toilet, etc.

On the 5th day one of the participants (playing the role of a prisoner) reports his will to abandon the experiment. He pronounces aloud his real name (in full), confirms his decision and says that he does not want to give any reasons for it. We see him leave the cell and move to the room where he can change to his own clothes; lastly, we witness his talk to one of the organizers.

This participant who leaves the prison comments on his feelings, saying 'Thanks a lot, I feel like a shit' (in Polish 'chujowo'). He is not the first one to leave. Out of twenty participants at the end only three prisoners and few guards remain. Next day (the 6th day of the experiment) another participant playing the role of a prisoner leaves the experiment, and two guards are discharged.

The rest of the crew decides to press ahead with another torture. They order the prisoners to cut their hair with the trimmer. Who can use this instrument? Prisoner number 810, the one who started all the protests, comes forward. He shaves his own head, but other prisoners disagree to cut or shorten their hair. The guards punish all the prisoners by not letting them sleep at night. In the morning (Day 7) number 810 talks to the chief warden and offers to him to encourage the prisoners to have their hair cut. After the approval he encourages one of fellow prisoners to allow him to conduct the operation. The other agrees, but only to shorten his hair on the back and by the sides. Number 810 starts the job. He begins by cutting the hair on the agreed-on parts of the head, but suddenly by mistake he cuts the hair from the middle of the scull, so that to complete the task he has to cut all the hair.

On the same day the guards communicate that the sewage system broke down and the inmates will not be able to go to the toilet. They are given metal buckets to be used inside their cells. They do not like it, but later they will play with the content of the buckets.

The guards return to their space. They are reading the note one of them has just got. It is a cramped piece of paper with the communication: THEY ARE GETTING READY TO TAKE OVER. LOCK THE DOORS.

In the afternoon, there arises a conflict over emptying the buckets, making the guards embarrassed. The chief guard is talking to the organizer (Żmijewski?). When he is criticised for his poor efficiency, he comes back to his office and takes the poster with the prison regulations. He reads the rules point by point, marking them with a tick or circling, and adding something at some points about violations. Finally, he writes on it: STATE OF CRISIS. The guards start talking on the crisis, aware of the need to solve the problem.

In the following scene the guards are writing something on a piece of paper, a proposal of new regulations perhaps. We can see words: CORRECTIONAL FACILITY, REHABILITATION, PRISONER, GUARD, WARDEN, DIGNITY, GOLDEN MEAN.

In the next scene, the inmates, staying in their cell, see the same list written on a piece of paper. Finally, they are going to leave the cell. In the following scene we see them sitting in a circle together with the guards. In the end, a guard asks the prisoners if they want to return to their cells. Prisoner 810 answers: 'I am leaving without giving my reasons,' and starts taking off his prison uniform. The guards and other prisoners follow him. After that they shake hands and introduce themselves with their first names.

The next scene shows an after-experiment meeting of the former participants, talking about what they have experienced. The video includes also fragments of a conversation (conducted later) with a participant about the effect of the experiment on an individual.

The letter to Zimbardo and his reply to Żmijewski

After the experiment Artur Żmijewski informed Philip Zimbardo about his project and received a reply from him on 7th October 2005. Zimbardo congratulated him on a good replication of the prison environment and praised the author of the video for the artistic values of his movie. At the same time he pointed to the lack of important details in the description of Żmijewski's project. Indeed, we do not know how the recruitment was organised, how many participants played the roles of guards, whether the participants actually lived closed in the simulated 'prison.' He also criticised the organiser for his influence on the interaction process going on within the experiment setting (one of the guards was chosen to play a special role, that of the experimenter's confederate?).

Zimbardo attributed the outcome different from the one he had obtained himself to the decreasing number of participants, including guards. Finally, he stressed that in the Polish study it was much easier to get out of the experimental social system than it was in the Stanford experiment. According to him, it was that difference that had such a strong effect on the results of Żmijewski's repetition of the original experiment. He also informed about exchanging emails with one of the participants (called 'Czarny Maniek') and suggested using the term 'replication' instead of 'repetition' (Zimbardo 2005). To me, his response seems fairly superficial and a bit patronizing. In fact, his comments lack an in-depth discussion of the results and their prospective significance.

Discussion of Żmijewski's 'Repetition'

As I have already mentioned, Żmijewski is not a scientist but an artist. His intention was to create a work of art that would impress the viewers. Had he designed his 'experiment' with the aim to learn the patterns of human behaviour that arise in social systems characterized by granting to one group extreme power over the other, he would have to create an experimental setting in which all the methodological requirements are fulfilled. However, he did not care too much about it.

However, there is another plausible explanation for why the artist faced unexpected resistance of the actors refusing to play the game he had told them to play as long as

possible. Such an outcome of the group process he had initiated may well have been a consequence of a different socio-cultural context in which his experiment was done.

To create an experimental prison and set it in motion by assigning *actors* to two *social positions* (guard and inmate) in the system, Żmijewski used human beings with a definite cultural background. They had undergone the *socialization process* in the country where *respect for authority* had never been positively valued.

The roots of this cultural pattern lie in the history of Poland, making the inhabitants of their country allergic to any authority. When the Polish state more than a hundred years ago was divided into parts under the control of three much stronger neighbouring states, the Poles who cooperated with foreign authorities – by taking positions allowing them to exercise legal power – did not enjoy the respect of their compatriots. For the same reason, over the whole post-war period of Soviet domination, which followed a short-lived independence (restored in 1918) and over 5 years of German occupation, the attitude of a typical Polish citizen toward his or her superiors or supervisors (in particular, those one has to obey in prison or a similar social institution) was characterised by low *respect* for and *distrust* of authority.

This situation might be diagnosed by resorting to the concept of *homo sovieticus*. Even though the concept, which has been applied to the Soviet society, can no longer be used to describe social attitudes in contemporary Poland, some elements of this model still persist in this country (Tischner 1995, p. 205). In fact, Polish society is often pointed out as a unique example of a society with the lowest level of social trust (Giczi, Sik 2009). In this respect it differs significantly from the society in which Zimbardo carried out his experiment.

There may exist other socio-cultural factors responsible for the differences, too. To enquire into the matter more thoroughly we would need to compare the results of repetitions done in many different socio-cultural contexts.

Another track to be followed to understand Polish results is the recruitment procedure that was used by Żmijewski to find volunteers for his 'repetition.' As Zimbardo points out, no detailed description of the recruitment procedure was given by his Polish colleague. Fortunately, we have some information on the subject, so a comparison is possible. In the case of Stanford Prison Experiment the participants were recruited from among students. In Poland it was a group of unemployed men representing various professions from qualified workers (an electro-mechanic) to an actor. Even if they were not older (we do not know anything about their age, we can only guess it from the video recording) than the Stanford participants, they certainly shared many traits that distinguished them as a social category from American students in 1971. They differed in social status, future prospects, and, last but not least, their employment experience. Those who were jobless for a long time probably experienced more or less serious economic and financial difficulties. If so, they should have been more economically oriented than American students. Each day they spent in the experimental 'prison' they earned an amount of money that probably counted in their daily and monthly budgets. If that was the case and the financial motivation were to play a more important role for Polish participants, they should be more willing to stay in the experiment longer than American students. But they still decided to leave the experiment – so the

economic factor, even if it certainly had a strong impact on the situation, in the last resort, did not count that much.

Another factor that might have produced the result different from that obtained in the original experiment was that Polish participants, apart from having a richer life and job experience, may have brought into the experimental situation their strong commitment to certain *general values* concerning work and employment. Polish workers might also have been faithful to certain more specific values that achieved a prominent position within Polish national variety of the working class ethos. 'Solidarity' movement, which had grown out of Polish cultural tradition and became its important component itself, stressed the importance of non-economic rights of the workers and political freedom with emphasis on the necessity to respect unalienable *dignity* of any employee. This value was incorporated into Polish cultural heritage due to the teaching of John Paul II on the 'dignity of human work' and the 'dignity of the subject of work.' The Pope's teaching was presented in his encyclical *Laborem exercens,* which was issued 14th September 1981 just before the second part of the 1st National Congress of the "Solidarity" trade union.

There are more cultural differences one can suspect of bearing on group processes in laboratory systems created within larger social systems having each a definite cultural identity. The aim of the study conducted under the schemes of European Values and World Values Surveys was to examine value differentiation across a number of different cultures. The fundamental work of Geert Hofstede (2001) originated in a simple observation that two groups of candidates for a position in a company (one group coming from the USA, the other from the Netherlands) differed with their attitude toward two recruitment procedures because they represented different cultures with different sets of values, even though both sets functioned within the same cultural circle dominated by mainly protestant values. This difference was disclosed when the recruitment procedure was being implemented. The effect of the discovery was launching a big research project, aimed at measuring cultural differentiation – first, on the level of a company, finally, on the global level (Hofstede 2001). Moreover, the project inspired further research that led to defining the concept of *dimension of culture*. The concept was introduced to render how different cultures differ in their preferable values. Six value dimensions were finally detected by means of factor analysis: *power distance*, *individualism*, *masculinity*, *uncertainty avoidance*, *long term orientation*, and *indulgence*.

According to The Hofstede Centre, Polish and American value systems differ considerably in most of these 6 dimensions. The largest difference (+47) between the two cultures occurs in the *Uncertainty Avoidance* dimension. Poles are more prone than Americans to avoid conflict situations such that the outcome of the interaction process is hard to predict, so the two sides are *uncertain* about possible effects of their actions.

If the guards did not follow the instructions of their supervisor (his intention was apparently to intensify 'class struggle') and refrained from inventing new punishments to increase efficiency and extent of their power over the prisoners – in other words, if they did not keep them in constant uncertainty about the next torture – the experiment would certainly last much longer. Note that the guards were also motivated to

hold the level of coercion in reasonable limits, for another reason: to avoid their own uncertainty about what would be the response of their victims to new disciplinary measures.

Table 1. Differences in cultural values between United States and Poland

Country / Dimension	United States	Poland	Difference PL–US
Power distance	40	68	+28
Individualism	91	60	-31
Masculinity	62	64	+2
Uncertainty avoidance	46	93	+47
Long term orientation	26	38	+12
Indulgence	68	29	-39

Source: The Hofstede Centre, http://geert-hofstede.com/countries.html

'What people most ardently strive for is a compromise.' It is the interpretation of the results given by the artist himself. It turned out consistent with our explanation invoking Hofstede's theory of multidimensional values systems.

Conclusion

To conclude, there are many variables which might have been at work in the experimental system created by the Polish artist trying to replicate the findings obtained by the famous American social psychologist. Some factors pertain to the very organization of the experiment. We do not have the necessary information to assess their importance. The position of the experimenter and his actions should not be ignored, either. In addition, even if he stays in hiding, it is clear that he has triggered off the interaction process. We may also suspect that he may have intervened in its course, say, by sending to the actors subtle cues prompting the 'solidarity' solution of the conflict situation.

Lastly, socio-cultural values varying across national cultures may have appeared the main factor responsible for diverging experimental results. It is not my intention, however, to offer a definite answer to the question of which factor played the decisive role but to invite social psychologists to a debate that might result in reconciling alternative explanations.

References

Giczi, J., Sik A. (2009). *Trust and Social capital in Contemporary Europe, Tarki European Social Report,* http://www.tarki.hu/en/research/european_social_report/giczi_sik_eng_2009.pdf

Hofstede G. (2001). *Culture's Consequences: Comparing Values, Behaviours, Institutions, and Organizations Across Nations*. Thousand Oaks, CA: SAGE Publications.

The Hofstede Centre, http://geert-hofstede.com/countries.html

Kowalczyk I. (2002). *Ciało i władza. Polska sztuka krytyczna lat 90.* Warsaw: Sic!

Tischner J. (2005) *Etyka solidarności oraz Homo sovieticus.* Kraków: Znak.

Wojtyła K. (Johannes Paulus II) (1981). *Laborem exercens.* http://w2.vatican.va/content/john-paul-ii/en/encyclicals/documents/hf_jp-ii_enc_14091981_laborem-exercens.html

Zimbardo P. (2015) *The Lucifer Effect by Philip Zimbardo.* http://lucifereffect.com/

Zimbardo P. (2007).*The Lucifer Effect: Understanding How Good People Turn Evil*. New York: Random House.

Zimbardo P. (2005). Commentary to 'Powtórzenie/Repetition' by Artur Żmijewski, http://www.obieg.pl/teksty/5776

Czy kontekst społeczno-kulturowy może wpłynąć na wyniki eksperymentu? Przypadek eksperymentu więziennego Zimbardo powtórzonego w Polsce przez Artura Żmijewskiego

Słynne badanie wykonane (1971) przez Phillipa Zimbardo, znane jako Stanfordzki Eksperyment Więzienny, około 10 lat temu zostało powtórzone w Polsce przez Artura Żmijewskiego, artystę uprawiającego sztukę wideo. Jego wyniki zasługują na uwagę psychologów społecznych prowadzących badania eksperymentalne. Nagranie wideo opublikowane przez Żmijewskiego (2005) pokazuje, że jego przedsięwzięcie zakończyło się zupełnie innym wynikiem niż oryginalny eksperyment. Uczestnicy sami zdecydowali o jego przerwaniu. Różnica mogła wynikać z niedostatecznego starania artysty o to, by jego działania były w pełni zgodne z regułami metodologicznymi, których przestrzeganie przy prowadzeniu eksperymentu uważają za konieczne badacze akademiccy. Innym możliwym wyjaśnieniem dlaczego wyniki Żmijewskiego odbiegały od otrzymanych przez Zimbardo jest wpływ szczególnego kontekstu społeczno-kulturowego. Kultura polska różni się w kilku wymiarach od kultury założonej przez Zimbardo przy tworzeniu układu eksperymentalnego i interpretacji wyników. Artykuł ten – zawierający opis projektu Żmijewskiego wraz z proponowanym wyjaśnienie wyników odwołującym się do porównań międzykrajowych – ma zachęcić do dyskusji na temat roli kontekstu społeczno-kulturowego w badaniach eksperymentalnych.

Słowa kluczowe: Phillip Zimbardo, stanfordzki eksperyment więzienny, Artur Żmijewski, kontekst społeczno-kulturowy, wymiar kultury

FOLIA 202

Annales Universitatis Paedagogicae Cracoviensis

Studia Sociologica VIII (2016), vol. 1, p. 117–142
ISSN 2081–6642

Zbigniew Karpiński
Institute of Philosophy and Sociology Polish Academy of Sciences

Kinga Wysieńska-Di Carlo
Institute of Philosophy and Sociology Polish Academy of Sciences
Albert Shanker Institute, Washington, DC

Modelling Social Situations: Trust and Cooperation Among Strangers of Unequal Status[1]

Abstract

This paper presents a process of constructing and testing a new theoretical model, one which explains how status differentiation affects cooperation between partners involved in one-shot Prisoner's Dilemma (PD) situations. Bridging claims from status characteristics and collective action theories, we predict that in a PD where actors are differentiated by salient status distinctions, the rate of cooperation will vary depending on the partner's status relative to the focal actor, as well as on whether it is a simultaneous or sequential game. In order to demonstrate the challenges involved in constructing and testing new models, we describe two different one-shot, two-person PD experiments, which are designed to test our predictions. We conclude with both substantive and methodological discussion.

Key words: status, trustworthiness, Prisoner's Dilemma, theory construction

Introduction

The cumulative growth of knowledge, which is the ultimate goal of science, occurs through relating abstract theoretical claims to each other, as well as to empirical observations. The element that allows for bridging the theoretical with the empirical, for coordinating data collection, and for guiding interpretation of both, is a *theoretical model* (Skvoretz 1998). Building theoretical models that provide explanations and exact predictions of previously unexplained or loosely explained phenomena is at the core of the scientific endeavour. In this paper, we describe step-by-step the process of constructing and testing a new theoretical model explaining how salient status differences among persons with no previous history of cooperation affect their behaviour in a one-shot Prisoner's Dilemma (PD). We start our presentation with a brief discussion of some methodological issues related to theory construction and the cumulative growth of knowledge in sociology. Next, we state the substantive problem at hand and present the theories that we apply to build our

[1] Research reported in this paper was supported by a grant from National Science Centre (Grant No. UMO-2012/05/D/HS6/03118). Authors are listed in alphabetical order and contributed equally.

model and derive testable hypotheses. Finally, we report on two experiments that we designed to test our hypotheses, and discuss limitations of our designs.

Cumulative growth in sociology

Cumulative theory growth in sociology can occur through building multilevel theories (Markovsky 1997) and/or through relating theories to each other either within theoretical frameworks (Berger, Zelditch 1997) or between them (Fararo, Skvoretz 1993).

Building multilevel theories fosters our understanding of macro implications for micro processes, and vice versa. Markovsky (1997), specifying criteria for multilevel theories, notes that theories of collective action are naturally multilevel, as they focus on explaining macro level (group) outcomes as a result of the micro level (individual) choices of mutually dependent actors. Expectation states theories, which explain the emergence of status structures out of collectively- and goal-oriented encounters between individuals, and account for their impact on stratification systems in society, represent another example of a multilevel approach (Lawler et al. 1993).

Multilevel theories grow when a set of statements meets the *containment and bridging conditions*. The containment condition is satisfied when theories are comprised of statements at two or more levels of analysis (containment) and the higher level statements contain multiple lower level units, such as body/cells, structure/positions, neighbourhood/residents. The bridging condition can be met either by: constructing a conditional statement in which the level of antecedent differs from that of the consequent (e.g., if actors are differentiated by gender, then the power and prestige hierarchy of the group will reflect unequal expectations held for different genders); or by formulating definitions, wherein the subject of the higher-level statement is defined in terms of the lower-level subject, (e.g., a group consists of at least two persons). Note that the latter type of bridge is a tautology and, as such, is untestable. Constructing and testing multilevel theories entails the same procedures as developing and verifying single level theories – that is, with special attention to avoiding the reification error, and to choosing carefully methods of data analysis (Markovsky 1997).

The multidimensional model of theory growth in sociology, put forth by Berger and associates (Berger, Wagner, Zelditch 1987), focuses on yet another aspect of cumulative theory growth, namely the relationships between empirical theories within a given general framework, paradigm, or *research program*. Five different *patterns of theoretical growth* have been identified (see Berger, Wagner, Zelditch 1987; Berger, Zelditch 1997; Shelly 2002), the most relevant of which for our purposes is *integration*. Integration entails merging two or more unit theories by identifying the interrelationships between theoretical arguments with the same range and domain of application, or by specifying conditions under which the process described by alternative mechanism operate (Berger, Wagner, Zelditch 1987; Berger, Zelditch 1997; Shelly 2002).

In Berger and colleagues' model, which is an intra-program model, the concern is the growth of theories that share a common set of concepts and assumptions.

What we propose here, however, is an attempt to relate multilevel theories originating in distinct general research programs. Fararo and Skvoretz (1993) call this *integration by consolidation*.

The feedback between theory and data plays an important role in the growth of scientific knowledge. Negative empirical evaluations indicate mistakes in the theoretical structure or its empirical instantiation. Progressive revision and retesting, in most cases, are a result of the interplay between theory and its tests (Wagner 2000). Models mediate between abstract theories and empirical data (Skvoretz 1998; Wysienska, Szmatka 2002; Lovaglia, Willer 2002; Karpiński, Skvoretz 2015). Constructing a model requires a formalized (logical or mathematical) representation of the relationships between variables of interest, instantiation of abstract theoretical elements, specification of scope conditions, and identification of empirical techniques (Berger, Zelditch 1997). Below we present all of these elements.

Substantive theories: group identity and cooperation

Sociologists and other social scientists have long been interested in the effect of *diversity* on *trust* and *cooperation*. By *diversity*, we mean differentiation of members of a collectivity with respect to some salient criterion that divides that collectivity into discrete groups. *Trust* is defined in terms of a person's expectations as to whether the behaviour of another person will be benign, these expectations being based on inferences about certain personal traits and intentions of the other actor (Yamagishi, Yamagishi 1994; Molm et al., 2000). And, finally, *cooperation* refers to the collective action by members of a group to pursue common goals. Cooperative behaviour has usually been modelled using the PD, whether two-person or multiple-person (Axelrod 1984, Kollock 1998a). In its standard application, a PD is a game between self-interested players who know nothing about one another except for their incentive structures.[2] The defining features of the PD game are: (a) mutual cooperation is most advantageous to the group as a whole; but (b) individually each player is better off defecting when the partner chooses to cooperate. Questions about the effect of group differentiation on the rate of cooperation amount to asking how the incentive structure is affected by the knowledge of one's own group membership, and that of the other. From the standard game-theoretic point of view, that knowledge is supposed to have little, if any, effect on trust and cooperation, unless the players believe that group membership carries some additional information about trustworthiness or willingness to cooperate. Research finds consistently higher rates of cooperation in PD games played by members of the same group, and lower rates in games played by members of different groups, compared with a baseline setting in which players are anonymous (Brewer, Kramer, 1986; Kramer, Brewer, 1984). This result holds for real-world groups (categories), such as ethnicity or nationality (see, for instance, Yamagishi et al., 2005; Habyarimana et al., 2007), and for situations created in the laboratory, such as preference for art by Klee vs. Kandinsky (Yamagishi, Jin, Kiyonari 1999; Yamagishi, Kiyonari 2000; Aksoy 2015). Also, shared

[2] See the Appendix for a detailed description of the two-person PD game.

group membership fosters cooperation and trust in two-person as well as multiple-person games (Brewer, Kramer, 1986; Dawes, Van De Kragt, Orbell, 1988).

There are several competing theoretical arguments explaining why shared membership elevates the rate of cooperation above the level found in the baseline setting. According to social identity theory, mere difference can lead people to evaluate their in-group more favourably than the out-group (Tajfel 1982). Consequently, actors come to perceive members of their own social category as more trustworthy than members of another social category (or categories). Acting on this perception, they are more likely to cooperate with an in-group member than with an out-group member (Brewer 1979; Brewer, Kramer 1986; Kramer, Brewer 1984). Research by Toshio Yamagishi and his collaborators, however, challenged this explanation by showing that: (a) shared membership is conducive to cooperation in simultaneous but not in sequential games, in which the expectation of direct reciprocity is a stronger predictor of cooperation (Yamagishi, Kiyonari 2000); and (b) shared membership promotes cooperation only when such sharing is common knowledge (Yamagishi, Jin, Kiyonari 1999). In other words, in games played by members of the same group, both players have to be aware of each other's group membership, since, otherwise, their common identity will not be sufficient to overcome the temptation to defect unilaterally. To illustrate, suppose two players, A and B, play against one another. Both belong to the same group, but only A is aware of that fact. Player A also knows that B is not aware of their common group membership. Because B is not aware that A is his or her fellow group member, B has no special incentive to treat A favourably. Expecting that, A also will not behave favourably towards B. As a result, the rate of trust and cooperation will be about the same as in the baseline setting. This is where Yamagishi's work departs from social identity theory, since the latter would predict that A – being aware of his or her common group identity – would choose to cooperate with B. In other words, social identity theory claims that actors cooperate with in-group members because they are in-group members, whereas Yamagishi claims that actors cooperate with their in-group members because they expect their in-group members to cooperate with them, but also that the expectation of cooperation has to be mutual, which only happens if both players are aware of their common identity. Research provides stronger support for Yamagishi's proposition than for social identity theory (Foddy, Yamagishi 2009; Yamagishi et al., 2005; Yamagishi, Kiyonari, 2000; Yamagishi, Jin, Kiyonari, 1999; Yamagishi, Mifune 2008).

Recently, Brent Simpson (2006) attempted to reformulate the social-identity theoretic account of cooperation by combining concepts from social identity theory (Tajfel 1982, Turner 1985) with the fear-and-greed approach to the study of social dilemmas (Ahn et al. 2001; Dawes et al. 1986; Rapaport, Chammah 1965; Simpson 2003). According to this approach, defection in a PD can be motivated by fear or greed, the former meaning the motivation to avoid being exploited by a non-cooperative partner, and the latter being the temptation to free ride on the partner's cooperation. Both of these motivations are defined in terms of game parameters (see the Appendix for a more formal treatment). More specifically, greed is equal to the difference between the payoff to unilateral defection (T) and the reward for cooperation (R). Fear, in turn, is measured as the difference between the payoff

for mutual defection (P) and unilateral cooperation (S). In other words, the larger is the payoff for unilateral defection relative to the reward for mutual cooperation, the greater one's greed, or the inclination to exploit the other. Similarly, the larger is the payoff for mutual defection in comparison with that for unilateral cooperation, the more afraid one is of the other's exploitation. By manipulating the payoffs, one can change the amount of fear and greed in the game, leading to more or less cooperation, depending on which payoffs are changed and in which direction.

Simpson couples the concepts of fear and greed with a central notion of social identity theory, the metacontrast principle. According to this principle, behavioural responses to the distinction between the in-group and the out-group are driven by: (a) maximization of inter-group differences; and (b) minimization of intra-group differences (Turner 1985, Hogg 1996). When a player is motivated by greed, cooperation with a fellow group member achieves these goals. But, when a player is motivated by fear, cooperation realizes the former goal, whereas defection realizes the latter. Thus, the two goals cancel each other out, and, consequently, we have no reason to expect any effect of identity on responses to the fear component of social dilemmas. This reasoning leads Simpson to propose that when players in a prisoner's dilemma make their decisions sequentially: (a) the player who moves first will not be affected by common identity, as the first decider in a sequential PD is motivated only by fear; but (b) the player who makes the second move will be affected by common identity, as he or she is motivated only by greed. This is where the fear-and-greed approach differs from Yamagishi's because Yamagishi proposed that common identity has no effect on cooperation in sequential games (Yamagishi, Kiyonari 2000). Research provides strong support for Simpson's reasoning (Simpson 2006).

To sum up, the three perspectives – social identity theory, Yamagishi's generalized reciprocity hypothesis, and Simpson's fear-and-greed approach: (a) agree that shared group membership elevates the rate of cooperation above the levels observed in anonymous settings, when decisions are made simultaneously; but (b) differ with respect to predictions concerning sequential games, with social identity theory claiming that common identity will increase cooperation in both simultaneous and sequential games, Yamagishi claiming it will have no effect in sequential game, and Simpson predicting that it will only affect the decision made by the second decider.

Substantive theories: Status and cooperation in PD

All the theories discussed in the preceding section are limited to the study of cooperation when the distinction between the in-group and the out-group is based on a *nominal* characteristic. However, characteristics that differentiate social actors may have not only nominal, but also *status* value (Berger, Webster 2006; Ridgeway 2001, 2014). Status distinctions are different from nominal distinctions in one important respect: while the latter give rise to in-group bias, i.e., preferential evaluation and treatment given to members of one's own group, regardless of the group's position in a social hierarchy, status distinctions give rise to *status beliefs* that accord more competence, esteem, and general social worth to one 'state' of the underlying

characteristic than to other state(s) (Ridgeway 2006). In addition, both high- and low-status actors accept the status beliefs.

Status distinctions can be argued to have an important implication for cooperation in social dilemmas. High-status actors are perceived as more competent, respected, and esteemed members of the collectivity, and also as more *group-oriented* and *cooperative* (Ridgeway 1982). Consequently, high-status actors can be expected to cooperate more in social dilemmas, compared with low-status actors. Thus, playing a PD game against a high-status actor reduces one's fear (of being exploited), but not necessarily one's greed (to take advantage of the other).

Choices made by high-status actors are not only more likely to be more cooperative, they are also more likely to be perceived as *normative*, thus providing low-status actors with cues to 'desirable' or 'appropriate' behaviour in a given situation. As a result, low-status actors will follow the behaviour of the high status actor due to the former's belief that it is the right thing to do. High-status actors are also more influential, so they can choose to cooperate expecting to influence low-status members (Sell 1997). Moreover, high-status actors experience positive emotions, (Lovaglia, Houser 1996; Lovaglia 1997) and positive emotions foster integration (Kemper 1987, 1991). Positive emotions induce behaviour that ties group members together, whereas negative emotions tend to draw group members away from each other. Positive emotions experienced by high-status members may therefore compel high-status actors to act in a manner that binds low-status actors to the group (Lovaglia, Houser 1996).[3]

To sum up, the reasoning presented above asserts that when actors who differ in status play a sequential PD game, the level of cooperation will depend on who initiates the game. If the high-status player makes the first move, his or her choice is likely to give the low-status player a cue as to what he or she should do himself or herself. So, if the high-status actor initiates the game by acting cooperatively – which we expect him or her to do, given the relationship between status and group orientation – the low-status player is more likely to respond in kind than he or she would respond if the information about status was unavailable. Furthermore, if the greater likelihood of cooperative behaviour on the part of the low-status actor reflects his or her belief that cooperation is the correct behaviour because this is what the high-status actor does, then one can expect the low-status actor to respond cooperatively to the behaviour of the high-status player, and to act cooperatively in future interactions involving different people.

On the other hand, the player who makes the first move in a sequential PD game is motivated by fear, while the second player is motivated by greed. As a result, the high-status actors' willingness to cooperate in a sequential PD played against a low-status actor can be reduced by the former's fear of being exploited. The high-status actor's fear can be further exacerbated by the expectation of the low-status actor's

[3] It is beyond the scope of this paper to review existing research on the link between status and contributions to public goods, but it should be noted that it demonstrates both that contributions to public good lead to higher status (e.g., Willer 2009), as well as that higher status actors, given a chance, initiate contribution more often than low status actors and they contribute more. Low status actors, in return, contribute more when following the lead of high status ones (e.g., Sell 1997; Kumru, Vesterlund 2010; Simpson, Willer, Ridgeway 2012).

being a less cooperative group member. In other words, the low-status actor is motivated by greed, and this motivation is even stronger than it would be in a situation where players had no information about their relative statuses. Consequently, the high-status actor has little incentive to make a cooperative move.

When a low-status actor initiates the game, however, the situation changes considerably. As a first decider, the low-status actor faces the fear of being exploited, but this fear is reduced by the knowledge that the partner is of higher status, and therefore is likely to respond cooperatively. Thus, the low-status actor has reason to expect the high-status actor to respond in kind to his or her cooperative move, which induces him or her to choose cooperation. This reasoning leads us to the following proposition:

> *Multilevel Hypothesis* 1: In a sequential PD game, if actors are differentiated by a salient status characteristic, the rate of cooperation will be higher when the low-status actor initiates the game than it is when the high-status actor does so.

Actors are predicted to adjust their behaviour in a PD game based on the information about their partner's status. Irrespective of whether the actor is of high or low status, knowing that his or her partner is of high status encourages the actor to behave cooperatively, expecting the partner to respond in kind. When the actor knows the partner to be of low status, he or she is less likely to cooperate, expecting the partner to behave selfishly. That is, actors are more likely to cooperate in the PD game when their partner is of high status than when he or she is of the low status. Thus, our next hypothesis can be stated as follows:

> *Multilevel Hypothesis* 2: In a sequential PD game, the rate of cooperation will be higher when the partner is of high status than when he or she is of low-status.

This prediction can also be presented in a somewhat different way. Let HH denote a pair of players, each of whom has a high status, and let HL represent a pair of players of which the first player has a high status, and the second player low status. Then, according to hypothesis 2, the rate of cooperation in pairs of type HH is predicted to be higher than in pairs HL. Similarly, we expect the rate of cooperation in pairs of type LH to be higher than in pairs LL.

We can also expect a higher rate of cooperation when both players are of high-status and a lower rate of cooperation when they are both of low status. Note that when both players are high-status, each can expect the other to behave in a cooperative manner, and so each has an incentive to act cooperatively as well. Similarly, when both players have low statuses, each has little motivation to choose cooperation, as each believes their partner will not cooperate. Therefore:

> *Multilevel Hypothesis* 3: In a sequential PD game, the cooperation rate among high-status actors will be higher than that among low-status actors.

Arguably, hypotheses 1 through 3 extend to simultaneous games as well. First, the type of game does not affect the nature of the expectations concerning the other's cooperativeness. All else being equal, the overall rate of cooperation may be

somewhat different in sequential games than in simultaneous ones, but the effect of status can still be said to operate in the same way in both types of games. As stated above, both fear- and greed-based motivations are present when actors make decisions in simultaneous games. Still, when status is salient in a prisoner's dilemma, high-status actors are more sensitive to the fear component, whereas low-status actors are more attuned to the greed component. This means their motivations are very similar to those in a sequential game in which the first player is driven solely by fear and the second solely by greed. Thus, introducing status into the situation can be expected to reduce differences in the rate of cooperation across simultaneous and sequential games. Note also that in simultaneous games both actors respond to expectations about the other's behaviour, whereas in sequential games this is true only of the first deciders, as the second deciders respond to actual behaviour (Yamagishi, Kiyonari, 2000). Insofar as, in the experiment that we report below, we focus on the behaviour of the first deciders, this is yet another reason to expect little difference in the effect of status on cooperation across sequential and simultaneous games.

Let us present a summary of our predictions. With two levels of the subject's status and two levels of the partner's status, there are four possible combinations. Let us use L to denote low status and let us use H to denote high status. The four combinations are HH, HL, LH, LL, with the first symbol in each pair referring to the subject's status. Thus, for instance, HH refers to pairs in which both the subject and the partner are of high status, HL means pairs in which the subject has high status and the partner has low status, and so forth. Finally, we use the symbol '>' to indicate the ordering of pairs in terms of the expected cooperation rate, such that HH>LL indicates that pairs in which both players are of high status are expected to have higher cooperation rates than pairs in which both are of low status.

Table 1. Summary of theoretical predictions

Hypotheses	Game type	Prediction
Hypothesis 1	Simultaneous and Sequential	LH>HL
Hypothesis 2	Simultaneous and Sequential	HH>HL, LH>LL
Hypothesis 3	Simultaneous and Sequential	HL>HH

Scope conditions

Particular theories apply to certain kinds of phenomena under certain conditions. *Scope conditions* specify the circumstances under which the relationships expressed in theoretical hypotheses are expected to hold true (Foschi 1997; Cohen 1980, 1989; Markovsky 1994; Walker, Cohen 1985). In other words, they are abstract and universal statements that define properties that must be present or absent in a situation in order for the theory to be applicable, testable, or assumed to be true. Scope statements (also referred to as boundary conditions, see Shelly 2002) assert nothing about the truth of theoretical principles. The importance of the scope

statements is simply that wherever scope conditions are met, a theoretical claim is applicable. Nothing in either the scope statement or the theoretical argument guarantees that any particular situation will meet the scope conditions (Cohen 1989, p. 83). Efforts to integrate two or more theories will involve a process of rectifying two or more sets of scope conditions (Shelly 2002).

Expectation states theories are limited in their applicability to goal-oriented groups in which group members value the task and share the motivation to achieve success and avoid failure, where standards for evaluating success and failure are at least partly under the control of group members, and where group members recognize that they are interdependent and thus must take into account other group members' actions. In addition, status characteristics theory specifies that group members must believe that some characteristics are instrumental to achieving the group's goal – that is, possessing a high state of the characteristic increases the likelihood of success. Collective action theories, on the other hand, apply to mixed-motive situations, which are situations in which there is an incentive to defect despite the fact that mutual cooperation is to the group's benefit (Sell 1997).

It appears that the two theoretical frameworks differ regarding the situations to which they apply, although these situations are not mutually exclusive. As actors are interdependent, they need to take each other's (potential) actions and motivations into account. As there is a mutually satisfying outcome that is better than when no cooperation occurs, and that is easily identifiable, we can also assume that task orientation is present, and that there are characteristics instrumental to achieving it. As our primary goal is to explain status effects in PD situations, the hypotheses summarized in Table 1 apply to situations in which actors are differentiated with respect to status and face a social dilemma. In the simplest case, the situation is comprised of only two actors who may be of the same or different status. Status is, of course, not the only dimension of differentiation in social groups. Distinctions other than status have powerful effects on cooperation in social dilemmas, and these distinctions can interact with status. The focus on status in our work is not to imply that we believe categorical distinctions and social identity effects to be unimportant. Rather, we propose to *abstract* status from other dimensions of differentiation to study its effect in *isolation* from them.

The experiment

In order to test our predictions, we needed subjects to play either sequential or simultaneous PD games in pairs that are either of unequal or equal status. Also, the latter pairs are further divided into those made up of high-status players and those comprised of low-status players.

There are three variables that we explicitly manipulated: subject's status (high vs. low); partner's status (high vs. low); and the type of the game (sequential vs. simultaneous). This yields a total of 8 treatments. The design of our experiment combines the standard experimental setting used in expectation states research (see Berger 2007 for a review of the setting) with a two-person investment game

setting (see Berg, Dickhaut, McCabe 1995; see also Sell 2007 for a review of social dilemma experiments in the social sciences).

The expectation-states experimental setting was employed to induce status differences between our subjects. To study the effect of these differences on the rate of cooperation, we use the investment game, which has the structure of a PD game, and has been used in many studies as an experimental 'operationalisation' of a PD game (see Simpson 2006; Yamagishi, Jin, Kiyonari 1999 for examples). Given our hypotheses and research interests, our focus in this experiment was on the investment game. In other words, because our hypotheses are not concerned with the emergence and reproduction of status hierarchies in task groups – a major theme in expectation states theory – we used the expectation-states setting instrumentally, in order to introduce status distinctions into our experimental situation. Beyond that, we did not make any substantive use of data generated in the part of our experiment that employed this setting.

The first part of the experiment: assigning ability

In order to test our hypotheses, we needed to introduce *status differences* into our experimental situation. In status characteristics theory, status differences determine levels of assigned competence and corresponding expectations regarding an actor's performance at a group task. In many social-scientific experiments investigating links between status and collective action, status differences are introduced using quasi-experimental measures, such as level of education (e.g., Simpson, Willer, Ridgeway 2012), gender (e.g., Sell 1997), or scores on a quiz administered at the beginning of the study (List 2007; List, Cherry 2008; Kumru, Vesterlund 2010). Using these characteristics as measures of status in an experiment is relatively easy, does not require cover stories, and usually avoids deception. The drawback, however, is that some of these characteristics, such as gender, give rise to social identity effects, and identity has been found to have a powerful effect on cooperation. Also, the quasi-experimental measures can be nontrivially associated with the dependent variable, i.e., cooperation, reducing the internal validity of the experiment.

To avoid such problems and study the 'pure' status effect, we decided to base our measure of status on the *standard experimental situation* in expectation states research (see Troyer 2002, Berger 2007 for an overview). In expectation states experiments, subjects first work individually on a series of problems that are said to involve a recently discovered perceptual or cognitive ability (named 'contrast sensitivity ability' and 'meaning insight ability,' respectively). Subjects are told that, according to collected evidence, these abilities have no association with other skills, abilities, or personal attributes, but other properties of these new skills have yet to be investigated. The problems are quite ambiguous, but the subjects are explicitly told that there is a correct answer and individuals that are more contrast-sensitive or have more meaning insight ability are more likely to answer correctly. In our experiment, we used a version of a setting designed by Lisa Troyer (2002), which utilizes visual problems related to 'contrast sensitivity.'

Once a subject completes all his or her tasks, he or she is assigned a 'score,' which is supposed to be equal to the number of correct answers he or she gave. The scores are either 'high' or 'low,' indicating their ability (or lack thereof) to solve most of the problems correctly. Contrast sensitivity is actually a fictitious ability and the tasks given to the subjects do not have a correct solution in any meaningful sense. The 'scores' are randomly assigned to the study participants, regardless of their responses to the problems. Thus, the classification of the subjects as having a high or low level of the ability is independent of their actual abilities or characteristics. The sole purpose of the first part of the experiment is to lead the subjects to believe that: (a) there is such an ability as contrast sensitivity; (b) the problems they just solved measure it in a valid way; and (c) they possess either a high or low level of the ability. With these beliefs, they proceed to the second part of the study, in which they are paired with a partner to work as a team on another set of contrast sensitivity problems.

The second part of the study: the link between status characteristic and cooperation

Two features of the second part of the experiment are crucial. First, the subjects know their own contrast-sensitivity score, as well as that of their ostensible partner. Second, as they are supposed to work as a team and take one another's propositions into account, they are given the opportunity to exchange their suggested answers to the task at hand before making their final decisions. At each trial, a subject first submits his or her initial solution and learns the initial solution proposed by his or her partner. If their propositions are different, the subject can either stay with his or her initial idea, or discard it and accept the alternative proposed by the partner. In expectation states theory, the probability of staying with the initial response is interpreted in terms of *rejection of influence* and modelled as a function of differences in performance expectations linked to status differences between team members. As mentioned earlier, we were not really interested in studying patterns of influence. Instead, we used the second part of the experiment to induce a link between status differences and cooperation by providing subjects with explicit information that their partner did or did not act as a 'good team player,' depending on the partner's status. More specifically, in line with the research by Ridgeway (1982) in which high-status actors were found to be more group-oriented and cooperative, the high-status partners were described as 'good team players' and the low-status partners as failing to act as such.

The 'partner' in the second part of the study was actually simulated. Each subject worked independently of the others, although they were led to believe that they collaborated in pairs. The 'partner' was *pre-programmed to disagree* with the subject on an initial solution for a specified percentage of the time.[4] The trials in which the 'partner' agreed with the subject were randomly distributed. The reason for using a simulated rather than a real partner is that it permits the researcher to manipulate explicitly the frequency of disagreements between the subject and the

[4] In expectation states experiments, the percentage is 75. In our study, we set the percentage to 80 per cent in our first experiment and to 70 in the second.

partner, and it is precisely the study of what happens when they disagree that has been at the centre of much research within the expectation-states tradition. Given that the objective of our study was not to test predictions derived from expectation states theory, we could let the subjects interact over the computer network with real rather than simulated partners. We decided to follow the original procedure, however, because it gave us the opportunity to observe if we were able to replicate the results of earlier studies on status characteristics. This is an important consideration. If we failed to *replicate* those results, it would imply that there is something wrong with our experimental procedures, instructions, or protocols.

The third part of the study: investment game under status differentiation

After the second part of the study came to an end, the subjects were asked to complete a short survey. They then received the feedback on their partner's group orientation. At this point, they were informed that the study had come to an end and they could go, but also that we were giving them an opportunity to take part in another, apparently unrelated experiment, which would allow them to earn extra money. Before they made their decision as to their participation in that additional experiment, however, we informed them about its objective and procedures.

The instructions that followed presented the new experiment as one in which the subjects are again required to work in pairs. It was emphasized, however, that they might have a different partner in the new experiment than they had previously. We did not want our subjects to think they would be paired with the same person, because the history of their relationship might have distorted the subjects' behaviour. Each member of a pair then received a pool of 'tokens' from the experimenter and had to decide whether to keep the endowment to him or herself, or divide it. Subjects were informed that the experimenter would double each token given to their partner. Participants could therefore benefit greatly by exchanging tokens – when one gave all his or her tokens to their partner and vice versa, both would end up having twice as much as at the beginning of the game. Knowing this, however, each of them was also tempted not to act in this manner because when one gave all his or her tokens and the other gave none, the latter would end up having three times what they started with and the former with nothing. As a team, participants were best off cooperating with one another, with cooperation meaning here the transferring of one's endowment to one's partner, but individually each benefited by unilaterally defecting, i.e., failing to share their resources with their partner. Unlike in the 'standard' PD game, the decision was not binary (cooperate or defect), as the subjects could give all or none of their tokens to the partner, or any amount in between.

Our instructions made it clear to the subjects that the decision they would face in the game was not a trivial one, as it would affect their payment, and that the tokens had a monetary value. Giving tokens to a trustworthy partner may be highly beneficial, because trustworthy partners are likely to honour trust placed in them and reciprocate, but misplaced trust may cost the subject dearly. The instructions

then stated that, to help them make their decisions, the information about their 'new' partner's performance in the earlier experiment would be displayed on screen, before they chose to participate in the game. Importantly, the performance of the 'new' partner matched closely that of their previous partner, as we wanted the subjects to 'project' onto their 'new' partner the link between their partner's scores and his or her willingness to cooperate.

One element of the instructions was varied randomly – namely, the information about the type of game. That is, half of our sample was informed they would be making their decision *simultaneously* with their partner, so that they could learn what the partner did only after they had made their own decision. The remaining subjects were informed that they would make their decisions *in sequence*, meaning that the first mover has no way of knowing their partner's decision, but the second mover makes his or her decision knowing what the partner did. It is, however, important to keep in mind that in sequential games the subject was always assigned the role of the first decider, although they were led to believe that this was due to a random choice. After the game-type manipulation, the subjects proceeded to two practice trials.[5] Once the practice trials were over, the subjects were matched with a potential partner, and could see the information about the partner's performance in the previous experiment. They then made a final decision about participation in the game. When they chose to participate, the game commenced. Note, again, that the partner was actually simulated and programmed to match the choice made by the subject, so that the subjects who gave nothing to their 'partner' received nothing and those who gave all their tokens received their 'partner's' whole endowment. When the game was completed, the subjects were asked to fill in a short survey and then summary information with their earnings was displayed on the screen.

Study 1

The first study was conducted in May and June 2014. The participants were students of public and private universities in Warsaw, Poland. They were recruited using email announcements, ads placed on social networking sites, and posters distributed in their schools. These ads and announcements emphasized the opportunity to earn money by taking part in a study of 'group decision making.'

After arriving in the laboratory, subjects were seated at computer stations. Next to each station was an informed consent form, which the subjects were asked to read and sign once they had gone through all the experimental instructions that were to be displayed on the screen. The experimenter also reminded them that they could leave the laboratory whenever they chose to do so, regardless of whether or not they completed the whole experiment, and without having to provide any reason. They were also told that leaving the experiment early would have no consequences for them except for its reducing the amount they would be paid, as their payment was dependent upon their performance in the experiment, and that leaving early could result in their

[5] In the sequential-game treatments, the subjects were the first deciders in one practice trial and the second deciders in the other.

receiving no payment at all. After that, they were asked to begin, and the experiment proceeded as described in the preceding section. However, a few details bear mentioning.

First, at the beginning of the session, the subjects were asked to provide information about their gender, age, and their year of study. Second, before the start of the second part of the experiment, in which subjects solved contrast-sensitivity problems in pairs, information about gender, age, and the year of study of their partner was displayed, along with the partner's score from the first part of the study. In order to control for the effects of these characteristics, the 'partner' was always of the same gender and a similar age as that of the subject. Furthermore, the subjects were given 20 contrast sensitivity tasks in the first part of the study and another 20 such problems in the second part, when they were matched with a partner. Thus, each subject was given as many as 40 problems in the course of the experiment. Finally, the information about age and gender was not displayed when an ostensibly new partner was assigned in the third part of the experiment.

A total of 118 subjects took part in the study. Each experimental session was comprised of an even number of participants, between four and eight, all of whom were in the same room, so that each of them could see how many other participants were present. Having an even number of subjects was necessary to validate the manipulation that they were working in pairs during the second and the third part of the study. When one of the participants failed to show up, a research assistant sat in his or her place. Also, because, as indicated above, at the beginning of the second part of the study, the subjects were informed that they were paired with a person of the same gender, there had to be at least two male participants and two female participants in each session to validate that information. After the second part, the subjects were informed that their remuneration was equal to 30 PLN. They could increase it by taking part in an additional study, in which their task would be to allocate 100 'tokens' between themselves and a partner, with each token worth 0.05 PLN. Thus, the tokens received from the experimenter were worth 5 PLN. Mutual cooperation could double the amount.

At the end of each session, the experimenter revealed all the deceptions used in the experiment, gave a detailed reason for why the deceptions were used, and responded to questions and comments from the subjects. Each session lasted about 40 minutes.

To sum up:

In the first part of the study, subjects solved 20 contrast-sensitivity tasks, after which they were randomly assigned either a high or low score;

In the second part, they were paired with a partner of the same gender and a similar age to work in pairs on another set of 20 contrast sensitivity tasks. The information on the partner's status was randomly varied. For the first two parts of the study, they earned 30 PLN (roughly, 10 USD at the time the study was conducted);

In the third part, they played an investment game by allocating 100 tokens between themselves and their partner, ostensibly a different person than their partner in the second part. Their new partner had similar contrast-sensitivity scores as the previous partner, but was of unspecified gender, year of study, and age. The tokens were worth 0.05 PLN each, so the endowment received from the experimenter amounted to 5 PLN.

Results of Study 1

Analyses of data from the second part of our study show that we were able to reproduce the results of earlier experiments using the standardized setting of the expectation states theory. That is, we observed that $P(S)$, or the probability of subjects staying with their initial decision was the highest for pairs HL – that is, situations in which a high-status subject was paired with a low-status partner – and it was the lowest for pairs LH. For status-equal treatment the probabilities were close to one-half. One-way ANOVA for $P(S)$ gives $F_{3,114} = 22.08$. The result is highly significant, well beyond the conventional significance levels. Thus, our status manipulation can be said to have been effective.

In Table 2, we present means and standard deviations of the cooperation rates across the conditions of our experiment. The results turn out to be inconsistent with what we predicted. First, the means and medians for pairs HL and LH are different, but the direction of the difference contradicts Hypothesis 1, as the average for the former pair is higher than that for the latter; also, the difference is not statistically significant ($p = 0.21$).[6]

Table 2. Descriptive statistics for the cooperation rate in Experiment 1

Game type	Subject's status	Partner's status	N	Median	Mean	SD
Sequential	H	H	17	.860	.838	.187
Sequential	H	L	13	.850	.754	.295
Sequential	L	H	13	.550	.636	.198
Sequential	L	L	13	.600	.669	.211
Simultaneous	H	H	13	.800	.719	.284
Simultaneous	H	L	18	.850	.761	.274
Simultaneous	L	H	19	1.000	.817	.231
Simultaneous	L	L	10	.900	.845	.171

For simultaneous games, the result is similar: although the difference in medians is larger, the means for pairs HL and LH are quite similar and not significantly different (p=0.55). Second, the means for pairs HH are somewhat greater than those for pairs HL, in line with hypothesis 2, but that difference is most likely due to an outlier, since the medians are almost identical. Also, the difference in means is not significant (p=0.196). As for the low-status subjects, the means and medians for pairs LH and LL are very similar (p=0.71). Consequently, we have no evidence to support Hypothesis 2, which predicts that subjects modify their behaviour in the investment game depending on the partner's status. As we can see, high-status

[6] The reported p values come from a series of significance tests based on linear combinations of coefficients from a linear regression model, with cooperation as the dependent variable and status and game type manipulations as the independent variables.

subjects cooperate at a relatively high-rate, regardless of their partner's status, whereas low-status subjects cooperate at a relatively low rate, regardless of their partner's status. Thus, as regards sequential games, some status effect is observed, but in a direction opposite of that we predicted. Interestingly, the results are quite different for the simultaneous games. As we can see, it is the low-status subjects that appear to cooperate more in simultaneous games. That is, the means for pairs LH and LL are higher than the corresponding means for pairs HH and HL. However, the differences are small and not significant (the HH-HL difference: p=0.81; the LH-LL difference: p=0.47).

As for the third hypothesis, the difference in the cooperation rate in sequential games between pairs HH and LL is quite large, positive and marginally significant (p=0.056). This is consistent with our prediction, but we failed to find a similar effect in simultaneous games (p=0.18).

In general, there are two possible explanations of our not finding support for our predictions: either the predictions were wrong or we failed to reproduce essential features of our theoretical model. The latter interpretation did not require of us to abandon or discard the hypotheses and the reasoning on which they were based, so we decided to consider this interpretation first. Considering flaws in our experimental design was also justified insofar as alternative hypotheses implying higher cooperation rates in both simultaneous games (because of positive emotions) and sequential games (because of high status actors influence) were only partially consistent with our observations. Recall that, even though high status subjects cooperated more than low status ones, regardless of who the partner was in the sequential games, there were no differences across status conditions in the simultaneous games.

We have identified several features of our study design that may have contributed to an unusually high cooperation rate and our inability to find the hypothesized status effects. First, we found a substantial and significant gender effect: male subjects turned out to have cooperated at a greater rate than female subjects. In comparison with otherwise identical women, men gave their partners about 14 tokens more, on average (p = 0.003). This was quite perplexing, as it suggested that gender was a salient distinction in our design, contrary to our intentions. Recall from our description of the study design that the subjects were made aware of their partner's gender at the beginning of the second part of the experiment. Even though no information about their 'new' partner's gender was provided in the third part, gender effects were still stronger than experimentally manipulated status differences.[7] We therefore conjectured that removing the reference to the partner's gender would remove the gender effect completely.

Second, the experimental sessions lasted quite long (more than 30 minutes each) and all subjects were physically present in one room. It is possible that the

[7] Note, however, that if we assume that in low risk games (as explained below), it is greed, not fear, that is the dominating motivation (despite both motivations being present in the game structure), we would expect women to cooperate at higher rates than men (Simpson 2003). If, on the other hand, gender acted as a status characteristic, men (high status) should only cooperate more when they believe they are paired with women (low status), and so we should still observe different rates of cooperation among subjects of unequal status (Sell 1997).

physical proximity and the relatively long time spent together working on the tasks together fostered some group-formation process, whereby subjects developed a sense of common identity and positive feelings for one another, which precluded them from taking actions that might 'hurt,' and instead encouraged them to take actions that could benefit the others, regardless of any information about their respective statuses. Importantly, common identity is conducive to the subjective transformation of a PD into an assurance game (Kollock 1998b), and the transformation is conducive to cooperation (Simpson 2004) because defection is the dominant strategy in the PD but not in the assurance game. It is possible that making the sessions shorter and moving to a bigger computer lab might mitigate or delay the group process.

Third, we suspected that our failure to find any status effect might have resulted from the relatively low stakes in our investment game setting. Recall that each subject was given a *guaranteed payment* of 30 PLN after the first two parts of the study, and that the endowment they received in the third part was worth 5 PLN. The risk associated with 'misplaced trust'– or with allocation of one's whole endowment to their partner – might therefore have seemed rather low to our subjects. Accordingly, for the third part, we decided to decrease the guaranteed payment, and to increase the amount to be invested by the subjects. Also, recall that the endowment received from the experimenter was described in terms of abstract 'tokens,' which might have diminished the subjects' ability to assess the risk appropriately, because such assessment required them to 'translate' the tokens into monetary values in order to estimate the financial effect of a particular 'investment.' Such calculations are quite difficult to process. Consequently, it may have been the case that if we described the endowment in terms of money rather than 'tokens,' subjects would have been more likely to estimate their risks more correctly. With these ideas in mind, we went on to carry out a second study with somewhat modified design.

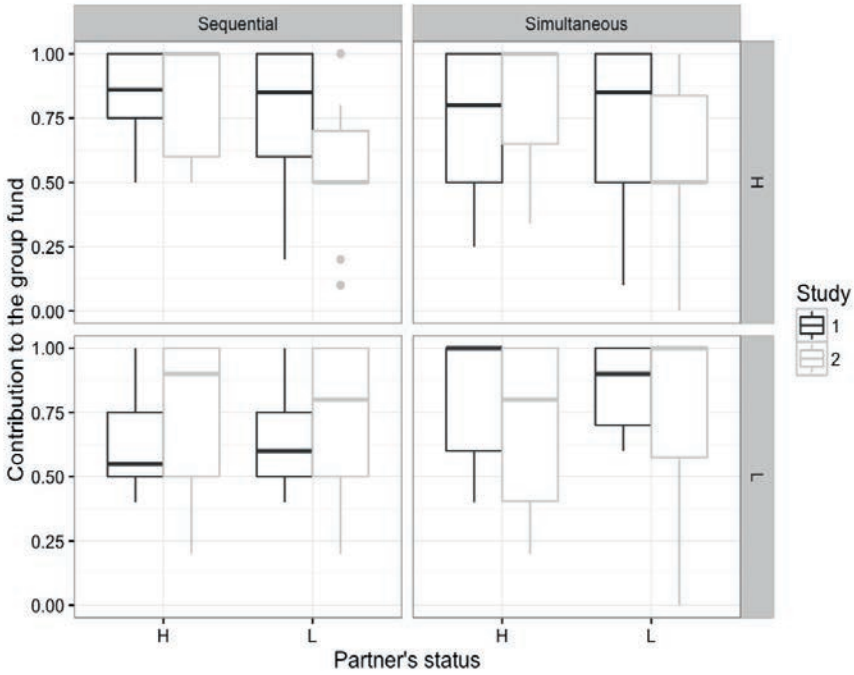
Study 2

The second study was conducted in February and March 2015. As previously, the participants were students from various departments of public and private universities located in Warsaw. We used the same channels of communication to recruit the participants. Once again, the ads and announcements informed students about the possibility of earning money by taking part in a scientific study of decision making processes in groups.

The general design of the second study was the same as in the first, but we changed a number of important details. First, we made the sessions considerably shorter by reducing to 10 the number of tasks in the first and second part of the experiment. Second, we simplified the instructions to make them shorter and clearer. Third, the only information about the 'partner' that we made available to the subjects was his or her contrast sensitivity score – i.e., there was no reference to the partner's age or gender. Fourth, the guaranteed payment after the first two parts of the experiment was reduced to 20 PLN and the endowment given by the

experimenter at the beginning of the third part was increased to 10 PLN. Thus, the guaranteed payment could be doubled by mutual cooperation or remain unchanged in the case of mutual defection. Fifth, the decisions in the investment game were described in terms of monetary stakes and not tokens.

Figure 1: Comparison of the results of Study 1 and Study 2



A total of 135 subjects took part in Study 2, four of whom were excluded due to violating scope conditions. Our final analyses thus include 131 observations. Figure 1, above, shows changes in the results between first and second experiment. In this figure, we make use of box-and-whisker plots to summarize the distribution of the dependent variable across experimental conditions. The rectangular boxes represent quartiles of the distribution, with the bottom of the box corresponding to the first quartile, the top – to the third quartile, and the thick segment inside the box to the median. Further, the horizontal whiskers adjacent to the boxes cover the distance of 1.5 times the inter-quartile range from the median. Any points beyond that range are outliers, represented as dots on the graph.

Figure 1 is divided into four panels and there are four box-and-whisker plots within each panel, corresponding to a total of 16 combinations of the subject's status (high vs. low), the partner's status (high vs. low), game type (simultaneous vs. sequential), and experiment (1 vs. 2). Figure 1 shows some notable differences between the results of the two experiments. The dependent variable in both experiments is the proportion of subjects' initial endowment that they transferred to their partners. The two top panels represent the behaviour of high-status subjects and the two bottom panels represent the behaviour of those who were assigned low

status. Further, the columns of the graph correspond to the type of game, and different colours were used to distinguish the results of our first study from those of the second.

The graph shows substantial differences in contributions across the two experiments, especially in the case of the high-status subjects. In the first experiment, high-status subjects' median cooperation was about the same, regardless of their partner's status or whether the game was simultaneous or sequential (although there is more variation in their contributions in the simultaneous games, as indicated by larger boxes). In the second study, their cooperation clearly varies depending on the partner's status (but not on the type of the game).

Low-status subjects' behaviour also changed from one study to the other, but the change is not as striking as for high-status subjects, and it occurred in sequential games only. In Study 1, low-status subjects' cooperation was rather moderate in the sequential game, regardless of whether their partner was of high or low status. In Study 2, their median cooperation increased, especially in games played against high-status partners.

Let us now return to our hypotheses to see if the data from the second experiment are more consisted with our predictions. First, the median for pairs LH is higher than for pairs HL, which is consistent with the first hypotheses. This result holds for both sequential and simultaneous games, which also confirms our expectations, although it is somewhat less pronounced in the latter compared with the former. Second, as indicated, high-status subjects do vary their behaviour depending on the status of the partner. As Figure 1 makes it clear, their median cooperation rate, in both simultaneous and sequential games, is 100% when their partner is of high status and only 50% if the partner is of low status. However, low-status subjects' behaviour fails to conform to the hypothesized pattern. In sequential games the low-status subjects do cooperate more with high-status partners than with low-status partners, but the difference is much smaller than the corresponding effect for high-status subjects. Also, in simultaneous games, the low-status subjects cooperate more when their partner is low-status than when their partners are high-status, in contrast with our prediction.

On the whole, the data from Study 2 are more consistent with our predictions. We also did not find any significant differences in the rates of cooperation between male and female subjects in the second experiment. This is an important result, because it suggests that gender was not a salient feature of our experimental design, and that is precisely what we anticipated. In addition, the fact that gender had no effect on cooperation in Study 2 bolsters our confidence in the results of the study, as it suggests that we succeeded in removing a potential confounding factor from the picture.

Discussion and conclusions

A new multilevel model explaining the effects of status differences on cooperation in one-shot 2-person PD was outlined in this article. It supports previous theoretical analyses that demonstrate the interrelationship between cooperation

and status phenomena. The results of the empirical test, however, suggest that the theoretical model must be revised and expanded. The observations from the second study are consistent with our hypotheses regarding the behaviour of high status actors. Our analyses show that high-status subjects indeed cooperate more with other high status partners and less with low status partners, which suggests that they form performance (cooperation) expectations based on salient status differences. Also, we did not find any interaction between status and the type of game; this too squares with our hypotheses. Our observations nevertheless are inconsistent with our hypotheses regarding the behaviour of low status subjects, whose decisions seem to be unaffected by information about their partners' status. Whether this means they did not form performance (cooperation) expectations, or that some other status related process (e.g., reward expectations) affected their decisions, requires further theorization.

The results presented above also illustrate the idea that the cumulative growth of knowledge can only be achieved when theoretical models and their empirical replicas are carefully constructed, and when empirical activity is theory-driven. The feedback between model, its experimental replica, and collected data stimulated additional questions and improvements to the experimental design. These improvements had a number of desired effects, including eliminating gender effects.

Note that we modified several features of the experimental design between study 1 and study 2. This precludes us from conclusively identifying the feature that contributed most to the differences in the results between the two experiments. Doing so was beyond the scope of our work, however, as we simply aimed to satisfy the scope conditions of our model, and to identify and remove any feature that might have compromised or confounded the results of the first experiment. Careful investigation into how a particular feature of our study design interacted with our major dependent variable – cooperation in the investment game – is interesting in its own right, as it may shed some light on how the interplay of status and other social mechanisms affect cooperation, but that question is well beyond the scope of our present study.

It was also beyond the scope of the present study to explain the high overall rates of cooperation in both of our experiments. Such high rates are not unusual. In studies of the effects of punishments (e.g., Fehr, Gintis 2007), and shame and honour (Jacquet et al. 2011) reported cooperation was also high. Our average cooperation rate was nevertheless higher than the average of 40-60% in classic PD settings. Again, investigating whether any particular feature of the setting (e.g., status manipulation, subjects solving a set of tasks before the game or simply sharing the same space) had such an effect is interesting in its own right. Our hypotheses did not specify particular cooperation levels, but rather the order of cooperation levels among pairs of interactants (differences between pairs). Therefore, regardless of whether the overall cooperation was high or low, our hypotheses regarding status differences' effects on the behaviours of participants could still be tested and indeed showed such effects for high status subjects.

A theory is never fully complete or finished. Our model can be further elaborated to include joint effects of nominal and status distinctions, as well as status

effects in coordination games and repeated PD. More importantly, however, as our theoretical model links status with expectations of trustworthiness using the concepts of fear and greed, future work on the model may benefit from employing variants of the prisoner's dilemma game in which only one of the two basic motivations are present. More specifically, in one variant of the game the fear component is removed, leaving only the greed component, while in another variant it is the other way around. Given that actors of different status respond differently to fear and greed component of the game, predictions concerning cooperation in the modified games can be developed, leading to a theoretical elaboration of the original model. The design described in the present paper can be thought of as a baseline or standardized setting with which results of the future studies can be compared.

Appendix

The description of the one-shot two-person Prisoner's Dilemma (PD) game[i]

In a PD game, actors choose between "cooperation" (C) and "defection" (D). The actors are assumed to be "rational" in the sense that they are motivated to maximize their own payoff. However, their payoff depends not only on the decision they make, but also on that made by their partner in the game. The payoff structure is conveniently presented in the form of a matrix, such as the one shown below.

Column / Row	Cooperation	Defection
Cooperation	R,R	S,T
Defection	T,S	P,P

There are two players and each of them has two options to choose from, so there are four possible combinations of their decisions, each such combination represented by a cell in the matrix. Each cell contains two numbers – the first corresponds to the payoff of the row player and the second to the payoff of the column player.

The payoffs in our matrix are represented by abstract symbols rather than specific amounts. The symbols are interpreted as *T*: the temptation to free ride on the partner's cooperation, *R*: the reward for mutual cooperation; *P*: the punishment for mutual defection; *S*: the sucker's payoff for unilateral cooperation.

The prisoner's dilemma game is defined by the following two conditions: (a)ii the ordering of payoffs is as follows: $T>R>P>S$; and (b)[iii] $2R>T+S$. Any set of numbers satisfying these constraints constitutes the payoff set in the PD.

Let us now consider what rational actors will do in the Prisoner's Dilemma game, beginning with Row's decisions. Suppose that Row expects Column to cooperate. Given this expectation, Row's best response is to defect, because $T>R$. Further, if Row thinks that Column will defect, Row's best response, again, is to defect, because $P>S$. Thus, defection is Row's dominant strategy in the game, as it gives Row a better payoff than cooperation, regardless of what Column does. Because the payoff matrix

is symmetric, the same reasoning applies to Column's decisions. Thus, rational decision makers arrive at (D,D) and each end up having P, which is clearly inferior to the group's optimal outcome (C,C) giving R to each player.

Given the payoff structure, we can define two parameters that represent the significance of two motivations for defection: *fear* and *greed*. Fear is defined as *P*–*S*, the difference between the payoff for mutual defection and unilateral cooperation, while greed is equal to the difference between the payoff for unilateral defection and mutual cooperation, *T*–*R*. In most application, the two parameters are equal, but it is possible to vary levels of fear and greed, as long as the defining constraints of the game are satisfied.

Notes added by the Editor

[i] Prisoner's Dilemma is a *two-person game in the normal (matrix) form.* The structure of such a game depends on what *outcomes* are assigned to all action (strategy) pairs; and on how the actors' *preference relations* on the set of all possible outcomes are related to each other.

[ii] This ordering of *payoffs* (outcome utilities) implies that *for Row* the (D,C) outcome, or the outcome produced by Row's unilateral defection, is better than the (C,C) outcome (the result of mutual cooperation). As a consequence, Row may yield to the temptation to defect if he has any reason to suspect that Column is going to cooperate. The (C,C) outcome is preferred in turn by Row to the (D,D) outcome (mutual defection). For Row, the worst (least preferred) of 4 outcomes is (C,D), or the case where Row (the 'sucker') gets exploited by Column. Column's preference relation on the set of outcomes is determined similarly. It is easy to see that the partners' preference relations, on the one hand, are in conflict (the best hope outcome of Row is the worst fear outcome of Column and conversely); on the other hand, the preferences of the players partially agree, as both Row and Column prefer (C,C) to (D,D). Therefore, both players are motivated to cooperate with each other, but at the same time they are tempted to seek each one's own most preferred outcome, which amounts to disregarding the interest of the partner. Prisoner's Dilemma is an example of a *mixed motive game.* The problem with PD is that *individual rationality* prompts to each player the use of *dominant strategy* D to the effect that the game ends up with the outcome (P,P) which is not Pareto optimal (because (R,R) is better than (P,P) for both players). An outcome is called *Pareto optimal* if there is no other outcome which makes at least one player better off without making the other player worse off.

[iii] The inequality (b) makes sense under the assumption that Row and Column use the same *utility scale* to evaluate the outcomes of their co-action. If so, the four *payoffs* T, R, P, S, having the same value for Row and Column, can be operationally defined in terms of a number of units of a certain resource that is valued similarly by both players. The numbers 2R=R+R, T+S, and 2P are then regarded as the amounts of that resource possible to be earned by the 2-person group. Notice that (b) and the inequality 2R>2P, which follows from (a), jointly imply that 2R is the *maximum total group payoff.* Under such an interpretation the PD game can also serve as a mathematical model for any social interaction system in which the actors, apart from pursuing their *individual goals*, may be *collectively oriented*, in other words, they may show concern for the benefit of the group as a whole.

References

Ahn T.K., Ostrom E., Schmidt D., Shupp R., Walker J. (2001). *Cooperation in PD Games: Fear, Greed, and History of Play*. Public Choice, 106, p. 137–155.

Aksoy O. (2015). *Effects of Heterogeneity and Homophily on Cooperation*. Social Psychology Quarterly, 78, p. 324–344.

Axelrod R. (1984). *The Evolution of Cooperation*. New York: Basic Books.

Berg J., Dickhaut J., McCabe K. (1995). *Trust, Reciprocity, and Social History*. Games and Economic Behaviour, 10, p. 122–142.

Berger J. (2007). *The Standardized Experimental Situation in Expectation States Research: Notes on History, Uses, and Special Features*. In: M. Webster Jr., J. Sell (eds.), *Laboratory Experiments in the Social Sciences*, San Diego: Academic Press, p. 353–377.

Berger J., Wagner D.G., Zelditch M., Jr. (1987). T*heory Growth, Social Processes, and Metatheory.* Working Paper no. 1987-2. Department of Sociology, Stanford University.

Berger J., Webster M., Jr. (2006). Expectation, Status, and Behavior. In: P.J Burke (ed.), Contemporary Social Psychological Theories. Stanford: Stanford University Press, p. 268–300.

Berger J., Zelditch M., Jr. (1997). *Theoretical Research Programs: a Reformulation*. In: J. Szmatka, J. Skvoretz, J. Berger (eds.), *Status, Network, and Structure. Theory Development in Group Processes.* Stanford, CA: Stanford University Press, p. 29–46.

Brewer M.B. (1979). I*n-group Bias in the Minimal Intergroup Situation: a Cognitive-Motivational Analysis.* Psychological Bulletin, 86, p. 307–324.

Brewer M.B., Kramer R.M. (1986). *Choice Behavior in Social Dilemmas: Effects of Social Identity, Group Size, and Decision Framing*. Journal of Personality and Social Psychology, 50, p. 543–549.

Cohen B.P. (1989). *Developing Sociological Knowledge: Theory and Method*, 2nd ed. New York: Nelson-Hall.

Cohen, B.P. (1980). *The conditional nature of scientific knowledge*. In: L. Freese (ed.), *Theoretical Methods in Sociology*. Pittsburgh, PA: University of Pittsburgh Press, p. 71–110.

Dawes R.M., Orbell J.M., Simmons R.T., Van De Kragt A.J.C. (1986). *Organizing Groups for Collective Action*. American Political Science Review, 80, p. 1171–1185.

Dawes R.M., Van De Kragt A.J.C, Orbell J.M. (1988). *Not Me or Thee but We: The Importance of Group Identity in Eliciting Cooperation in Dilemma Situations: Experimental Manipulations*. Acta Psychologica, 68, p. 83–97.

Fararo T., Skvoretz J. (1993). *Methods and Problems of Theoretical Integration and the Principle of Adaptively Rational Action*. In: J. Berger, M. Zelditch (eds.), *Theoretical Research Programs. Studies in the Growth of Theory.* Stanford, CA: Stanford University Press, p. 416–450.

Fehr E., Gintis H. (2007). *Human Motivation and Social Cooperation: Experimental and Analytical Foundations.* Annual Review of Sociology, 33, p. 43–64.

Foddy M., Yamagishi T. (2009). *Group-Based Trust*. In: K.S. Cook, M. Levi, R. Hardin (eds.), *Whom Can We Trust.* New York: Russel Sage Foundation, p. 14–41.

Foschi M. (1997). *On Scope Conditions*. Small Group Research, 28, p. 535–555.

Habyarimana J., Macartan H,, Posner D.N., Weinstein J.M. (2001). *Coethnicity. Diversity and the Dilemmas of Collective Action.* New York: Russell Sage Foundation.

Hogg M. (1996). *Intragroup Processes, Group Structure and Social Identity*. In: W.P. Robinson (ed.), *Social Groups and Identities.* Oxford: Butterworth-Heinemann, p. 65–93.

Jacquet J., Hauert C., Traulsen A., Milinski M. (2011). *Shame and Honour Drive Cooperation*. Biological Letters, 7, p. 899–901.

Karpiński Z., Skvoretz J. (2015). *Repulsed by the 'Other': Integrating Theory with Method in the Study of Intergroup Association*. Sociological Theory, 33, p. 20–43.

Kemper T.D. (1991). *Predicting Emotions from Social Relations.* Social Psychology Quarterly, 54, p. 330–342.

Kemper T.D. (1987). *How Many Emotions Are There? Wedding the Social and Autonomic Structure*. American Journal of Sociology, 93, p. 263–289.

Kollock P. (1998a). *Social Dilemmas: The Anatomy of Cooperation*. Annual Review of Sociology, 24, p. 183–214.

Kollock P. (1998b). *Transforming Social Dilemmas: Group Identity and Cooperation*. In: P.E. Danielson (ed.), *Modelling Rational and Moral Agents.* Oxford: Oxford University Press, p. 186–210.

Kramer R.M., Brewer M.B. (1984). *Effects of Group Identity on Resource Use in a Simulated Commons Dilemma.* Journal of Personality and Social Psychology, 46, p. 1044–1057.

Kumru C.S., Vesterlund L. (2010). *The Effect of Status on Charitable Giving*. Journal of Public Economic Theory, 12, p. 709–735.

Lawler E.J. Ridgeway C.L., Markovsky B. (1993). *Structural Social Psychology and the Micro-Macro Problem*. Sociological Theory, 11, p. 268–290.

List J.A. (2007). *On the Interpretation of Giving in Dictator Games*. Journal of Political Economy, 115, p. 482–493.

List J.A., Cherry T.L. (2008). *Examining the Role of Fairness in High Stakes Allocation Decisions*. Journal of Economic Behavior & Organization, 65, p. 1–8.

Lovaglia M. (1997). *Status, Emotions, and Structural Power*. In: J. Szmatka, J. Skvoretz, J. Berger (eds.), *Status, Network, and Structure. Theory Development in Group Processes.* Stanford, CA: Stanford University Press, p. 159–178.

Lovaglia M.J., Houser J.A. (1996). *Emotional Reactions and Status in Groups*. American Sociological Review, 61, p. 867–883.

Lovaglia M., Willer R.B. (2002). *Theory, Simulation, and Research: The New Synthesis*. In: J. Szmatka, M.J Lovaglia, K. Wysieńska (eds.), *The Growth of Social Knowledge*, Westport, CT: Praeger, p. 247–262.

Markovsky, Barry. 1994. *The Structure of Theories*. In: M. Foschi, E. Lawler (eds.), *Group Processes: Sociological Analyses*. Chicago, Ill: Nelson-Hall, p. 3–24.

Markovsky B. (1997). *Building and Testing Multilevel Theories*. In: J. Szmatka, J. Skvoretz, J. Berger (eds.), *Status, Network, and Structure. Theory Development in Group Processes.* Stanford, CA: Stanford University Press, p. 13–28.

Molm L.D., Takahashi N., Petersen G. (2000). *Risk and Trust in Social Exchange: An Experimental Test of a Classical Proposition*. American Journal of Sociology 105, p. 1396–1427.

Rapoport A., Chammah A.M. (1965). *Prisoner's Dilemma: A Study in Conflict and Cooperation*. Ann Arbor: University of Michigan Press.

Ridgeway C.L. (2014). *Why Status Matters for Inequality*. American Sociological Review, 79, p. 1–16.

Ridgeway C.L. (2006). *Status Construction Theory*. In: P.J. Burke (ed.), *Contemporary Social Psychological Theories.* Stanford: Stanford University Press, p. 301–323.

Ridgeway C.L. (2001). *Inequality, Status, and the Construction of Status Beliefs*. In: J.H. Turner (ed.), *Handbook of Sociological Theory*. New York: Springer, p. 323–340.

Ridgeway C.L. (1982). *Status in Groups: The Importance of Motivation.* American Sociological Review, 47, p. 76–88.

Sell J. (2007). *Social Dilemma Experiments in Sociology, Psychology, Political Science, and Economics*. In: M. Webster Jr., J. Sell (eds.), *Laboratory Experiments in the Social Sciences.* San Diego: Academic Press, p. 459–481.

Sell J. (1997). *Gender, Strategies, and Contributions to Public Goods.* Social Psychology Quarterly, 60, p. 252–265.

Shelly R.K. (2002). *How Scope and Initial Conditions Determine the Growth of Theory*. In: J. Szmatka, M.J. Lovaglia, K. Wysieńska (eds.), *The Growth of Social Knowledge*, Westport, CT: Praeger, p. 115–130.

Simpson B. (2006). *Social Identity and Cooperation in Social Dilemmas*. Rationality and Society, 18, p. 443–470.

Simpson B. (2004). *Social Values, Subjective Transformations, and Cooperation in Social Dilemmas*. Social Psychology Quarterly, 67, p. 385–395.

Simpson B. (2003). *Sex, Fear, and Greed: A Social Dilemma Analysis of Gender and Cooperation*. Social Forces, 82, p. 35–52.

Simpson B., Willer R., Ridgeway C.L. (2012). *Status Hierarchies and the Organization of Collective Action*. Sociological Theory, 30, p. 149–166.

Skvoretz J. (1998). *Theoretical Models: Sociology's Missing Links.* In: A. Sica (ed.), *What Is Social Theory? The Philosophical Debates*. Malden, MA: Blackwell Publishers, p. 238–252.

Tajfel H. (1982). *Social Psychology of Intergroup Relations*. Annual Review of Psychology, 33, p. 1–39.

Troyer L. (2002). S*ES V7.2.* A Computerized Version of the Expectation States Research Program's Standardized Experimental Setting.

Turner J.C. (1985). *Social Categorization and the Self-Concept: A Social Cognitive Theory of Group Behavior*. Advances in Group Processes, 2, p. 77–121.

Wagner D. (2000). *On the Irrationality of Rejecting Falsified Theories*. Sociological Focus, 32, p. 27–39.

Walker H.A., Cohen B.P. (1985). *Scope Statements: Imperatives for Evaluating Theory*. American Sociological Review, 50, p. 288–301.

Willer R. (2009). *Groups Reward Individual Sacrifice: The Status Solution to the Collective Action Problem*. American Sociological Review, 74, p. 23–43.

Wysieńska K., Szmatka J. (2002). *Positivism and Theory Construction in Group Processes*. In: J. Szmatka, M.J. Lovaglia, K. Wysienska (eds.), *The Growth of Social Knowledge.* Westport, CT: Praeger, p. 77–96.

Yamagishi T., Yamagishi M. (1994). *Trust and Commitment in the United States and Japan*. Motivation and Emotion, 18, p. 129–166.

Yamagishi T., Kiyonari T. (2000). *The Group as the Container of Generalized Reciprocity.* Social Psychology Quarterly, 63, p. 116–132.

Yamagishi T., Mifune N. (2008). *Does Shared Group Membership Promote Altruism? Fear, Greed, and Reputation*. Rationality and Society, 20, p. 5–30.

Yamagishi T., Jin N., Kiyonari T. (1999). *Bounded Generalised Reciprocity: Ingroup Boasting and Ingroup Favoritism*. Advances in Group Processes, 16, p. 161–197.

Yamagishi T., Makimura Y., Foddy M., Matsuda M., Kiyonari T., Platow M.J. (2005). *Comparison of Japanese and Australians on Group Based Cooperation.* Asian Journal of Social Psychology, 8, p. 173–190.

Modelowanie sytuacji społecznych: zaufanie i kooperacja pomiędzy nieznajomymi o nierównym statusie

Niniejszy artykuł opisuje proces konstruowania i testowania nowego modelu teoretycznego wyjaśniającego, jak nierówności pod względem statusu wpływają na poziom kooperacji między aktorami uczestniczącymi w jednokrotnie rozgrywanym Dylemacie Więźnia (DW). Łącząc idee pochodzące z teorii charakterystyk statusu i działań zbiorowych, przewidywaliśmy, że w DW, w którym aktorzy są widocznie zróżnicowani pod względem statusu, poziom kooperacji będzie zależał od statusu aktora w relacji do statusu jego partnera, a także tego, czy gra ma charakter symultaniczny czy sekwencyjny. Aby zademonstrować wyzwania związane z konstruowaniem i testowaniem nowych modeli teoretycznych, przedstawiamy dwa różne skonstruowane przez nas eksperymenty dotyczące dwuosobowych jednokrotnie rozgrywanych sytuacji DW. Artykuł wieńczy dyskusja o charakterze substantywnym i metodologicznym.

Słowa kluczowe: status, spolegliwość, Dylemat Więźnia, budowa teorii

FOLIA 202

Annales Universitatis Paedagogicae Cracoviensis

Studia Sociologica VIII (2016), vol. 1, p. 143–165
ISSN 2081–6642

Szymon Czarnik
Jagiellonian University, Poland

Reading Minds of Experimental Subjects. Insights from Pre- and Post-Experimental Surveys in a Redistribution Game Experiment[1]

Abstract

Subjects' responses in pre- and post-experimental questionnaires are utilized to elucidate their behaviour in an asymmetric Prisoner's Dilemma redistribution game with no communication between players. We find experimental subjects to be predominantly negative in their assessment of intentions behind their partners' decisions while describing their own motivations as rationally self-interested, reciprocal, and efficiency-oriented. Thus, in the absence of communication, negative intent attribution may be one of the crucial reasons behind failure to establish lasting cooperation, even in situations where both players are well aware of its benefits and the behaviour necessary to achieve it. We also find some evidence that, on average, players conscious of Pareto-optimizing potential brought about by mutual adoption of Tit-for-Tat strategies make more friendly decisions in the game. Lastly, we consider the study's ecological validity in the light of subjects' post-experimental statements.

Key words: behavioural game theory, Prisoner's Dilemma, tax redistribution, cooperation, interpretation of intentions

The experiment

In May 2005, I conducted an experimental study to investigate redistributive behaviour in a situation where unequal incomes, subject to subsequent redistribution, were rightfully earned by participants (rather than allotted randomly or on the basis of some kind of disputable criterion like a score in a quiz).

A few days before the experiment proper, all subjects attended a single group meeting with the researcher. Apart from introducing everyone to the basic framework of the experiment, the meeting provided an opportunity to collect some relevant additional data. Subjects filled in a questionnaire containing a number of items related to real-life redistribution themes. They made a series of monetary choices wherewith their aversion to payoff inequality was measured. They were put in a position to reveal ('behind the veil of ignorance') their beliefs about what constituted a fair initial payoff distribution in the experiment. Finally, they had to

[1] The research on which this paper is based was financed by the Polish Ministry of Education and Science as part of the project 1 H02E 046 28.

do the hard work to earn money to be subsequently used in redistribution games. The job to be done was deciphering either one, or two, or four pages of coded text, with 15 PLN paid per page. Thus 72 subjects, 36 male and 36 female, were randomly divided into three 12 person male groups and three 12 person female groups with income levels of 15, 30, and 60 PLN, respectively. Twelve experimental sessions (six all-male and six all-female) were scheduled, with two persons from each income level taking part in any single session. Within each session twelve-round redistribution games were to be played between subjects of unequal initial payoffs, i.e. 15–30, 30–60, or 15–60 PLN.

Directly before playing experimental games subjects watched a graphical presentation explaining the mechanism of tax redistribution and the rules of the game in an intuitive way. It is important to stress that terms like 'game,' 'players,' 'strategy,' etc. were never used in communication with the subjects. The whole research situation was explicitly described as an investigation of real-life monetary decisions affecting both one's own and someone else's earned money.

The redistribution mechanism

Redistribution mechanism used in the game may be thought of as an implementation of a negative income tax (NIT) proposed by Milton Friedman back in the 1960s (Friedman 2002). The basic idea is that taxes would be paid only by people with incomes above a certain threshold value, while those below the threshold would pay a 'negative tax,' i.e. receive a subsidy from the budget. The amount of tax paid (subsidy received) would in turn depend on how much person's income exceeds (falls short of) the threshold (with those at the threshold breaking even).[2]

An implementation proposed here would operate in three steps:

1. Everyone pays a linear income tax on his or her initial income.
2. Part of total tax revenues is 'lost,' i.e. taken away to cover the cost of tax collecting and redistribution.
3. What remains in the budget is divided equally among all persons in a form of lump-sum subsidies.

To analyse the logic of this redistribution mechanism, let us denote person's *i* initial payoff by p_i, a linear tax rate by T and a fiscal cost, or share of revenues lost in the process, by C. Then each person's final income, i.e. income after paying a tax and receiving a subsidy, may be construed as consisting of two parts:

$$p'_i = (1 - T)p_i + (1 - C)T\bar{p}$$

[2] A rationale behind NIT was to help low-income workers in a way that would minimize distortions in the market. As Friedman put it, 'Like any other measures to alleviate poverty, it reduces the incentives of those helped to help themselves, but it does not eliminate that incentive entirely' (p. 192). However, the idea met with severe criticism from the word go (see, e.g. *Fallacies of the Negative Income Tax* in Henry Hazlitt's *Man vs. the Welfare State*, 1969).

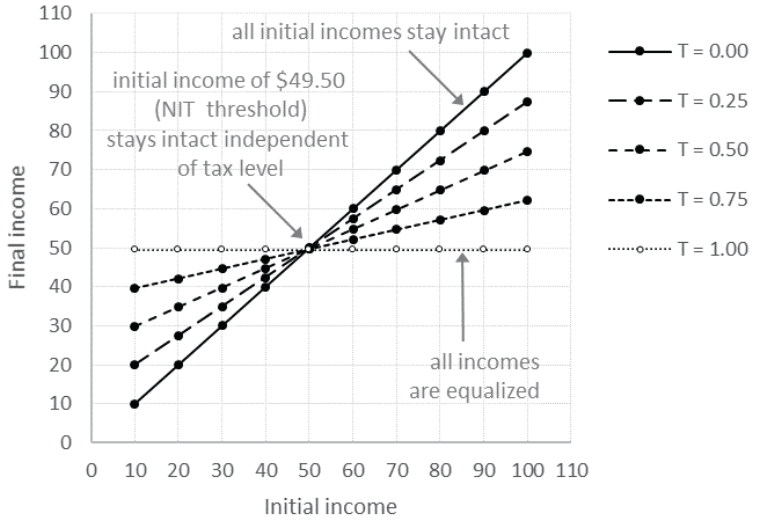
$(1- T)p_i$ is a part that depends on person's *i* initial income. To be exact, it is a part of the person's initial income that is kept after paying a linear tax. $T(1 - C)\bar{p}$ is a lump-sum subsidy from the budget, which is the same for all persons. Obviously, it does not depend on individual initial income but instead it depends on $\bar{p}$, the average initial income in the population.[3] With T equal to 0, all persons would stay with their initial incomes, and with T equal to 1 all incomes would be equalized at the level of $(1 - C)\bar{p}$.

The effective amount of tax paid by player *i* is a difference between his initial income and his final income, or alternatively a difference between the amount paid by him in the form of a linear tax, Tp_i, and the lump-sum subsidy received from the budget, $(1 - C)T\bar{p}$:

$$\tau_i = p_i - p'_i = Tp_i - (1 - C)T\bar{p}$$

A straightforward calculation shows that tax redistribution benefits a person, namely $p'_i > p_i$ (which is tantamount to $\tau_i < 0$), if and only if $p_i < (1 - C)\bar{p}$. $(1 - C)T\bar{p}$ is then a threshold value of negative income tax.[4] At the same time, the amount lost in the process of tax redistribution, which we shall henceforth call a *net social loss*, is $\lambda = CT\sum p_i$.

Figure 1. Tax redistribution at different tax levels and fiscal cost 10%



[3] Tax revenues available for redistribution are equal to $(1 - C)\sum Tp_i$. As they are divided evenly among all persons, a lump-sum subsidy may be expressed as $(1 - C)\sum Tp_i/n$, or $T(1 - C)\bar{p}$.

[4] It may be noted in the passing that with sufficiently large fiscal cost *C* only few persons, or even nobody at all, would benefit from tax redistribution. With *C* equal to 1, all tax revenues would be lost and no subsidies sent back.

By means of an example, Figure 1 shows how redistribution at various tax levels (and at fixed 10% fiscal cost) would affect initial incomes of ten persons, earning \$10, \$20, \$30, ..., \$100, respectively. Average initial income in the group is \$55, and break-even NIT threshold is \$49.50 (0.9 times \$55).

The experimental redistribution game

In the experiment, dyads formed by persons of unequal initial payoffs were playing a redistribution game with NIT mechanism described previously. Three combinations of initial payoffs were possible: 15 PLN vs. 60 PLN, 15 PLN vs. 30 PLN, and 30 PLN vs. 60 PLN. In each round of the game, both a high- and a low-earner (henceforth denoted by H and L, respectively) had to make two choices. First, they had to cast a secret vote on the preferred level of linear tax redistribution, with 0% tax leaving the initial payoffs intact and 100% tax making both payoffs equal. The votes were then revealed and effective tax level was set to the average of the two proposals, $T = (t_L + t_H)/2$. After incomes had been redistributed accordingly, players had an opportunity to make free gifts to one another. The crucial element of the game was that voluntary transfers were fully efficient, whereas tax transfers involved a fiscal cost leading to either 10 or 30% of 'tax revenue' leaked in the process.

In all experimental games, L's initial payoff was below NIT threshold and therefore L was in a position to gain from tax redistribution at the expense of H.[5] Namely, low earners were effectively paying a negative tax in the amount of $\tau_L < 0$, whereas high earners were effectively paying a positive income tax in the amount of $\tau_H > 0$.[6] As a matter of fact, τ_H was necessarily utilized both to subsidize L and cover the inherent net social loss, $\tau_H = |\tau_L| + \lambda$.

Charts in Figure 2 show the outcomes of tax redistribution for each type of dyad with tax level T at 50% (being the result of L voting maximum 100% tax, and H voting no tax at all), and fiscal cost either 10, or 30%. Naturally, with no redistributive taxation (T=0), players would stay with their initial payoffs, no matter the fiscal cost.

Finally, after NIT mechanism had been put into effect, players could offer free monetary gifts to one another and thus their final incomes were $p''_i = p'_i - g_i + g_j$, where g_i and g_j are voluntary gifts offered by players i and j respectively.

Now we should recognize that the structure of the experimental redistribution game is essentially that of a *repeated sequential asymmetric continuous-strategy Prisoner's Dilemma*. Cooperative decisions in the game consist in L voting first for low or no taxes, and then H offering him a sufficiently high voluntary gift. By virtue of such cooperation, net social loss can be minimized, or even eliminated. Thus,

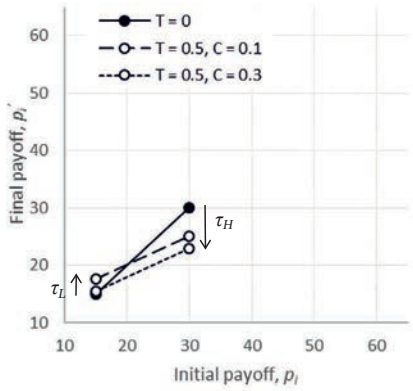
[5] With L's payoff underlined, NIT thresholds for 15–30 PLN dyad are 20.25 (fiscal cost 10%) and 15.75 (30%); for 15–60 PLN they are 33.75 (10%) and 26.25 (30%), and for 30–60 PLN they are 40.50 (10%) and 31.50 (30%).

[6] The exact amount is given, as in the general case, by $\tau_i = Tp_i - (1-C)T\bar{p}$, with $\bar{p} = (p_L + p_H)/2$.

as characteristic of Prisoner's Dilemma, it is in the best interest of both parties to enter into mutually beneficial cooperation by substituting costly tax redistribution with efficient voluntary redistribution. This, however, is not a trivial exercise, as it requires L to renounce his tax benefits, and H to give up a portion of his rightfully earned incomes.[7] In effect, under standard assumptions of rational egoism, the unique equilibrium outcome in a single round of this game is 50% tax redistribution with no voluntary redistribution at all. This is brought about by L voting for maximum tax ($t_L = 1$) and offering no gift ($g_L = 0$), and H voting for minimum tax ($t_H = 0$) and offering no gift either ($g_H = 0$). The equilibrium is clearly Pareto suboptimal due to net social loss λ inherent in tax redistribution.

Figure 2. Tax redistribution in the experimental dyads

15 PLN vs. 30 PLN



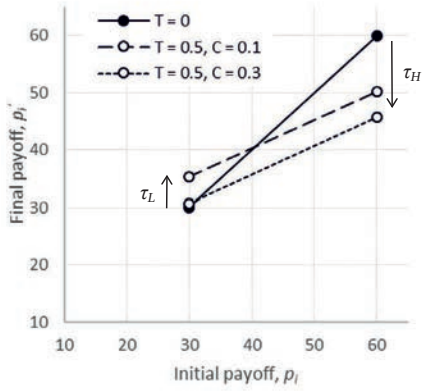
$\tau_i = p_i - p_i'$

High-earner pays a positive tax in the amount of $\tau_H > 0$

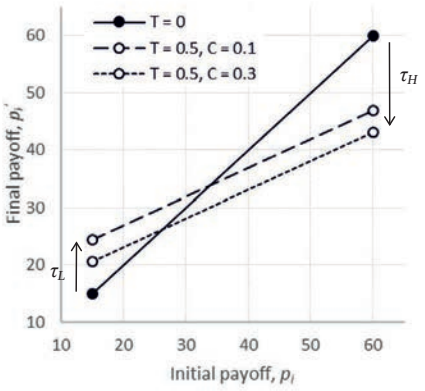
Low-earner pays a negative tax in the amount of $\tau_L < 0$ (receives a net subsidy)

$\lambda = \tau_H - |\tau_L|$
The difference between amount paid by H and amount received by L is a net social loss, λ.

30 PLN vs. 60 PLN



15 PLN vs. 60 PLN



[7] Remember that initial incomes were directly proportional to the amount of work done at the pre-experimental meeting. This created a strong sense of entitlement as most subjects ('behind the veil of ignorance') considered proportional compensation to be a fair allocation rule.

All experimental encounters between low- and high-earners consisted of 12 rounds of thus defined redistribution game. Since the total number of rounds was common knowledge, by means of backward induction the dismal high-taxes no-charity status quo was the unique equilibrium of the whole 12-round game as well.[8] Players' equilibrium gains and losses in 6 experimental types of games are juxtaposed in Table 1 (graphically, these results were presented in Figure 2).

Table 1. Theoretical equilibrium outcomes of experimental redistribution games

Dyad types (p_L–p_H)	15–30 PLN		30–60 PLN		15–60 PLN	
Fiscal cost (C)	10%	30%	10%	30%	10%	30%
L's final payoff (p''_L)	17.63	15.38	35.25	30.75	24.38	20.63
H's final payoff (p''_H)	25.12	22.87	50.25	45.75	46.87	43.12
L's gain ($\lvert\tau_L\rvert$)	2.63	0.38	5.25	0.75	9.38	5.63
H's loss ($\tau_H = \lvert\tau_L\rvert+\lambda$)	4.88	7.13	9.75	14.25	13.13	16.88
Net social loss (λ)	2.25	6.75	4.50	13.50	3.75	11.25
Efficiency of tax transfer ($\lvert\tau_L\rvert/\tau_H$)	54%	5%	54%	5%	71%	33%

In the experiment proper, a member of any payoff category X played first a doubleheader against a member of Y category (one game at 10%, the other at 30% fiscal cost), and then proceeded to a doubleheader with a partner from Z category.[9] For instance, one possible path for a 15 PLN earner was to play games in the following order:

15–30 PLN (cost 10%) → 15–30 PLN (cost 30%) → 15–60 PLN (cost 10%) → 15–60 PLN (cost 30%)

Main experimental results

Table 2 lists main results from actual experimental redistribution games. These were succinctly commented by the author in his earlier paper (Czarnik 2006), while a detailed analysis is to be found in his PhD thesis (Czarnik 2007).

[8] To be sure, this is characteristic of all finitely-repeated versions of the classic prisoners' dilemma. As Kreps et al. put it back in 1982, 'This game has a unique Nash equilibrium path, which involves each player choosing to fink at every stage… This outcome is clearly and dramatically inefficient.' Then they go on to contrast this with actual empirical evidence of human subjects' behaviour: 'This uniqueness result is disturbing in light of experiments with this game, of which there have been a very large number… A common pattern in these experiments is that, at least for some time, both players cooperate and, in the process, end up with payoffs that are strictly greater than they would obtain under equilibrium play.' (Kreps et al. 1982, p. 2).

[9] Doubleheaders were played within fixed-pairs. However, subjects had not been informed that they played two consecutive games against the same person. As revealed in a post-experimental survey, most of them actually believed they were matched with a new person each time.

Table 2. Main results from experimental redistribution games

Dyad types (p_L–p_H)	15–30 PLN		30–60 PLN		15–60 PLN	
Fiscal cost (C)	10%	30%	10%	30%	10%	30%
L's mean tax vote (t_L, %)	56.73	40.92	58.82	53.99	78.55	74.79
H's mean tax vote (t_H,%)	1.33	2.22	1.77	1.79	4.43	5.12
L's final payoff (p''_L, PLN)	16.90	15.80	35.36	32.76	23.14	19.80
H's final payoff (p''_H, PLN)	26.80	26.29	51.91	49.72	48.76	46.21
H's mean gift (g_L, PLN)	0.38	0.64	2.18	2.34	0.35	0.30
Mean social loss (λ, PLN)	1.30	2.91	2.73	7.52	3.10	8.99
Number of games	12	11*	12	12	11*	12

*One game was lost due to technical problems

Briefly, the results can be summarized as follows:

1. Subjects' behaviour deviated substantially from equilibrium play. Nonetheless, for the most part their decisions were narrowly self-interested and far from reaching Pareto-optimal cooperation. Moreover, most of low-earners' self-serving tax decisions were in direct violation of their fairness judgments, as elicited in the pre-experimental procedure under the Rawlsian 'veil of ignorance'.[10]
2. Tax-imposed redistribution was most severe where initial income disparity was the largest, namely in 15–60 PLN dyads.
3. Increased inefficiency of tax redistribution (30% versus 10% fiscal cost) had next to no effect in games involving the highest earners (60 PLN). Only in 15–30 PLN games larger fiscal cost induced somewhat more restrained tax voting.
4. Except for 15–60 PLN games, in majority of rounds voluntary donations tended to be crowded out by fiscal redistribution. This crowding-out effect was evidenced by statistically significant negative correlation between the level of Ls' tax votes and the amount of Hs' free gifts.

   Results 1–4 refer to the data from the first two games played within fixed pairs, one under 10% and the other under 30% tax cost. After these two games subjects were matched again to play a doubleheader with another person. Players involved in 15–30 PLN games were matched against 60 PLN, those involved in 30–60 PLN games were matched against 15 PLN, and those in 15–60 PLN games against 30 PLN.
5. In second-partner games, a powerful history effect was revealed. The relationship between income inequality and amount of tax redistribution (see point 2 above) was completely reversed. 15 PLN earners who experienced a large income disparity in their first doubleheaders against 60 PLN and acquired a taste for heavy tax redistribution, had no qualms about voting high taxes in their

[10] For more details, see Czarnik 2009.

subsequent encounters against 30 PLN. On the other hand, those 15 PLN earners who practiced moderate tax voting in their first doubleheaders against 30 PLN, just went on with their moderate approach in subsequent games against 60 PLN.

Insights from pre- and post-experimental surveys

Pre-experimental survey included questions about causes behind wealth and poverty, income redistribution by government, relations between the rich and the poor, as well as some personal data (sex, academic major, year of studies, financial situation). The post-experimental survey consisted of six open-ended questions:

1. All in all, are you satisfied to have taken part in the experiment?
2. Did you know personally any person co-participating in your experimental session?
3. Do you think you were matched with the same person twice?
4. Please, summarize shortly the behaviour of persons you were matched with.
5. What were your concerns when making decisions about tax level and amount of free transfer to the other person?
6. If you could communicate with the other person to establish a common way of conduct, what would you propose?

Satisfaction was universal among experimental subjects, with only two out of seventy-two participants responding ambiguously. Eighteen subjects had known personally someone who participated in their experimental session, but only three of them thought they had been matched with that person. Nearly half of the experimental subjects thought they had been paired with a different partner each time. Thus, we may consider experimental instruction to be successful in dissuading subjects from treating two consecutive games as one meta-game. If pairs were drawn randomly, the probability of being matched with a different person each time would be only 25%, and yet nearly each second subject thought this had been the case.

Questions 3 to 5 provide much more substantial information as they allow us an insight into how experimental subjects interpreted their own and their partners' behaviour.

Intentions attributed to other parties

Each subject's description of other people's behaviours was dissected into distinct themes and categorized according to the same coding scheme. Some respondents were quite laconic while others had their statements categorized into as many as four different groups. Those detailed categories were in turn collapsed into three major groups depending on the effect other people's actions had on the respondent:

- *positive*, i.e. suggesting other party's friendly attitude or readiness to cooperate;
- *negative*, i.e. suggesting other party's hostile intent, disregard for cooperation, or erratic behaviour;

- *neutral*, i.e. neither beneficial not detrimental per se, mainly referring to rational pursuit of other party's own interest.

It should be noted that subjects were making a single statement concerning all their partners. For this reason, it is not surprising that some statements included what would otherwise seem contradictory opinions, i.e. both positive and negative.[11]

Table 3 contains the frequencies of persons subscribing to particular themes, separately for subjects with initial earnings of 15, 30, and 60 PLN.

Table 3. Interpretations of motives behind partners' behaviour provided by 15, 30, and 60 PLN earners*

–	negative	16
·	neutral	8
+	positive	6

–	negative	16
·	neutral	10
+	positive	8

–	negative	16
·	neutral	7
+	positive	9

	15 PLN earners (n=22)	**N**
–	stinginess of high-earners	11
·	reciprocation	6
+	(positive)	3
–	(negative)	3
·	own interest	4
–	lack of rationality	4
+	generosity of high-earners	3
–	lack of reciprocation	3
·	rationality	2
·	minimization of own losses	1
–	chaos	1
·	(it depends)	1

	30 PLN earners (n=23)	**N**
–	covetousness of low-earners	11
+	reciprocation (positive)	6
–	lack of reciprocation	6
·	(it depends)	5
–	stinginess of high-earners	4
–	lack of rationality	4
·	rationality	3
–	inefficiency	3
·	minimization of own losses	2
·	own interest	1
+	generosity of high-earners	1
+	restraint of low-earners	1
·	superficial lack of rationality	1
·	caution	1
+	group efficiency	1
·	income equalization	1

	60 PLN earners (n=24)	**N**
–	covetousness of low-earners	11
·	reciprocation	8
+	(positive)	6
–	(negative)	2
–	lack of rationality	7
·	(it depends)	3
·	own interest	2
–	inefficiency	2
+	restraint of low-earners	2
+	generosity of low-earners	1
·	rationality	1
·	minimization of own losses	1
–	lack of reciprocation	1
·	caution	1

* Multiple categories possible. Last column in each table displays numbers of persons whose comments fitted particular categories.

First thing to notice is the prevalence of negative sentiments in all earning groups.[12] In the 15 PLN group, the main complaint was about the 'richer' subjects unwilling to share (*'Generally, persons who had most money were more stingy'*). On the other hand, a typical comment from a 60 PLN earner complained that *'most*

[11] For example, one 60 PLN earner's description of his partners found its way to both 'covetousness of low-earners' and 'generosity of low-earners' categories. He stated that some of his partners '*acted as if they were willing not so much to improve their own payoffs but instead to make me lose as much as possible*,' which referred to a doubleheader with a low-earner trying to extract maximum amount in taxes, irrespective of the cost it inflicted on his partner. Then he also said that '*the last one was completely different – low taxes and on top of that he was transferring some money to me even though he had less than I.*'

[12] Not only the largest number of persons provided negative responses but also subjects were most likely to give more than one negative label to their partners' behaviour. All in all, 48 persons offered 73 negative labels (on average 1.52 per person), 25 persons offered neutral labels (on average 1.20 per person), and 23 persons offered 24 positive labels (on average 1.04 per person).

participants employed tactics to rob me of my hard-earned possessions.' Both complaints certainly had substance to them. In the 15 PLN group, 6 out of 11 accusations of the 'rich' being stingy came from subjects who had received the lowest average gift transfers from high-earners (ranging from 0.02 to no more than 0.19 PLN). Similarly, in 60 PLN group, 6 out of 11 accusations of the 'poor' being covetous came from subjects who had suffered the highest average tax vote by low-earners (ranging from 82 to 94%).

Subjects in the 30 PLN group were in a somewhat special position – in the course of experimental session they switched between being low-earners (against 60 PLN) and high-earners (against 15 PLN). It is telling that 11 of them were willing to disparage behaviour of low-earners whereas merely 4 of them condemned high-earners for their stinginess. This may serve as an anecdotal illustration of Kahneman and Tversky's prospect theory (1979): gains and losses are not perceived symmetrically. In this case, losses incurred by high level of enforced redistributive taxation gave rise to more negative comments than gains unrealized due to better-off partners' reluctance to share.

Apart from 'stinginess of the rich' and 'covetousness of the poor,' the third most frequent observation referred to reciprocal actions of one's partners. More often than not, reciprocation mentioned was cooperative in character and consisted in some kind of higher-gifts-for-lower-taxes agreement.

Motivations behind subjects' own actions

As far as subjects' own motivation is concerned, three themes came to the fore in all three earning groups:

- own interest ('*I was willing to secure myself the highest possible gains*');
- reciprocity ('*by transferring small amounts I tried to induce other participants to lower their tax proposals*'; '*if somebody was kind enough to give me some amount freely I was lowering my tax vote, and vice versa*');
- group efficiency ('*let's not introduce tax or we'll lose some part of our common pool of money*').

It is noteworthy that subjects in the lowest earning 15 PLN group were most explicit about being guided by their own self-interest (see Table 4). As a matter of fact, they admitted self-interest three times more often than reciprocity (in proportion 17:5), whereas in the other two groups self-interest and reciprocity were on the same footing (9:9 in 30 PLN, and 11:13 in 60 PLN). This seems to be derivative of the particular structure of the redistribution game used in the experiment. It was in high-earners' interest to uphold the status quo of initial earnings and so their self-interest did not manifest itself in any payoff-changing actions. For low-earners, on the contrary, any positive tax vote disturbed status quo to their advantage and thus was a self-evident example of acting on their own interest[13]. By the same

[13] One may also convincingly argue that status quo, completely apart from being in the interest of high-earners, constituted a fair distribution of payoffs. During the pre-experimental meeting subjects, yet unaware of their particular roles in the experiment, were asked to

token, initiating reciprocal substitution of high-taxes with free-giving was clearly a self-defending action on part of high-earners, even as it could be to the benefit of the other party as well.

Table 4. Motivations behind subjects' decisions

15 PLN earners (n=23)	**N**
own interest	17
group efficiency	6
reciprocity	5
fairness	2
equalization of incomes	2
reducing harm to others	1
minimization of own losses	1
partner's income[13]	1

30 PLN earners (n=23)	**N**
own interest	9
reciprocity	9
group efficiency	8
minimization of own losses	4
helping low-earners	3
(it depends)	3
partner's income[13]	2
equalization of incomes	2
fairness	1

60 PLN earners (n=24)	**N**
reciprocity	13
own interest	11
group efficiency	6
minimization of own losses	5
helping low-earners	1
reducing harm to others	1
partner's income[13]	1
equalization of incomes	1

It is sobering to realize that fairness as a guiding principle was invoked explicitly only by a single subject:

> [I based my choices] *on actual income difference and the amount of work performed (to make it more or less fair).*

Merely two more subjects implicitly hinted at respecting, at least to some extent, other people's initial incomes as fairly earned:

> *In the course of time I came to the conclusion that my partner's 60 PLN wasn't a 'windfall gain' and that he had to do some work on it so after the second round I started setting tax at low level.*

> [I sought] *equal split but not in the absolute sense as if the person who had earned her sum of money had to share with me only because I was less lucky.*

Scant presence of direct fairness considerations in subjects' statements provides much food for thought as certainly everybody was well aware that initial payoffs were earned in proportion to participants' own efforts instead of being distributed haphazardly. It is even more revealing given that one of the pre-experimental tasks was to decide upon fair allotment of initial payoffs in the experiment.

---

divide 105 PLN between three persons burdened with decoding 1, 2 and 4 pages of coded text (tasks subjects themselves were soon afterwards asked to perform to earn their initial payoffs). 60% decided that 15:30:60 was a fair distribution in such a situation and another 15% thought it fair to make distribution of payoffs even more unequal (e.g. 10:25:70).

[14] This category includes statements to the effect that subjects conditioned their choices on the their partners' initial incomes. However no motive (either explicit, or implicit) to equalize incomes was mentioned.

Hypothetical communication between players

The last question in the post-experimental survey was about hypothetical communication between players. What common way of conduct would they propose to their partners?

It is clear from subjects' responses that their primary concern in hypothetical negotiations would be for reaching some sort of Pareto-optimal agreement. All in all, 49 persons made comments to that effect, with 35 of them explicitly mentioning a trade-off between free gifts and taxes as a mechanism for achieving Pareto-optimality[15]. The tax-gift substitution may be construed here as a friendly Tit-for-Tat strategy in Iterated Prisoner's Dilemma (Axelrod 1984).

The following are three examples of comments that either express concern for group efficiency in general, or directly propose a reciprocal solution, or both:

- [I would suggest] *optimization of profits, i.e. my objective would be that both I and the other person get as much as possible, and the state* [= tax collector] *gets the least*;
- *that she votes no taxes, and I give her as much as I would have to pay in taxes anyway*;
- *transferring money instead of taxes – then together we don't waste anything on tax costs – more money could be distributed in a more satisfying way.*

These and other hypothetical communications are categorized in Table 5.

Table 5. Common way of conduct proposed in hypothetical communication*

Ideas proposed	Subjects' initial income			Total (n=70)
	15 PLN (n=23)	30 PLN (n=23)	60 PLN (n=24)	
substituting taxes with gifts (Tit for Tat)	11	12	12	35
concern for group efficiency	4	9	7	20
equalization of incomes	4	2	2	8
(no sensible idea)	3	1	2	6
fairness	1	1	1	3
upholding status quo	0	3	0	3
undefined consensual solution	0	1	1	2
(it depends)	1	0	1	2
minimization of own losses	1	0	0	1
more rationality	1	0	0	1

* Multiple categories possible for a single person.

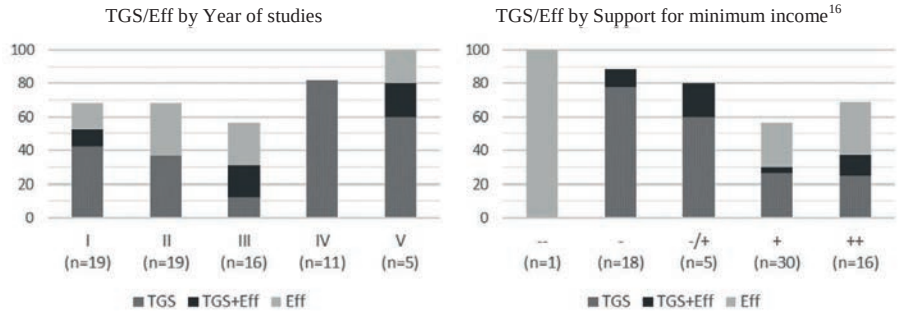
[15] To be sure, desire for group efficiency is implicit in tax-gift substitution proposal. On the other hand, concern for efficiency does not necessarily imply tax-gift reciprocation (for instance, a high-earner could think of sticking to status quo payoffs as a means of avoiding inefficiency inherent in the tax system).

Pareto-optimizing reciprocal solution, i.e. substituting taxes with gifts, was explicitly recognized in all earning groups with equal frequency. However, it is interesting that it depended heavily on two factors: age (as measured by year of studies) and subjects' opinion on state-guaranteed minimum income[16]. In Figure 3, we analyse those relationships by categorizing efficiency-concerned subjects into three groups:

1. Those who simply stated that they would like to implement tax-gift substitution [TGS];
2. Those who elaborated on TGS being a means for achieving efficiency [TGS+Eff];
3. Those who merely declared their desire for efficiency without explicit reference to TGS [Eff].

First two of the above categories comprise of experimental subjects explicitly advocating mutual adoption of Pareto-optimizing Tit-for-Tat strategies.

Figure 3. Prevalence of tax-gift substitution (TGS) and concern for group efficiency (Eff) in hypothetical communication by year of studies and support for state-guaranteed minimum income



Clearly, we observe a rapid increase in advocacy for Tit-for-Tat strategies after third year of studies which may be interpreted as an indication of older subjects being more strategically sophisticated.

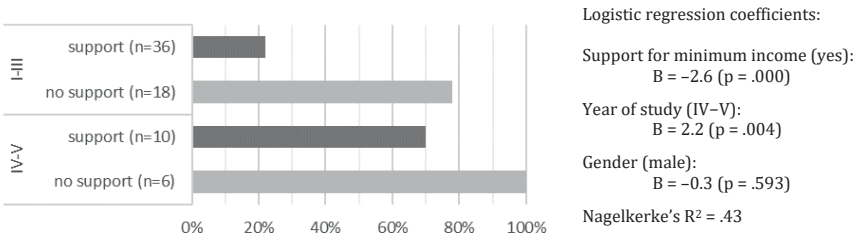
More interestingly, there has been a vast disparity in TGS reciprocation between supporters of state-guaranteed basic income and those uncertain or opposed to it. Explicit Tit-for-Tat advocacy was more than twice less prevalent in the former group's hypothetical communication.

The effect of support for state-guaranteed minimum income (as presented in Figure 4) turns out to be statistically significant predictor of Tit-for-Tat advocacy in logistic regression model with control for subjects' gender and year of study. Certainly, no definite answer can be given to explain why such an effect should take place. Nonetheless, one may surmise that persons conceiving of basic income as

[16] In the pre-experimental questionnaire subjects were asked to express their opinion on the following statement (adopted from Polish General Social Survey): 'The government should provide everyone with a guaranteed basic income.' Available responses were: 'strongly agree (++) / agree (+) / neither agree, nor disagree (-/+) / disagree (-) / strongly disagree (--).'

a kind of human right could be somewhat reluctant to discuss the redistribution problem in terms of bargaining. They would be more inclined to invoke efficiency in general[17], or even to make a direct appeal to equality[18]. Now all this, of course, may be a peculiarity of the present study. Still, it is plausible that subjects' socio-political beliefs about real-life issues concerning income distribution should affect their attitude in experiments, especially in a game overtly construed as a redistributive tax system under democratic rule.

Figure 4. Tit-for-Tat advocacy by year of study (I–III vs. IV–V) and support for minimum income law



What if communication was possible?

There was no experimental condition allowing for communication between players but it is instructive to analyse if there is any difference in actual behaviour between players who (post factum) advocated the adoption of Tit-for-Tat strategies and those who did not. It turns out that there is some. At the same time there are clear-cut examples of how inability to communicate may be a hardly passable stumbling block on the path to cooperation. This is in accord with evidence from Prisoner's Dilemma and public good experiments, where 'preplay communication, which should have no effect in theory, is the non-payoff variable that raises the rate of cooperation by the most' (Camerer 2003, p. 46).

In Table 6, we compare average decisions made in redistribution game by Tit-for-Tat advocates (TFT) with those made by other players (non-TFT). As Tit-for-Tat is equivalent to tax-gift substitution, we focus our attention on low-earners' tax votes and high earners' voluntary gifts. To be sure, the 15 PLN were low-earners throughout the experiment just as the 60 PLN were high-earners. In contrast, the 30 PLN were switching roles: they were low-earners when matched against the 60 PLN (in such a setting we analyse their tax vote) and high-earners when matched against the 15 PLN (in such setting we consider their voluntary gift).

[17] Indeed, 13 out of 14 persons expressing concern for efficiency without direct reference to Tit-for-Tat strategies ('Eff' boxes in Figure 3) were proponents of state-guaranteed minimum income.

[18] Out of 8 hypothetical communications calling for 'equalization of incomes,' 7 came from proponents of basic income law.

Table 6. Low-earners' tax vote and high-earners' voluntary gift. Decisions made by Tit-for-Tat advocates (TFT) and the rest (non-TFT)

Decision		**Tax vote**			**Voluntary gift**		
Decision-maker: Player		15 PLN		30 PLN		60 PLN	
Partner		30 PLN	60 PLN	60 PLN	15 PLN	15 PLN	30 PLN
Player's type	TFT	50.4	39.2	59.3	0.58	3.55	3.84
	non-TFT	68.1	79.9	45.9	0.78	0.51	0.39
Difference		-17.7	-40.7	13.4	-0.2	3.0	3.4
*p**		0.193	0.007	0.288	0.683	0.062	0.078

*Exact significance (2-tailed) in Mann-Whitney test. The size of each group was between 10 and 12.

In Table 6, we compare average decisions made in redistribution game by Tit-for-Tat advocates (TFT) with those made by other players (non-TFT). As Tit-for-Tat is equivalent to tax-gift substitution, we focus our attention on low-earners' tax votes and high earners' voluntary gifts. To be sure, the 15 PLN were low-earners throughout the experiment just as the 60 PLN were high-earners. In contrast, the 30 PLN were switching roles: they were low-earners when matched against the 60 PLN (in such a setting we analyse their tax vote) and high-earners when matched against the 15 PLN (in such setting we consider their voluntary gift).

In the 15 PLN group, we observe that TFT players were voting lower taxes than non-TFT, especially against the 60 PLN, where the difference amounts to whole 40 percentage points and is statistically significant. On the other end of the payoff scale, in the 60 PLN group, TFT players were offering substantially higher gifts than non-TFT and the differences are on the verge of statistical significance[19].

Of course, we should allow for reverse interpretation of causal path for it might also be the case that it had been actual experience of reciprocity that subsequently prompted subjects to invoke Tit-for-Tat strategy. If so, there should be some differences in partner's behaviour experienced by TFT and non-TFT players.

As evidenced by the data in Table 7, only in the 15 PLN group Tit-for-Tat advocates received significantly better treatment (from the 60 PLN). Other differences were far from significant, and some – like tax vote experienced by the 30 PLN from the 15 PLN – had the opposite direction, i.e. it was TFT players who had suffered larger tax burdens imposed by low-earners. Furthermore, it should be noted that TFT advocates were describing their partners' behaviour in no better terms than non-TFT. As a matter of fact, they were more likely to complain of other people's 'lack of rationality' and 'lack of reciprocation.'

[19] One should bear in mind that we have very small samples here and thus it takes a really strong effect to turn out statistically significant.

Table 7. High-earners' voluntary gift and low-earners' tax vote. Partner's decisions experienced by Tit-for-Tat advocates (TFT) and the rest (No TFT)

Decision		Voluntary gift			Tax vote		
Player		15 PLN		30 PLN		60 PLN	
Decision-maker: Partner		30 PLN	60 PLN	60 PLN	15 PLN	15 PLN	30 PLN
Player's type	TFT	0.47	3.41	2.48	67.1	55.2	58.6
	No TFT	0.84	0.72	1.91	45.6	64.9	49.3
Difference		-0.37	2.68	0.57	21.5	-9.7	9.3
p*		0.911	0.053	0.727	0.118	0.880	0.799

* Exact significance (2-tailed) in Mann-Whitney test. All sample sizes were between 10 and 12.

Misadventures of mute cooperation – a case study

A perfect illustration of how the inability to communicate could thwart cooperation is provided by games from one experimental session where 5 out of 6 subjects were aware of the mutually beneficial tax-gift substitution. First, let us consider their motivations and hypothetical communication proposals (subject are labelled by their experimental nicks).

Table 8. Own motivations and hypothetical communications by subjects in session #11 (males)

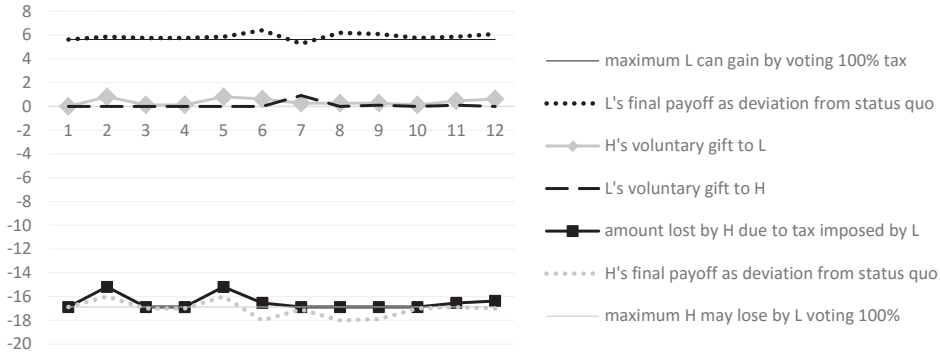
Kadr (60 PLN)	[I was basing my decisions on] *the tax rate imposed by the other person in a previous round and my willingness to initiate 'dialogue'… Let's profit together.*
Kana (60 PLN)	[I wanted to keep] *tax at its lowest because of the inherent cost. I was willing to pass 10 to 20% of my income.*
Kent (30 PLN)	[I would propose to have] *0% tax and share thus acquired 'profits' fifty-fifty.*
Klon (15 PLN)	*I would propose we both vote 0%… I would then expect my partner to transfer an amount making up for my losses… I would gain a good deal and the other person would lose less than if I voted 100%.*
Koch (15 PLN)	*Zero taxes, equalization through free giving.*

Now, it is obvious from the above statements that subjects, irrespective of their initial income position, were all interested in a sort of Tit-for-Tat arrangement with free giving and no taxes. However, this is very far from what actually happened in their games, as evidenced by the following series of game charts in Figures 5-8. Charts visualize both the players' decisions and the payoffs through 12 rounds of the redistribution game[20]. In the charts, topmost and bottommost thin horizontal lines indicate, respectively, the amount that a low-earner (L) could gain and a high-earner (H) lose due to low-earner's uncooperative voting for 100% redistributive

[20] Each one of the charted games was played at high 30% cost of tax redistribution and followed immediately after a game at 10% cost played with the same partner. However, the latter was unknown to the subjects and most of them surmised that they had been matched against a new partner.

tax. Solid grey and black lines with square markers indicate, respectively, H's free gifts to L, and amounts lost by H due to tax imposed by L. Dashed line traces L's free gifts to H, and dotted lines represent final payoffs in particular rounds expressed as deviations from initial incomes. If players voted 0% tax-rate and offered each other no gifts throughout the whole game all lines would coincide and run horizontally at zero-level (that is except for the 'maximum' lines).

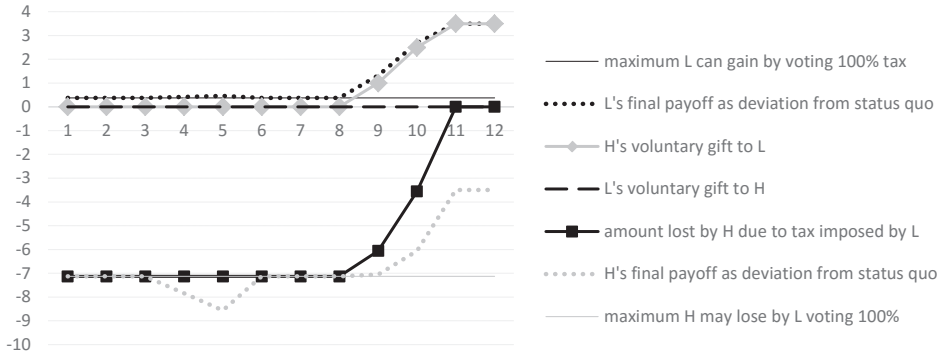
Figure 5. Game #108 (first partner): Koch (15 PLN) – Kadr (60 PLN)



In Figure 5 we witness faint and utterly failed attempts at cooperation by both players. High-earner, Kadr, makes varied little gifts (less than 1 PLN), not very attractive to Koch who in this game can assure himself as much as 5.63 PLN gain in tax benefits. This is exactly what he does most of the time, thus imposing on Kadr a concomitant loss of more than 16 PLN. Every now and then Koch is probing his partner's reaction with slightly diminished tax-vote (e.g. in rounds 2 and 5) but, in the last analysis, both players' frail and erratic cooperative gestures are futile and the game never strays too far from its dismal equilibrium track.

Simultaneously, another game, visualized in Figure 6, was played between Kent (30 PLN) and Koch (15 PLN). They both used strategies that amounted to 'suspicious Tit-for-Tat,' namely 'don't cooperate unless the other player initiates cooperation.' Accordingly, in the first eight rounds, Koch was receiving no free transfers from Kent and was himself voting 100% tax rate, which basically mimicked the uneventful history of their first game at 10% cost. Thus he was inflicting 7.13 PLN tax on his better-off partner only to gain a meagre 0.38 PLN tax benefit for himself. Finally, in round 9 Koch ventured to lower his defences a little bit and slightly diminished his tax vote. Kent responded in kind, and step by step they established a full-blown cooperation by round 11. They enjoyed the fruits of tax-gift substitution in the twelfth round as well, and then the game was over. A peculiarly bitter happy end it was as they must have reflected upon the uncooperative stalemate that ruled supreme for most of the game.

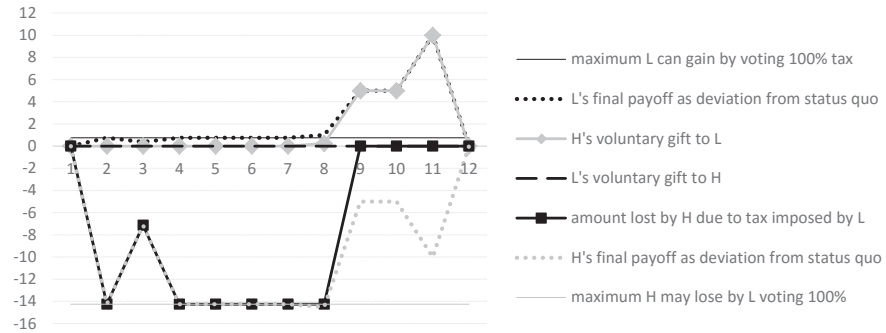
Figure 6. Game #12 (first partner): Klon (15 PLN) – Kent (30 PLN)



It is instructive to follow Kent and Klon to their games with newly matched partners. Would they capitalize on their recently experienced late ascent to cooperation? Well, they tried. In his next two games, Kent was a low-earner facing 60 PLN player nicknamed Kadr and indeed he did open both games with a 0% tax vote. However, Kadr failed to make an immediate cooperative response and the game deteriorated to no-gift maximum-tax equilibrium with only two occasional attempts at cooperation on part of Kent, again unreciprocated.

In Figure 7, we see the second game between Kent and Kadr.

Figure 7. Game #84 (second partner): Kent (30 PLN) – Kadr (60 PLN)



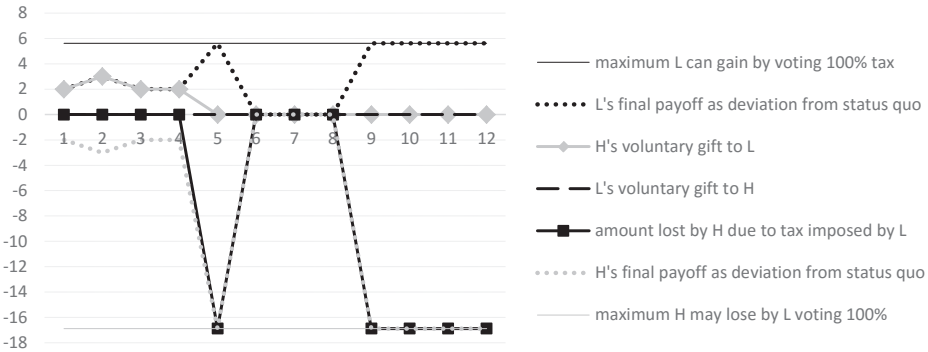
Again, Kent initiated the game with a 0% tax vote and yet again Kadr failed to respond immediately. In effect, rounds 2 and 4 through 8 were devoid of any trace of cooperation. Unexpectedly, in round 9 Kent dropped his tax-vote to zero and finally managed to coax Kadr into mutually beneficial cooperation which lasted for two more rounds. In the last round, Kadr defected on Kent, probably willing to recover half of the double-sized gift he offered him in round 11 as an incentive to vote no taxes in round 12.

Finally, let us follow Klon's adventures in his last two games played against Kana (60 PLN). In the first game, he consistently voted a 0% tax-rate and was receiving gifts of ca. 5 PLN. The problem with this arrangement was that Klon could

secure himself as much as 9.38 by voting 100% tax and so efficient cooperation was really an exercise in self-sacrifice for him. Worse, Kana tried to take advantage of his partner's good will as in the last four rounds he cut his donations to be less than half of what they used to be so far.

Figure 8 visualizes their second game. It started as if it was a continuation of the previous one: no taxes and insufficient gifts. In round 5, Klon's resolve to sustain disadvantageous cooperation finally broke down and he voted 100% tax. Kana responded by withholding the gift. In rounds 6 through 8, Klon tried to re-establish cooperation but to no avail – and since round 9 the game deteriorated to a grim uncooperative equilibrium routine.

Figure 8. Game #132 (second partner): Klon (15 PLN) – Kana (60 PLN)



These games provide an ample illustration of how difficult it may be to initiate and sustain cooperation in a Prisoner's Dilemma type of situation where participants have no effective means of communication. Even with all participants concerned about social losses due to entrenched defection and privately willing to establish mutually beneficial Tit-for-Tat agreement, long-lasting cooperation is a far shot.

Some remarks on ecological validity

To be succinct: 'Ecological validity is the extent to which research findings would generalize to settings typical of everyday life' (Wegener, Blankenship 2007). Certainly, it would be rather silly to claim that experimental framing of a two-person redistribution game made it a straightforward model of real-life redistribution processes in a democratic society. That granted, we may still expect certain factors of the situation to work on a similar basis. Half a century ago, Morris Zelditch (1969) asked 'Can you really study an army in the laboratory?'. His conclusion boiled down to an observation that 'while you cannot take an army into a laboratory, you certainly can study important theoretical features of armies' (Webster, Sell 2014, p. 20). By the same token, in the present experiment attention was focused on crucial circumstances under which redistributive policies are shaped and implemented. One such aspect is pre-tax income inequality, another is the scope of wastefulness inherent in the system.

Maximum efforts were taken to provide clues for experimental subjects that what they participated in was not a parlour game but a session of serious decision-making with financial consequences for all involved. Thus both in recruitment ads and introductory pre-experimental speech subjects were informed that

> The aim of the experiment . . . is to better understand processes related to wage setting, tax voting and decision-making concerning monetary transfers to other persons.

More concretely, their role in the experiment was explained in the following terms:

> Today you will be asked to fill out a questionnaire concerning various issues that refer to earning money, taxes, income inequalities, etc. Then, we have a certain set of tasks prepared for you, a job you will be paid for . . . . Next time we meet in the computer laboratory where you will make decisions concerning money you will have earned today.

The tasks that were to be performed on spot were time-consuming and paid by piece rate. Therefore all subjects knew well that all money in the experiment was earned and that earnings were in direct proportion to the amount of job done. They were even asked, 'behind the veil of ignorance,' to make fairness judgments about experimental payoff scheme. Hence we know that to a large extent they believed that initial payoffs of 15, 30, and 60 PLN had been fairly earned. Now did this no-nonsense framing of the experiment become a part of how subjects perceived the whole situation? Did their experimental choices have anything to do with their political beliefs and moral convictions concerning real-life social processes? They were not asked about it in the post-experimental survey for fear that such questions, when publicized, could influence behaviour of other persons in the future sessions. But it is worthwhile to conclude this article with a number of comments that subjects spontaneously made to that effect.

Thus Biel (30 PLN) complained of 'ingratitude' of her partners:

> *They didn't want to change their tax proposals even though I encouraged them with 'symbolic' transfers of certain sums of money. Such behaviours may be observed in everyday life – people would like some, e.g. the rich, to pay taxes and make donations – 'give them an inch, and they'll take a mile.'*

Haft (60 PLN) offered one of the most elaborate accounts of player's own motivations in the game:

> *I was guided by a general economic conviction that taxes should be minimized. . . . Besides, I consider the very idea of equalizing taxation to be socially, economically, and ethically misguided. I was transferring money out of a simple need to share (with persons 3 and 4), for even though the disparity was rooted in disparate amounts of job done, I was willing to compensate it somehow. The fact that the amount of job to be done was determined by chance (rather than by choice) certainly had some relevance here. My second motivation was a desire to persuade my partner (persons 1 and 2) to lower the taxes. I didn't want money to be dissolved in procedural costs, and at the same time I wanted somehow to show those people that a higher tax imposed on the better-off does not lead to enriching*

the poor, and that general lowering of the rates may bring about much better result. Thus I simply cared for my own interest, but also tried to teach others a deal.

Horn (15 PLN) took the survey opportunity to offer a meticulous self-critique:

The study was certainly a good idea and I think the findings will be descriptive of the whole society, unfortunately with certain exceptions. As I mentioned before, I had 15 PLN so I was always the poorer [of the two] and each time I felt the urge to choose the highest equalizing tax in order to grab some part of my partner's income… such behaviour is quite typical of Polish paupers who prefer, out of envy, to take from the rich only because the latter are better-off. So in the first two rounds my behaviour could exemplify such an attitude. But as I mentioned before, in subsequent rounds I decided to vote lower taxes so as not to take money from persons who had to work for it after all.

Chem (30 PLN), on the other hand, felt he should justify his choices in high-earner's position:

Allow me to write a few words. If I am correct, I think the experiment was testing whether people declaring certain attitudes are faithful to them in reality. If so, I think it doesn't fully reflect the phenomenon, i.e. declaration vs. behaviour. For example, in the survey I declared rather left-wing views. I wrote that income disparities should be smaller, that poor people should be assisted, etc. During the session I didn't give money to a person I was matched with. But (in my opinion) the level of the whole society, the state, etc. is different from managing money at micro-level, where my main concern was to get a few zlotys for a ticket home, or for kilograms of photocopies that I need to make for the quickly oncoming examination period. I knew that by giving nothing at all to the other person I'm not depriving him of life's necessities, and that either way he will leave the experiment with a certain sum of money.

For Etna (60 PLN), taking part in the experiment was *'a stimulus to reflect upon* [her] *attitude to financial issues'*. Finally, Cedr (30 PLN) mentioned that after the experiment subjects went on discussing their choices in the redistribution game:

Some of them played similarly to me, which I liked, others did not. It seems that some of them didn't feel strategy but after we had left the laboratory it turned out that they had their strategies and could argue them all right ☺

Indeed, it is the case that certain behaviours in the experimental game were seemingly irrational until explained by the subjects. Why should the high-earners vote for positive taxes if they could transfer the same amount freely without incurring the additional cost? Or why should the low-earners vote high taxes and then make gifts to their better-off partners? Well, from the subjects' post-experimental statements we learn that these were not necessarily symptoms of insanity. They could have been premeditated, even if desperate, attempts at signalling friendly attitude and spurring cooperation. So in the end it seems that in the social world there is a method to every madness.

Conclusions

Data from pre- and post-experimental surveys may provide us with clues about factors affecting the subjects' behaviour in the experiment, as well as help us better understand and interpret motivations behind their actions. Two such factors that turned out to have some influence on the subjects' perception of the optimal behaviour in the redistribution game were beliefs about state-guaranteed minimum income and year of studies. Older students were much more likely to formulate a hypothetical negotiated solution for the game in terms of mutual use of Tit-for-Tat strategies characteristic of repeated Prisoner's Dilemma. This suggests that students' strategic sophistication is linked to age. Since many game-theoretic experiments have students for experimental subjects, it would therefore be advisable to perform such experiments on groups within narrow age-brackets, or at least to gather information on subjects' age for the purpose of statistical control.

Allowing for communication between players could immensely affect the likelihood of establishing a full-blown cooperation. With all channels of communication blocked, even cooperation-conscious players find it hard to overcome mutual defection which constitutes unique equilibrium in a repeated Prisoner's Dilemma when the total number of rounds is common knowledge among players. Problems caused by the inability to negotiate common way of conduct are further exacerbated by the fact that people will tend to read hostile, reckless, or irrational actions into other people's decisions that affect them negatively.

References

Axelrod R. (1984). *The Evolution of Cooperation*. Basic Books. New York.

Camerer C.F. (2003). *Behavioral Game Theory. Experiments in Strategic Interaction*. New York: Russell Sage Foundation.

Czarnik S. (2009). *Proportionality and Two Kinds of Fairness*. In: J. Bacher, J. Górniak, M. Niezgoda (eds.), *Selected Research Papers in Education, Labour Market and Criminology*, vol. 1. Linz: Trauner Verlag. p. 103–109.

Czarnik S. (2007). *Sprawiedliwość i równość w podziale dóbr wypracowanych. Formalny model redystrybucji w warunkach przymusu i dobrowolności* (in Polish). Ph.D. thesis. Uniwersytet Jagielloński, Kraków.

Czarnik S. (2006). *Justice in the Shadow of Self-Interest. An Experiment on Redistributive Behavior.* Acta Physica Polonica B, 37, p. 2967–2977.

Friedman M. (2002). *Capitalism and Freedom. 40th Anniversary Edition*. Chicago and London: The University of Chicago Press.

Hazlitt H. (1969). *Man vs. the Welfare State*. New Rochelle, N.Y.: Arlington House.

Kahneman D., Tversky A. (1979). *Prospect Theory: An Analysis of Decisions under Risk.* Econometrica, 47, p, 263–291.

Kreps D.M,, Milgrom P., Roberts J., Wilson R. (1982). *Rational Cooperation in the Finitely-Repeated Prisoners' Dilemma*. Institute for Mathematical Studies in the Social Sciences, Stanford University.

Webster M., Jr., Sell J. (2014). *Why Do Experiments?* In: M. Webster Jr., J. Sell (eds.), *Laboratory Experiments in the Social Sciences*, 2nd ed. San Diego, CA: Academic Press, p. 5–21.

Wegener D.T., Blankenship K.L. (2007). *Ecological Validity*. In: R.F. Baumeister, K.D. Vohs (eds.), *Encyclopedia of Social Psychology*. Thousand Oaks, CA: Sage Publications, p. 275–276.

Zelditch M. (1969). *Can You Really Study an Army in the Laboratory?* In: A. Etzioni (ed.), *Complex Organizations*, 2nd ed. New York: Holt, Rinehart, Winston, p. 528–539.

Czytanie w myślach uczestników eksperymentu. Wnioski z kwestionariuszy wypełnianych przed i po eksperymencie z grą redystrybucyjną

Wypowiedzi badanych z kwestionariuszy przed- i poeksperymentalnych wykorzystane zostały do zinterpretowania ich zachowań w grze redystrybucyjnej o strukturze asymetrycznego Dylematu Więźnia bez komunikacji między graczami. Badani w przeważającej mierze negatywnie oceniali intencje swoich partnerów, podczas gdy swoje własne decyzje określali jako motywowane racjonalnym interesem własnym, wzajemnością i troską o efektywność. Przy braku komunikacji atrybucja złych intencji może być więc jedną z głównych przeszkód na drodze do ustanowienia trwałej kooperacji, nawet wówczas, gdy obydwaj gracze w pełni zdają sobie sprawę z tego, jakie korzyści niesie ze sobą współpraca i jakie działania są konieczne do jej osiągnięcia. Tym niemniej zebrane dane świadczą o tym, że gracze świadomi Pareto-optymalizującego potencjału tkwiącego w obustronnym przyjęciu strategii wet za wet na ogół podejmują w toku gry bardziej przyjazne decyzje. Wreszcie w świetle poeksperymentalnych wypowiedzi badanych, oceniamy trafność ekologiczną eksperymentu redystrybucyjnego.

Słowa kluczowe: behawioralna teoria gier, Dylemat Więźnia, redystrybucja podatkowa, kooperacja, interpretacja intencji

FOLIA 202

Annales Universitatis Paedagogicae Cracoviensis

Studia Sociologica VIII (2016), vol. 1, p. 166–178
ISSN 2081–6642

Marcel Kotkowski
Jagiellonian University, Poland

Psychophysiological Techniques for Measuring Emotion in Social Science

Abstract

This article briefly presents 6 techniques of measuring emotion: functional magnetic resonance imaging (fMRI), electroencephalography (EEG), electromyography (EMG), galvanic skin response (GSR), Facial Action Coding System (FACS) and infrared thermography (IRT). A note on each technique points out the dimension of emotion (valence or arousal) that is measured with a given technique, and informs on its previous use in sociology, as well as its major advantages and disadvantages. Limitations common to all techniques are discussed in the concluding section.

Key words: sociology of emotions, emotion measurement, research on emotion, measuring techniques, physiological indicators

Introduction

The history of research on emotions in social science is rather short. 'Fathers' of sociology, August Comte, Emile Durkheim, Georg Simmel or Max Weber, granted, to be sure, to the concept of *emotion* a relatively high position in their theoretical systems (Shilling 2012, p. 188–205), but psychologists have always had much greater experience in studying this field than sociologists. For the latter emotions have been for many years a subject of secondary interest, too often forgotten or underestimated. It is only over the last forty years that theoretical systems that stressed the importance of emotions have appeared again. Unfortunately, the number of new *sociological* studies dealing with emotions is gradually declining (Turner, Stets 2006, p. 25), with the majority of these contributions being focused on theoretical analysis rather than the practice of empirical research.

Historically, while sociology has been preoccupied mainly with theorising on emotions, leaving aside the question of how to investigate this important phenomenon (Lively 2014, p. 1), psychology has taken over some of the sociologists' duties and got involved in the study of emotions also in the context of *social interaction*. Psychologists have invented tools and research techniques allowing the *measurement* of the affective states of a human being. In addition, the dynamically developing *neuroscience* aided in the research on emotions, making it more effective and

reliable. The so-called 'neuroscience turn' occurred in economy, philosophy, anthropology, law, and in the aforementioned psychology. However, it did not yet happen in sociology (Scheve 2003, p. 3). The reason for this lies in the sociologists' concern over the impact of reductionism and biological determinism. The debate on *sociobiology*, which took place in the 1970s, did not lead to the emergence of a new paradigm in social science. An explanation of why things have gone so is beyond the scope of this article. Nonetheless, the consequences of a revival of naturalism have been far-reaching. Sociologists can no longer ignore the influence of biological factors on emotions. They have to acknowledge that their claim that emotions have social origin is losing its prominent status within sociological theorising. Indeed, social researchers are becoming more and more aware that further research on both theoretical and empirical aspect of emotions cannot do without *physiological* factors and indicators to be taken into account along with the social and cultural variable (Stets 2010, p. 266).

The implementation of techniques widely used in medicine, criminology and cognitive science provides social science with a huge opportunity to gain more extensive and deeper knowledge of emotions and to empirically verify certain hypotheses, which, due to the lack of adequate tools, have so far been shrouded in doubt. Modern science has to consolidate and use its resources effectively. For this reason, I found it useful to present to fellow sociologists certain research techniques applied in other sciences, the techniques which have already been used with success in sociological studies of human emotionality. In this paper, a description of each technique employed in the area of emotion research is presented with some examples of its use in sociology, followed by an account of its major drawbacks, advantages, and dangers involved. The focus is solely on *measurement.* It would require writing another more extensive article to provide an overview of general models, theories, methods and results of empirical studies in this area.

Dimensions of emotions and measurement techniques

Before proceeding to an overview of techniques, we recall the distinction between two dimensions, *valence* and *arousal*, that the researchers studying emotions have identified (Lottridge, Chignell, Jovicic 2011, p. 201). Valence, the dimension which is more difficult to measure, pertains to the quality of experienced affect that is described in terms of a position on a bipolar continuum extending from *positive* emotions on one end, and *negative* on the other end. Research on emotions usually makes use of the so-called *Big Six*, a catalogue of six emotions considered basic: happiness, sadness, anger, fear, disgust, and the sixth one as to which the researchers' views differ (Scherer 2004, p. 677).

The arousal dimension describes the *strength* of an experienced affect varying from very strong to unnoticeable. Detecting the exact strength of the emotional state of a test participant is possible but far more difficult than ascertaining if it is positive or negative. Differentiating between the dimensions of measurement is vital because most research techniques provide information about only one of them.

In the sections of this paper that follow, six techniques used in emotion research are described one by one. Many other relevant measurement techniques tools have

been skipped, however; these include: positron emission tomography (PET), magnetoencephalography (MEG), near infrared spectroscopy (NIRS), transcranial magnetic stimulation (TCMS), or reading emotion from the tone of voice or pulse. The reason for the omission was the priority given by the author to the techniques which have already been used in social studies, as well as to those which are feasible, or easy to implement, from the technical, financial, and teleological point of view.

Functional magnetic resonance imaging

Functional magnetic resonance imaging (fMRI) apparatus is a highly-specialized device, which proved suitable for measuring valence of experienced emotions. Its functioning relies on the observation and measurement of the BOLD effect (blood-oxygen-level dependent contrast), which consists in the increase of oxygen saturation in certain parts of the brain in response to stimulation. The increased demand for oxygen (and glucose) that arises with the increased activity of neurons in a given area, is being satisfied by an inflow (visible in the magnetic field) of blood to that area.

The use of fMRI, among other things, allowed for the identification of *mirror neurons*, or the areas of the brain which are being activated when a person performs a goal-oriented activity or observes such activities being performed by others. Intensive studies revealed the importance of this area for primary socialization, understanding other people's emotions, or empathy; a connection between dysfunctions of this area and autism was also found (Iacoboni, Dapretto 2006, p. 942; Iacoboni 2009).

Immordino-Yang and Damasio (2007) conducted a study on patients with brain damage. Using fMRI, they proved that emotions play a substantial role in the decision-making process and learning. Patients with lesion of a particular part of the frontal lobe, while maintaining the ability to think logically, could not predict the consequences of their decisions and appeared unable to learn on their mistakes. Another often quoted research in which social science interweaves with neuroimaging, is the one devoted to pinpointing the areas of the brain, responsible for focusing one's attention during interactions. The results of that study contributed to a better understanding of the structure of human interaction process (Redcay et al. 2010).

Without doubt, fMRI is a powerful diagnostic tool. Operated by competent personnel, it can provide invaluable data on human cognitive processes and the anatomical foundations of social existence. It is characterized by non-invasiveness and high spatial resolution, which guarantees a quality image with a high level of detail.

Major drawbacks of using fMRI in social science are the high running costs (about 1000–1500 PLN per test) and the need for close cooperation with the personnel operating the apparatus. Another limitation, which significantly decreases the usability of the device, is the fact that the person undergoing a test needs to remain still in a tube-shaped structure. This obviously precludes the observation of interaction in natural conditions. In some cases, the disqualifying factor is the

relatively low time resolution – in order to see the changes on the screen, one has to wait for about 1–2 seconds. In this case, grasping the dynamic, momentary or ephemeral affective states may be impossible.

For these reasons, the technique is unlikely to be used extensively in sociological investigations. Nevertheless, I decided to include this technique in my account because it has become an invaluable tool for psychological studies still having a great impact on the social sciences.

Electroencephalography

Electroencephalography (EEG) can be employed, depending on the research assumptions, to determine the valence and/or arousal of experienced emotions.[1] It is an imaging technique for the electrical activity generated by brain structures. A set of electrodes is attached to the head (scalp) of the test participant. When neurons are activated, an electrical current flows through them and the activity is recorded on the device. The weak signals are amplified and written into the computer memory (Teplan 2002, p. 1).

EEG has been used to understand and anticipate antisocial behaviours. Many papers suggest that aggressive individuals, likely to commit offences, are characterized by abnormal readings. Raine, Venables, and Williams (1990) concluded that antisocial behaviour observed at the age of 24 could have been predicted on the basis of lower frequencies found in the readings obtained from the measurement of 15 year-olds. The same line of argument was put forward by Barratt et al. (1997). They claim that that lower values of the brain response to stimulation distinguish the individuals with antisocial inclinations from the rest of society.[2] Of course these findings do not imply that human behaviour is determined solely by biology but they encourage social scientists to explore new possibilities of analysing deviance.

By far the biggest advantage of using EEG is its large time resolution, measured in milliseconds (Teplan, 2002, p. 4). It allows for even momentary emotions to be captured. Unfortunately, in the context of sociological research on emotion, EEG has many disadvantages. Undoubtedly, the biggest one is the nuisance of the measurement process. The test participant has to sit still, without blinking and keeping facial expression fixed because a slightest muscle contraction can influence the results (Bahari, Janghorbani 2013, p. 228). The necessity of background noise control and the cumbersome equipment rule out the possibility of performing tests outside the laboratory. Other difficulties are connected with low spatial resolution, which means

[1] We distinguish two types of analysis: EEG, which is a test to detect problems in the electrical activity of the brain, and ERP (Event-Related Potentials), which is a stereotyped electrophysiological response to a stimulus, for instance, situation, object or thoughts.

[2] ERP (event-related potentials) studies conducted with the use EEG under the so-called oddball paradigm have shown that the participants whose P300 amplitude (caused by a stimulus engaging attention and eliciting an orienting response) was lower were characterized by higher susceptibility to deviation. Since the P300 wave is generated mainly in the parietal lobe, the weakening of its amplitude can be associated with aggressive behaviour, indecisiveness and non-compliance with social rules.

that only relatively large areas of active neurons leave traces in the recorded data, the recording being limited to cerebral cortex with no representation of subcortical structures (Teplan 2002, p. 7). Furthermore, the recognisability of emotions with algorithms is relatively low. Best results so far, with 83.33% recognition of 6 basic emotions (according to Ekman) were achieved by Panagiotis and Leontios (2010).

Electromyography

Electromyography (EMG) is a diagnostic technique based on measuring electrical activity that appears as a result of muscle contraction. By placing the electrodes at the right points on the participant's face, it is possible to assess the valence of experienced emotions. Tone of the muscle responsible for frowning (Latin *musculus corrugator supercilii*) increases inversely with the decrease on the valence scale, which means that the test participant is experiencing negative emotions. In contrast, positive affective state is indicated by the activity of the *zygomaticus major muscle*. The tone of this muscle increases proportionally to the increase on the valence scale (Brown, Schwartz 1980; van den Broek et al. 2006, p. 4). A later study (Larsen, Norris, Cacioppo 2003, p. 777) shows that experiencing positive emotions corresponds not only to the increase in the activity of the zygomaticus major muscle but also to the decrease of the tone of the muscle responsible for frowning. The relation also works the other way – negative emotions are correlated with the decrease in the readings of zygomaticus major muscle.

Until now EMG has been used mainly in experiments and quasi-experiments. It is useful in the assessment of emotional response to a stimulus and in investigating the influence of independent variables on dependent variables. In this context, the technique was used in the study of the emotional response to text, image and sound (Larsen, Norris, Cacioppo 2003, p. 783), and video (van den Broek et al., 2006). Equally positive results were achieved during the analysis of affective states connected with human interaction mediated for instance by Internet communicators and with technology per se (Calvo, Member, Mello 2010; Mandryk, Atkins 2007).

A signal, recorded in the form of a frequency (Hz), is sent to an amplifier via cables and later to the computer (Larsen, Norris, Cacioppo 2003, p. 778). The quantitative data obtained clearly indicates any alterations of the affective state of the test participants over time. With the use of a correct database, EMG allows not only to assess whether the emotions experienced are positive or negative, but enable us to recognize exactly the type of emotion (Petrantonakis 2010, p. 190). Similar to other methods mentioned, this one is also characterized by automated measurement and result interpretation.

Like any other technique, the one in question has its drawbacks too. When you use EMG, you must remember that the measurements of positive and negative emotions are conducted independently. In general, the correlation between muscle tone and experienced emotions is stronger in the muscle responsible for frowning than in zygomaticus major muscle (Larsen, Norris, Cacioppo 2003). In other words, it is easier to detect negative than positive emotions. Substantial difficulties stem from the fact that EMG can only be used in laboratory conditions, which eliminates the

possibility of observing emotions in their natural environment. The necessity of connecting the electrodes to the participants' faces also needs to be taken into consideration, as it may significantly reduce comfort and produce negative emotions, leading to biased research results.

Galvanic Skin Response

Galvanometer is a device for measuring human skin conductivity. Conductivity increases abruptly during mental activity, such as solving mathematical problems, and stays on a relatively low level during rest. Sudden emotional arousal is one of the factors triggering galvanic reactions (Picard, Scheirer 2001, p. 1). The measuring of conductivity and resistance relies on the Ohm law, according to which the intensity of an electrical current is proportional to the voltage between two ends of a conductor, in this case two electrodes (Dawson, Schell, Filion 2007, p. 204). D'Mello, Dowell, and Graesser (2013) indicate that studying the galvanic response of human skin (GSR) is the fastest and most reliable way of measuring the intensity of human emotions. Electrodes can be placed at various points of the body, but the best results are achieved by measuring the electric potential of feet and hands. For obvious reasons, the latter possibility is used more frequently. Wires are usually connected to two fingers of the non-dominant hand (Boucsein et al., 2012).

An attempt to create a responsive user interface is an example of using GSR as a technique for collecting data on emotional arousal of the participants. The data was later used to stimulate them by triggering the correct algorithms (Villon, Lisetti 2006). Other researchers (Wang, Prendinger, Igarashi 2004) achieved satisfactory results with GSR by measuring the strength of emotional arousal during the interaction between two individuals, mediated by an Internet communicator.

GSR is a cheap method. The purchase of an affordable apparatus allows for an unlimited number of tests and the equipment requires only one person to operate. Portable GSR devices have been in development for some time. One of them is the *Galvactivator* – an ergonomic measuring device resembling a glove (Picard, Scheirer 2001). It enables doing research outside the laboratory, so studying the level of emotional arousal in natural conditions becomes possible. It is equally important that many individuals can be tested simultaneously if an adequate number of devices is available.

Some authors (Cacioppo, Tassinary 1990, p. 17; Ward, Marsden 2003, p. 210) point out that the results of a test for an individual may depend to a significant degree on the number and default level of activity of the individual's sweat glands. These characteristics vary across individuals, which requires that the measuring devices are calibrated before each test in order to set the so-called zero level for each subject. Some researchers report that better results can be achieved by measuring conductivity rather than electric resistance. Age and sex of tested individuals also matter – older participants achieve lower results; women's reaction to unpleasant stimuli is stronger, whereas men are more sensitive to erotic arousal. External variables, which are likely to affect measurement results, include body temperature, and temperature and humidity in the room in which the test is carried out (Boucsein et al. 2012, p. 1030). Administering medications in the course of a GSR test brings

into the measurement process another heterogeneous set of factors which may distort results. A potential weakness of this technique, however not always crucial, is a relatively low time resolution of 1–3 seconds (Dawson et al. 2007, p. 211).

Facial Action Coding System

Facial Action Coding System (FACS) is a technique of reading emotions from the face. In the 1970s, Paul Ekman[3] established a number of Action Units (AU) conceived of as a kind of measurement units. Each of them is related to a change in the tone of face muscles, which are attributed to particular emotions, e.g., raising the corner of the mouth and eyebrows is interpreted as smile, a sign of happiness. AU are located mainly on eyebrows, nose and mouth.

Many observation patterns, similar to FACS have been established. One of the most popular is the Facial Animation Parameters (FAPs). A breakthrough in the development of FACS occurred in the last decade of the 20th century, when the first algorithms automating the decoding process were created (Bettadapura 2012, p. 1–7). The test procedure is simple – a video camera, aided by appropriate software, has to be pointed (assuming the appropriate angle and distance) at the face of a person whose emotions currently experienced are going to be recognized and named by the computer system (Sánchez et al. 2011, p. 1272). Depending on the software used and the type of emotion to be detected, the accuracy of the technique oscillates between 70% and 100% (Bettadapura 2012, p. 10; Karpouzis et al. 2007, p. 7).

At present, recognizing emotions from the face has become very popular in the commercial world (in this way companies and corporations study emotional reactions to new products or solutions) and in science. Because of its high usability, the method is still being developed. Although it has many varieties now, all of them are based on the work of Paul Ekman. The educational value of his legacy is priceless, as he was able to describe (see Ekman et al. 1987) universal emotions for particular cultures. Ekman and Cordaro (2011) proposed a list of *basic emotions* and investigated their characteristics. Ekman's book (1985, revised edition 2009) is a kind of manual for all who want to learn about the situations in which the other speaker is not telling the truth. Studies conducted by other researchers using FACS (Tsai, Chentsova-Dutton 2003) proved that Americans of Scandinavian and Irish descent express emotions differently, e.g., the first group was characterized by lower expressiveness in reference to happiness and love. The differences detected in those studies support the results of previous theoretical work. What is more, these empirical findings show that cultural differences are more visible for positive than negative emotions (Matsumoto et al. 1998).

Reading emotions with the use of a video camera equipped with necessary software is a cheap technique that enables carrying out an unlimited number of tests at any time and place. Once a video camera has been set up correctly, no additional actions are requited from the researcher, which makes this technique simple and

[3] Paul Ekman, considered the founder of FACS, continued the research initiated by Carl-Herman Hjortsjö.

effective. The biggest advantage of FACS, however, is full automation, making the presence of a researcher unnecessary for data collection. The test participant can be placed in another room or building, or even continent. There are no contraindications for the test not to be carried out remotely and in a fully automated manner, like the tests conducted for commercial purposes. Moreover, the right choice of equipment and its proper placement make it possible to study entire groups in various environments.

The choice of an appropriate database on the basis of which the software interprets the movement of each point as a sign of a particular emotion is important from both theoretical and methodological point of view (Karpouzis et al. 2007, p. 2). The problem lies in the number of different databases available and their incomparability. At present, the most popular are MMI Facial Expression database and the Cohn-Kanade database (Pantic et al. 2005; Sánchez et al., 2011, p. 1276).

Despite relatively high effectiveness of the technique, not every emotion can be recognized easily. Happiness, surprise and disgust are easier for the software to read from the face because these emotions find expression largely in evident AU of the mouth. Detecting anger is more difficult, which emotion for that reason is often confused with other affects. Covering the face, especially lips (e.g. with a scarf, facial hair, makeup), considerably decreases the chance of a correct reading, even by 50% (Bettadapura 2012, p. 10–22).

Infrared thermography

Infrared thermography (IRT) is a technique designed for detecting and recording temperature displacement on the surface of an object through the measurement of the infrared radiation emitted by the object. In emotion research, the test object is the human face. A video camera, pointed at the test participant's face, creates a colourful image, in which brighter colours indicate high temperature and darker shades represent lower temperature. It is worth mentioning that an infrared video camera reads the temperature of each pixel individually (Clay-Warner, Robinson 2014, p. 3). In essence, the theoretical basis of this technique stems from the aforementioned FACS and EMG (Robinson et al., 2012, p. 14). Like in the cases previously described, micro-expression, connected with experiencing positive or negative emotions, plays here a key role too. Latest research (Jarlier et al. 2011) indicated that interpreting thermographic data in terms of AU brings promising results. Although the technique has been used mainly to recognize emotions, it also performs remarkably well in establishing their level of arousal (Clay-Warner, Robinson, 2014, p. 5).

The research by Robinson et al. (2012) was the first one in which the effectiveness of infrared thermography in recognizing emotion was empirically proven. The results indicate that the temperature of the human face is different when experiencing positive and negative affects. The biggest differences were visible in the temperature of the cheeks and forehead. In the same year, another team of researchers (Wesley et al. 2012) compared the effectiveness of two techniques for detecting emotions with algorithms based on AU – using a thermal imaging video camera and a regular one – in different environmental conditions. As expected by

the authors, tampering with the temperature negatively influenced the results of the infrared test. At the same time, switching off the light in the test room decreased the effectiveness of regular video cameras, while the effectiveness of IRT remained at a steady level. As indicated by the research of the Pavlidis, Eberhardt, and Levine (2002), IRT can be used not only for recognizing emotion by face analysis. The researchers found that the temperature measured with this instrument around the eyes of a person, in 83 percent of cased provided a correct prediction of whether the person was telling the truth. After lying, this area of the face got significantly hotter, indicating guilt.

The advantages of IRT are similar to those of FACS, namely: low costs, automation and the possibility of group remote testing. Robinson et al. (2012, p. 15, 36) point to the possibility of conducting interaction tests in a dynamically-changing environment under any light conditions, even in complete darkness. However, the authors of most research incorporating IRT claim that the best results can be achieved by employing another measuring technique (techniques).

As mentioned above, the biggest drawback of the IRT technique is the necessity to maintain a relatively steady temperature in the test room (the same applies to humidity). The result of the test can be also affected by anything that covers the face, even partially (hair on the face or covering it, any kind of glasses, scarves or hats). It is also highly unadvisable for the test participant to move excessively and change the distance from the video camera (Clay-Warner, Robinson, 2014, p. 5), although these disadvantages depend on the software used, which is constantly being improved.

Conclusions

All the techniques presented above using different mechanisms and ways of data collection share a few significant limitations. All of them arise from methodological assumptions and the imperfection of measuring devices. First of all, each technique is suitable for studying a limited number of emotions only. The researchers generally agree that the number of possible emotional states corresponds to the number of possible models of assessment, which makes it virtually endless (Scherer 2000, p. 149). For pragmatic reasons, even those who criticise the conception of basic emotions have to resort to it, for otherwise any analysis would be inconclusive or incomplete.

Another common setback is the inability to study mixed emotions. A particular emotional state experienced by an individual can be a combination of different or even contradictory states (Davidson 2003, p. 1; Panksepp, Watt 2011, p. 3). Very often we cannot discriminate between two emotions, for instance, uncertainty and frustration, nor can we devise an indicator for the case, which happens fairly often, where an individual experiences anger and happiness at the same time.

The third limitation (not applicable to the last two techniques described) goes together with the biggest advantage of the measurement based on physiological indicators – relying on a reading which is independent from the person's will because of being regulated by the autonomous nervous system (Westerink, Broek, Schut 2008, p. 151). It may be risky to rely solely on these techniques, without asking

the participants themselves to describe their emotional states. For that reason, the use of psychophysiological techniques is often complemented by taking self-reports from the subjects.

The techniques described in this paper, in spite of their weaknesses and limitations, have been used for many years and they enjoyed growing credit due to being constantly developed and improved. Since their introducing to psychology they have been widely recognized in the world of science and gained strong and extensive methodological underpinning. Sociology of emotions – if it aims at becoming a truly empirical science – may only benefit from vast research experience and knowledge accrued over time in the disciplines that deal with the biophysical aspect of human existence. For social scientists, who are interested in studying social interactions and relations, the techniques such as FACS or IRT can be useful, first of all, in so far as they enable testing entire groups during the course of interaction in laboratory and naturals settings.

References

Bahari F., Janghorbani A. (2013). *EEG–Based Emotion Recognition Using Recurrence Plot Analysis and K Nearest Neighbour Classifier.* Conference paper: Biomedical Engineering, p. 228–233.

Barratt E.S., Stanford M.S., Kent T.A., Felthous A. (1997). *Neuropsychological and Cognitive Psychophysiological Substrates of Impulsive Aggression.* Biological Psychiatry, 41, p. 1045–1061.

Bettadapura V. (2012). *Face Expression Recognition and Analysis: The State of the Art.* [online] http://arxiv.org/ftp/arxiv/papers/1203/1203.6722.pdf, access: April 20, 2015.

Boucsein W., Fowles D.C., Grimnes S., Ben-Shakhar G., Roth W.T., Dawson M.E., Filion D.L. (2012). *Publication Recommendations for Electrodermal Measurements.* Psychophysiology 49, p. 1017–1034.

Brown S.L., Schwartz G.E. (1980). *Relationships Between Facial Electromyography and Subjective Experience During Affective Imagery.* Biological Psychology, 11, p. 49–62.

Cacioppo J.T., Tassinary L.G. (1990). *Inferring Psychological Significance from Physiological Signals.* American Psychologist, 45, p. 16–28.

Calvo R.A, Member S., Mello S.D. (2010). *Affect Detection: an Interdisciplinary Review of Models, Methods, and Their Application to Learning Environments.* IEEE Transactions on Affective Computing, 1, p. 1–23.

Clay-Warner J., Robinson D.T. (2015). *Infrared Thermography as a Measure of Emotion Response*. Emotion Review, 7, p. 157–162.

Davidson R.J. (2003). *Seven Sins in the Study of Emotion: Correctives from Affective Neuroscience*. Brain and Cognition, 52, p. 129–132.

Dawson M.I., Schell A.M., Filion D.L. (2007). *The Electrodermal System*. In: J.T. Cacioppo, L.G. Tassinary, G.G. Berntson (eds.). *Handbook of Psychophysiology.* New York: Cambridge University Press, p. 159–181.

D'Mello S.K., Dowell N., Graesser A. (2013). *Unimodal and multimodal human perception of naturalistic non-basic affective states during human–computer interactions*, Affective Computing, IEEE Transactions, 4(4), p. 452–465.

Ekman P. (2009). *Telling Lies: Clues to Deceit in the Marketplace, Politics, and Marriage*. New York-London: W.W. Norton & Company.

Ekman P., Cordaro D. (2011). *What is Meant by Calling Emotions Basic.* Emotion Review, 3, p. 364–370.

Ekman P., Friesen W.V., O'Sullivan M., Chan A., Diacoyanni-Tarlatzis I., Heider K., Ricci-Bitti S.E. (1987). *Universals and Cultural Differences in the Judgments of Facial Expressions of Emotion.* Journal of Personality and Social Psychology, 53, p. 712–717.

Iacoboni M. (2009). *Imitation, Empathy, and Mirror Neurons.* Annual Review of Psychology 60, p. 653–670.

Iacoboni M., Dapretto M. (2006). *The Mirror Neuron System and the Consequences of Its Dysfunction.* Nature Reviews. Neuroscience 7, p. 942–951.

Immordino-Yang M.H., Damasio A. (2007). *We Feel, Therefore We Learn: The Relevance of Affective and Social Neuroscience to Education.* Mind, Brain, and Education, 1, p. 3–10.

Jarlier S., Grandjean D., Delplanque S., N'Diaye K., Cayeux I., Velazco M.I., Scherer K.R. (2011). *Thermal Analysis of Facial Muscles Contractions.* IEEE Transactions on Affective Computing, 2, p. 2–9.

Karpouzis K., Caridakis G., Kessous L., Amir N., Raouzaiou A., Malatesta L., Kollias S. (2007). *Modelling Naturalistic Affective States via Facial, Vocal, and Bodily Expressions Recognition*. In: T.S. Huang, A. Nijholt, M. Pantic, A. Pentland (eds.) *Artificial Intelligence for Human Computing.* Berlin: Springer Berlin Heidelberg.

Larsen J.T., Norris C.J., Cacioppo J.T. (2003). *Effects of Positive and Negative Affect on Electromyographic Activity over Zygomaticus Major and Corrugator Supercilii.* Psychophysiology, 40, p. 776–785.

Lively K.J. (2014). *Comment on 'Methodological Innovations from the Sociology of Emotions – Methodological Advances'.* Emotion Review, 7, p. 2014–2015.

Lottridge D., Chignell M., Jovicic A. (2011). *Affective Interaction: Understanding, Evaluating, and Designing for Human Emotion.* Reviews of Human Factors and Ergonomics, 7, p. 197–217.

Mandryk R.L., Atkins M.S. (2007). *A Fuzzy Physiological Approach for Continuously Modelling Emotion during Interaction with Play Technologies.* International Journal of Human Computer Studies, 65, p. 329–347.

Matsumoto D., Takeuchi S., Andayani S., Kouznetsova N., Krupp D. (1998). *The Contribution of Individualism vs. Collectivism to Cross-national Differences in Display Rules.* Asian Journal of Social Psychology, 1, p. 147–165.

Panagiotis C.P., Leontios J.H. (2010). *Emotion Recognition from EEG Using Higher Order Crossings.* IEEE Transactions on Information Technology in Biomedicine, 14, p. 186–197.

Panksepp J., Watt D. (2011). *What is Basic about Basic Emotions? Lasting Lessons from Affective Neuroscience.* Emotion Review, 3, p. 387–396.

Pantic M., Valstar M., Rademaker R., Maat L. (2005), *Web-based Database for Facial Expression Analysis.* Conference paper: Multimedia and Expo.

Pavlidis I., Eberhardt N.L., Levine J.A. (2002). *Seeing through the Face of Deception.* Nature, 415 (6867), p. 35.

Petrantonakis S.C. (2010). *Emotion Recognition from EEG Using Higher Order Crossings.* IEEE Transactions on Information Technology in Biomedicine, 14, p. 186–197.

Picard R.W., Scheirer J. (2001) *The Galvactivator: A Glove That Senses and Communicates Skin Conductivity.* Conference paper: 9th International Conference on Human–Computer Interaction.

Raine A., Venables S., Williams M. (1990). *Autonomic Orienting Responses in 15-year Old Male Subjects and Criminal Behaviour at Age 24.* American Journal of Psychiatry, 147, p. 933–937.

Redcay E., Dodell-Feder D., Pearrow M.J., Mavros S.L., Kleiner M., Gabrieli J.D.E., Saxe R. (2010). *Live Face-to-Face Interaction during fMRI: A New Tool for Social Cognitive Neuroscience*. NeuroImage, 50, p. 1639–1647.

Robinson D.T., Clay-Warner J., Moore C.D., Everett T., Watts A., Tucker T., Thai C. (2012). *Toward an Unobtrusive Measure of Emotion during Interaction: Thermal Imaging Techniques.* Advances in Group Processes, 29, p. 255–266.

Sánchez A., Ruiz J.V., Moreno A.B., Montemayor A.S., Hernández J., Pantrigo J.J. (2011). *Differential Optical Flow Applied to Automatic Facial Expression Recognition.* Neurocomputing, 74, p. 1272–1282.

Scherer K.R. (2004). *Ways to Study the Nature and Frequency of Our Daily Emotions: Reply to the Commentaries on 'Emotions in everyday life'.* Social Science Information, 43, p. 667–689.

Scherer K.R. (2000). *Psychological Models of Emotion*. In: J. Borod (ed.), *The Neuropsychology of Emotion*, New York: Oxford University Press, p. 137–162.

Scheve C. von (2003). *Sociology of Neuroscience or Neurosociology*? Advances in Medical Sociology, 12, p. 255–278.

Shilling C. (2012). *Dwie tradycje w socjologii emocji*. In: M. Rajtar, J. Straczuk (ed.). *Emocje w kulturze.* Warszawa: Wydawnictwa Uniwersytetu Warszawskiego, p. 181–210.

Stets J.E. (2010). *Future Directions in the Sociology of Emotions*, Emotion Review, 2, p. 265–268.

Teplan M. (2002) *Fundamentals of EEG Measurement.* Measurement Science Review, 2, p. 1–11.

Tsai J.L., Chentsova-Dutton Y. (2003). *Variation among European Americans in Emotional Facial Expression.* Journal of Cross-Cultural Psychology, 34, p. 650–657.

Turner J.H., Stets J.E. (2006) *Sociological Theories of Human Emotions.* Annual Review of Sociology, 32, p. 25–52.

Van den Broek E.L., Schut M.H., Westerink J.H D.M., van Herk J., Tuinenbreijer K. (2006). *Computing Emotion Awareness through Facial Electromyography*. In: T.S. Huang, N. Sebe, M.S. Lew, V. Pavlović, M. Kölsch, A. Galata, B. Kisačanin (eds.), *Computer Vision in Human - Computer Interaction.* Graz: Springer Berlin Heidelberg, p. 53–63.

Villon O., Lisetti C. (2006). *Toward Building Adaptive User's Psycho-physiological Maps of Emotions Using Bio-sensors*, 29th Annual Conference on Artificial Intelligence.

Wang H., Prendinger H., Igarashi T. (2004). *Communicating Emotions in Online Chat Using Physiological Sensors and Animated Text.* Extended abstracts of the 2004 Conference on Human Factors and Computing Systems.

Ward R., Marsden S. (2003). *Physiological Responses to Different Web Page Designs.* International Journal of Human-Computer Studies, 59, p. 199–212.

Wesley A., Buddharaju P., Pienta R., Pavlidis I. (2012). *A Comparative Analysis of Thermal and Visual Modalities for Automated Facial Expression Recognition*. In: G. Bebis, R. Boyle, B. Parvin, D. Koracin, C. Fowlkes, S. Wang, M-H. Choi, S. Mantler, J. Schulze, D. Acevedo, K. Mueller, M. Papka (eds.), *Advances in Visual Computing*, vol. 2, Rethymnon: Springer Berlin Heidelberg, p. 51–60.

Westerink J.H.D.M., Van den Broek E.L., Schut M.H. (2008). *Computing Emotion Awareness through Galvanic Skin Response and Facial Electromyography*. In: J.H.D.M. Westerink, M. Ouwerkerk, T.J.M. Overbeek, W.F. Pasveer, B. de Ruyter (eds.), *Probing Experience.* Netherlands: Springer Netherlands, p. 149–162.

Psychofizjologiczne techniki pomiaru emocji w nauce społecznej

W artykule tym przedstawiono w skrócie 6 technik pomiaru emocji: obrazowanie funkcjonalnego rezonansu magnetycznego (fMRI), elektroencefalografię (EEG), elektromiografię (EMG), reakcję skórno-galwaniczną (GSR), system kodowania ruchów twarzy (FACS) oraz termografię podczerwieni (IRT). Dla każdej techniki podano wymiar emocji (walencję lub pobudzenie), który mierzy dana technika, oraz informacje o jej wcześniejszym użyciu w socjologii, jak również przedstawiono jej główne zalety i wady. W zakończeniu omówiono ograniczenia wspólne dla wszystkich technik.

Słowa kluczowe: socjologia emocji, pomiar emocji, badanie emocji, techniki pomiarowe, wskaźniki, fizjologiczne

FOLIA 202

Annales Universitatis Paedagogicae Cracoviensis

Studia Sociologica VIII (2016), vol. 1, p. 179–189
ISSN 2081–6642

RECENZJE/REVIEWS

Anna Karnat-Napieracz
Uniwersytet Pedagogiczny im. KEN w Krakowie

Bez wzajemności albo w stronę czystego daru. Refleksje socjologiczne wokół decyzji podejmowanych przez żywych dawców narządów

Slavenka Drakulić, *Ciało z jej ciała: O banalności dobra*, tłum. Dorota Kozińska, Wydawnictwo W.A.B., Warszawa 2008, 252 s.

O książce i jej autorce

Książkę Slavenki Drakulić, chorwackiej dziennikarki, eseistki i pisarki, można odczytać jako reporterski zapis spotkań i rozmów z osobami, które, jako tak zwani „żywi dawcy", zdecydowały się dobrowolnie przekazać swoją nerkę osobie niespokrewnionej i nieznanej. Jednak, tylko i wyłącznie, takie potraktowanie omawianej publikacji nie oddaje w pełni jej sensu i głębi przekazanych treści. Jest tak zapewne dlatego, że autorka, wykorzystując swoje naukowe przygotowanie i życiowe doświadczenie, wzbogaciła reporterską narrację również rozważaniami o naturze człowieka, społecznych aspektach daru i wymiany oraz wreszcie samej istocie dobra, którego rozpoznanie było dla niej punktem wyjścia do zebrania materiału służącego do napisania książki. Moralno-etyczne aspekty medycyny transplantacyjnej przeplatają się tutaj z rozważaniami filozoficznymi i analizami ukazującymi socjologiczną wyobraźnię autorki.

Drakulić ukończyła studia literaturoznawcze i socjologiczne w Zagrzebiu i sama jest biorcą – uzyskała nerkę do przeszczepu od młodej kobiety ze Szwecji, Christine Svenson, której zadedykowała swoją książkę. Od opisu spotkania ze swoją dawczynią Drakulić rozpoczyna swoją opowieść, swoistą „podróż" po meandrach ludzkich emocji, społecznych uwikłań i niejednoznacznych motywów działań, szukając w nich ukrytych pokładów człowieczeństwa, gdzie rodzą się akty dobrej woli i altruistycznych zachowań. Autorka zna zatem problem z autopsji, zna go jednak ze strony biorcy i dlatego też ukierunkowuje swoją ciekawość badawczą na drugą stronę relacji, a więc dobrowolnych dawców.

Na treść książki składa się kilkanaście opisów historii dawców, z którymi Drakulić spotkała się, odbywając w tym celu w 2005 roku podróże, m.in. do Nowego Jorku, Filadelfii, Kansas City i Burlington. Korzystała przy tym z funduszy

pozyskanych w ramach grantu fundacji H.A. Johnsons. Tytuły kolejnych rozdziałów książki, to po prostu imiona dawców, z jednym wyjątkiem[1]. Historie te są w zasadzie odrębnymi relacjami z odbytych spotkań, aczkolwiek niektóre z nich łączą się w sposób całkiem przypadkowy. Odkrycie tych powiązań stanowi dodatkowy atut książki, ponieważ ich ukazanie pozwala uświadomić sobie, że życiowe decyzje (w dosłownym sensie – dające życie innej osobie) wynikają nie tylko z dobrowolnych aktów dobrej woli, lecz stanowią również pewien łańcuch nieprzewidywalnych zależności i uwarunkowań. Tak było w przypadku dawczyni nerki dla Drakulić. Podjęła swoją decyzję po przeczytaniu artykułu o innej młodej osobie, która podarowała nerkę niespokrewnionemu biorcy.

Warto tutaj dodać, że na początku lat dziewięćdziesiątych dwudziestego wieku Drakulić wyemigrowała czasowo z powodów politycznych do Szwecji. Planowany nieco wcześniej przeszczep nerki, pochodzącej od jej przyjaciela, nie doszedł do skutku z powodu nagłej choroby dawcy. Po kilkuletnich dializach drugi przeszczep stanowił jedyne rozwiązanie.

W skomplikowaną procedurę przygotowawczą, a dla jednej ze stron tak znaczącą z punktu widzenia ratowania jej życia, wdziera się więc przypadkowość, która w konsekwencji rodzi zależność, jaka łączy dawcę i biorcę. Urzeczywistnienie tej relacji staje się o tyle możliwe (aczkolwiek nie jest konieczne) po dokonaniu przeszczepu, o ile poznanie dawcy i biorcy jest dopuszczone przez prawo regulujące transplantację narządów od żywych dawców[2].

Należy podkreślić, że rys przewodni książki Drakulić, dotyczącej głównie motywacji żywych dawców narządów, ma swoje ugruntowanie we wcześniejszych jej rozpoznaniach dotyczących natury zła czynionego podczas wojen w byłej Jugosławii. Ta kontradykcja wymaga krótkiego wyjaśnienia. Otóż, tym razem, po przeszło dziesięciu latach od tragicznych wydarzeń wojennych autorka zastanawia się nad naturą dobra i stawia fundamentalne pytanie czy dobro jest równie banalne jak zło? Jej obserwacje zbrodniarzy wojennych sądzonych w Hadze podczas procesów przed Międzynarodowym Trybunałem Karnym dla byłej Jugosławii spowodowały

[1] Chodzi o multimilionera Kravinsky'ego, który ofiarował zarówno majątek wart 45 milionów dolarów oraz nerkę. Sprawa ta zyskała duży rozgłos w mediach. W tym jedynym przypadku autorka nie oparła swojego opisu na bezpośredniej rozmowie, a jedynie na rozmowach telefonicznych i wymianie listów elektronicznych.

[2] Z państw europejskich jedynie w Szwecji obowiązuje ustawa, która dopuszcza jednoznacznie możliwość przeszczepów od żyjących dawców (anonimowych), którzy nie są spokrewnieni z biorcą. Jednak ewentualne spotkanie dawcy z biorcą byłoby według tych przepisów niezgodne z prawem. Polskie ustawodawstwo nie dopuszcza przeszczepów altruistycznych od osób obcych. Tak więc, bezinteresowne oddanie narządu nieznanemu (obcemu) biorcy nie jest możliwe. Natomiast w przypadku głębokiej więzi emocjonalnej pomiędzy dawcą a biorcą, jeśli brak jest pokrewieństwa, konieczne są odpowiednie zezwolenia: Komisji Etycznej Krajowej Rady Transplantacji oraz sądu. W krajach, gdzie dokonuje się tego rodzaju transplantacji, zazwyczaj nie ma uregulowań dotyczących potencjalnych spotkań dawcy i biorcy, przed i/lub po zabiegu. Regulują to najczęściej przepisy konkretnych ośrodków transplantacyjnych, bądź też decyzje pozostawia się zainteresowanym stronom.

bowiem, że kwestii zła gotowa była przyznać przymiot banalności i skonstatować, że zło wojenne było w gruncie rzeczy wynikiem działań, jak to określiła, zwykłych ludzi.[3] Raz jeszcze została więc potwierdzona słynna teza Hanny Arendt o „banalności zła". W omawianej książce Drakulić napisała: „(…) miałam okazję zaobserwować i odczuć na własnej skórze dwoistość ludzkiej natury – jej dobrą stronę w chorobie i złą na wojnie." (s. 25). Dlatego też, pod brzemieniem własnej choroby (cykl dializ i powtórna operacja przeszczepu po piętnastu latach od pierwszego przeszczepu nerki od zmarłego dawcy) ponownie zadała sobie pytanie o naturę człowieka i ludzkie motywacje, jednak tym razem, motywacje do czynienia dobra. Czy i ono, analogicznie do zła, wykazuje u swojego podłoża banalność? Co może kierować postępowaniem skrajnie altruistycznym – dobrowolnym oddaniem swojego narządu („części siebie") nieznanej osobie?[4] Jakim człowiekiem jest ten, kto decyduje się ofiarować dar życia innej ludzkiej istocie? I wreszcie, w jaki sposób biorca może podziękować za otrzymany dar i czy myślenie w kategoriach podziękowania czy jakiejkolwiek gratyfikacji może być odpowiedzią adekwatną wobec wartości uzyskanego dobra? Tym dobrem jest wszak życie, możliwe dzięki udanemu przeszczepowi narządu, umożliwiającemu sprawne funkcjonowanie. Te i inne pytania stawia Drakulić wprost lub pojawiają się jako konsekwencja jej rozmów z żywymi dawcami.

W tym miejscu należy podkreślić raz jeszcze, że tutaj właśnie dochodzi do głosu Drakulić jako badaczka, której nieobcy jest warsztat socjologa i socjologiczny sposób patrzenia na rzeczywistość relacji międzyludzkich. Nie wystarcza jej dotarcie do dawców i odkrycie powierzchownych, deklaratywnych motywacji ich czynów, nawet, jeśli uzyskuje je w klasycznych sytuacjach interakcyjnych, *face-to-face*. Nie poprzestaje na reporterskim relacjonowaniu przeprowadzonych osobiście wywiadów. Ryzykując przekroczenie granicy prywatności, nie jest w stanie oprzeć się pokusie poznania rzeczywistych i prawdziwych powodów decyzji ofiarowania „części siebie". Przeprowadzając rozmowy z dawcami, rozpoznaje ich środowisko życia, relacje rodzinne, wcześniejsze koleje losu oraz wytworzoną relację z biorcą, jeśli taka powstała. Asumptem do zebrania większej ilości materiału badawczego stała się dla niej pierwsza, naznaczona osobistym doświadczeniem rozmowa z jej dawczynią. Drakulić przyznaje, że w jej rozumieniu, otrzymanie nerki od żywego dawcy, było przekroczeniem jej wyobrażeń o granicach dobra, jakiego można doświadczyć od drugiego człowieka. Otrzymanie nerki od żywego dawcy wiązała nawet z poczuciem

[3] Tego problemu dotyczy inna książka autorstwa Drakulić: *Oni nie skrzywdziliby nawet muchy: zbrodniarze wojenni przed Trybunałem w Hadze* (2006), tłum. Jakub Szacki, Warszawa: Wydawnictwo W.A.B.

[4] Należy tu wyjaśnić, że recenzowana książka ma w opisie bibliograficznym następujący tytuł: *Ciało z jej ciała. O banalności dobra*. Jednak na okładce książki (w najnowszym wydaniu z roku 2012, z którego korzystałam; do niego także odsyłam, podając strony w cytowaniach), pod pierwszym członem tytułu: *Ciało z jej ciała*, umieszczono podtytuł: *Historie ludzi, którzy podarowali komuś część siebie*. Takie brzmienie tytułu z okładki ma z pewnością zachęcić potencjalnego czytelnika do zapoznania się z nią. Jednak to, co istotne, „część siebie" stanowi wyrażenie, jakim autorka często posługuje się w opisie relacji dawca – biorca. Jest w tym sformułowaniu także dosłowne odwołanie do socjologicznej kategorii „daru" („podarowali").

niesprawiedliwości, ponieważ uznawała, że naturalnym wyjaśnieniem przeszczepu narządów od dawców zmarłych jest przekonanie, że narządy te stają się dla nich zbędne. Nie można natomiast tego samego powiedzieć odnośnie do dawców żywych, a co więcej, w tym przypadku poza zagrożeniem dla ich życia i zdrowia (nawet, jeśli niewielkim dzięki wysoce specjalistycznym procedurom medycznym, to jednak możliwym) istniała w niej obawa, a wręcz lęk, że nie będzie w stanie spłacić długu wdzięczności, w istocie niemożliwego do spłacenia. Bycie obdarowaną w tak wielkim stopniu, stwierdza Drakulić, powodowało u niej niechęć wobec transplantacji narządów pochodzących od osób żywych, bez względu na to, czy były to osoby znane czy nieznane biorcy. Relacja biorcy z dawcą jest w takim przypadku narażona na nieprzewidywalne konsekwencje ewentualnego spotkania obu stron, a zasada wzajemności (*reciprocity*), która stanowi podstawę wielu stosunków społecznych, zostaje w niej ewidentnie zachwiana.

Wzajemność, wymiana, dar

Obawa, że dokonanie przeszczepu wytworzy taki rodzaj zależności, który „skazuje" biorcę na pozycję, w której zdaje sobie sprawę, że nie jest możliwe bezpośrednie i równoważne odwzajemnienie, wydaje się – w kontekście omawianego problemu – zrozumiała. Obawa ta jest podyktowana wpisaniem zasady wzajemności w relacje międzyludzkie i znaczeniem, jakie przypisuje się jej dla utrzymania stabilności systemu stosunków społecznych.[5]

Trzeba podkreślić, że z tej „niewypłacalności" biorców zdają sobie również sprawę dawcy, którzy jednak nie oczekują w bezpośredni sposób żadnej gratyfikacji za swój czyn. Głębokie przekonanie, że taki rodzaj pomocy jest czymś naturalnym, czy wręcz wskazanym, często powtarza się w uzasadnieniach decyzji o podarowaniu „części siebie". Oto przykładowe odpowiedzi na podstawowe pytanie: „dlaczego"? „Tysiące ludzi umiera. Ktoś umrze, jeśli tego nie zrobię." (Drakulić 2012, s. 74). Lub inne wyjaśnienie, wręcz lakoniczne: „Po prostu trzeba to było zrobić." (s. 76). Nawet, jeśli uzasadnienie własnej decyzji jest bardziej rozbudowane, to i tak sprowadza się do stwierdzenia, że najbardziej istotną kwestią jest pilna potrzeba innej osoby: „Po pierwsze, mogłam to zrobić. (…). Po drugie, jestem zdrowa i w mojej sytuacji finansowej stać mnie na to, żeby komuś pomóc. Jestem spełniona zawodowo i mam rodzinę, więc w zasadzie niczego mi nie brakuje. Po trzecie, ktoś potrzebuje nerki." (s. 151) Rozmówcy autorki wskazują też na fakt, który ich zdaniem stanowi wystarczające uzasadnienie altruistycznego czynu. Jest to przekonanie, medycznie potwierdzone, że posiadając jedną nerkę można wieść życie dobrej jakości. Taki wydźwięk mają słowa jednej z dawczyń: „Nikt nie potrzebuje dwóch nerek, a przecież ktoś umrze, nie doczekawszy się dawcy. Bezczynność oznaczałby w tej sytuacji, że jestem osobą samolubną, a nie chcę za taką uchodzić." (s. 76). Przy podejmowaniu

[5] Komplementarność i wzajemność definiuje i omawia Alvin W. Gouldner (1992), wskazując na problem różnic stopnia obopólności i symetrii we wzajemności.

decyzji, czynnikiem hamującym nie okazuje się być myślenie „asekuracyjne", to jest świadome zatrzymanie swojego narządu na wypadek zaistnienia potrzeby przeszczepu w rodzinie. Oddanie organu niespokrewnionej osobie wyklucza, w sposób oczywisty, ewentualną pomoc komuś z osób najbliższych (dziecku, rodzicowi, siostrze lub bratu). Uzasadnienie takiego myślenia niesie w sobie interesujące przesłanie, które chyba najlepiej wyraża jedna z wypowiedzi: „Zrobiłam to, bo pomyślałam, co by było, gdyby coś takiego przytrafiło się któremuś z moich dzieci. Chciałabym wówczas, żeby cały świat przyszedł nam z pomocą." (s. 36). Własne postępowanie altruistyczne pozwala sądzić, że także inni zachowają się podobnie (empatycznie), ponieważ owo „inni", to po prostu ludzie tacy sami jak „ja" – osoba podejmująca decyzję o ratowaniu czyjegoś życia. Takie założenie tylko po części jest konsekwencją braku wzajemności wpisanej w relację dawca – biorca. Kryje się za nim głęboka wiara w dobroć tkwiącą w ludziach oraz ich ludzkie, empatyczne odruchy w sytuacji, gdy konieczna jest pomoc. Jednocześnie w jakimś sensie jest tu również idea rozpropagowywania tego rodzaju zachowania, co wynika z samej oceny dokonanego czynu. Jak przedstawia to jedna z dawczyń: „Zostałam dawcą, bo mogłam nim zostać, i mam nadzieję, że inni też sobie uświadomią, jak łatwo uratować komuś życie." (s. 224). Inna z osób mówi natomiast: „Oto ja, taka maleńka, a naprawdę mogę komuś pomóc." (s. 46).

Warto w tym miejscu zaznaczyć, że omawiana kwestia podarowania „części siebie" wymyka się podejściom proponowanym na gruncie socjologicznej teorii odwzajemnionych wymian.[6] Należy zatem rozważyć, czy fakt braku wzajemności o bezpośrednim charakterze zbliża akt podarowania „części siebie" do „czystego daru".

„W swej formie najczystszy dar wręczony zostaje każdemu, kto go potrzebuje, z racji tylko i wyłącznie owej potrzeby. U źródeł czystego daru leży uznanie człowieczeństwa drugiej osoby, która skądinąd jest dla darczyńcy anonimowa, nie zajmuje żadnego wyraźnego miejsca na jego poznawczej mapie świata." (Bauman 1996, s. 98).

W opiniach dawców, ten akt, wysoce altruistyczny, nie jest w gruncie rzeczy czymś nadzwyczajnym. Jest ważny z uwagi na niego samego, na sam akt darowania. Oczywiście, zdają sobie oni sprawę z wagi swojego czynu dla biorców i ich rodzin, jednak kierując się wewnętrznym przekonaniem o zasadności takiego działania nie są skłonni przypisywać mu nadmiernej wagi. Dość dobitnie wyrażają to słowa jednego z dawców: „Ofiarowałem nerkę, żeby ulżyć cierpieniu tego świata, przynajmniej odrobinę. W skali globu to pewnie niewiele, za to ogromnie dużo dla dwudziestojednoletniej dziewczyny, której udało się pomóc." (Drakulić 2012, s. 185).

Tym, co zbliża takie altruistyczne czyny do socjologicznej kategorii daru, jest niewątpliwie zasada anonimowości, jaką chcą zachować i którą kierują się żywi dawcy narządów. Chodzi tu mianowicie o niechęć wobec jakiegokolwiek kontaktu

[6] Dla przykładu, kontinuum różnych typów wymian opracowane przez M.D. Sahlinsa przewiduje na jednym jego krańcu altruistyczne oferowanie pomocy, jednak i ono zakłada wzajemność. Trudno tu także mówić o zrównoważonej wzajemności (punkcie środkowym tego kontinuum), a także o odwzajemnionych transakcjach o charakterze negatywnym. Więcej na temat zasady wzajemności i jej zastosowania w świecie relacji ludzkich zob. Bierówka (2009).

z potencjalnym biorcą przed zabiegiem. Niektórzy dawcy unikają spotkania także po dokonaniu przeszczepu, co z kolei trudno zrozumieć biorcom. Przedmiotem relacji dawca – biorca jest dar z siebie dla nieznanej osoby. Większość rozmówców powoływała się na zasadę anonimowości również z praktycznego i psychologicznie zrozumiałego powodu, jakim jest chęć uniknięcia sytuacji, w której ich decyzja byłaby warunkowana jakimikolwiek informacjami dotyczącymi potencjalnego biorcy. W wypowiedziach dawców daje się zauważyć, że przywiązują oni ogromną wagę do tej zasady, ponieważ jej zachowanie daje poczucie komfortu, że przekazany narząd otrzymuje ktoś zupełnie nieznany, nie opisywany kategoriami socjologicznymi takimi jak: wiek, płeć, rasa, klasa społeczna, narodowość, itp. Biorca jest tylko i wyłącznie człowiekiem, ale jednym z tych, który należy do *kategorii ludzi w potrzebie*. Wszelkie potencjalne informacje, dotyczące biorcy, uznane zostają przez dawców, już na wstępie, jako mogące zakłócić proces podejmowania decyzji lub mogące znacząco zaważyć na ich osobistym nastawieniu i kondycji psychicznej przed zabiegiem medycznym. Z jakichś względów dawcy, przyjmując postawę altruistyczną, można rzec, mają też na uwadze swoje dobro, a jest nim samozwrotne utwierdzenie się w przekonaniu, że jest to decyzja właściwa, autonomiczna i nie pozostająca pod wpływem żadnych czynników zewnętrznych. Za istotne uznają jedynie swoje nastawienie na sam akt darowania narządu („części siebie") komuś, kogo definiują jako innego, w rozumieniu obcego. Wydaje się, że zachowanie anonimowości, zwłaszcza przed zabiegiem, pozwala obu stronom uniknąć napięcia, ponieważ „oczyszcza" sytuację ich zależności od wszelkich uprzedzeń, rozterek i wahań, które dotyczyć mogą zarówno dawcy, jak też biorcy.[7]

Pytanie o jakiekolwiek spodziewane korzyści (zyski) dla darczyńcy w przypadku daru wydaje się przeczyć samej jego istocie. Jednak nie można pochopnie sądzić, że akt taki nie ma dla niego znaczenia. Bezinteresowność daru nie eliminuje możliwości uzyskania korzyści moralnych, które jednak wymykają się logice zyskowności. W omawianej publikacji dawcy, pytani o skutki ich decyzji dla nich samych, wskazują na takie profity, które jednak dalekie są od jednoznaczności i trudno byłoby je sklasyfikować. Jest to, na przykład, samospełnienie. Trzeba tu zauważyć, że zazwyczaj chęć ofiarowania narządu jest tak silna, że dokonanie tego aktu utwierdza jednostkę w jej własnym postępowaniu, sposobie myślenia, daje głębokie poczucie samorealizacji, Czasami odbywa się to nawet wbrew opiniom i przy braku zrozumienia osób bliskich. W niektórych przypadkach darowanie narządu ma moc terapeutyczną, okazuje się znamienne dla uzyskania szacunku dla własnej osoby, pozwala uwierzyć, że konsekwencją dobrego uczynku może być kolejne dobro, co niekoniecznie było wcześniejszym doświadczeniem dawcy. Można wreszcie dodać, że

[7] Dla przykładu, zupełnie nie planowane i przypadkowe spotkanie potencjalnego biorcy z dawcą, jakie miało miejsce przed zabiegiem, wywołało u tego drugiego akt sprzeciwu. Jego powodem był fakt, że dawcą była młoda dziewczyna (studentka), a biorcą dojrzały (ponad sześćdziesięcioletni) mężczyzna. W pierwszym odruchu uznał on, że nie może przyjąć takiego daru. Do przeszczepu jednak doszło, po mocnym uzasadnieniu decyzji ze strony dawczyni, wskazującej na to, że ona i tak odda „część siebie", natomiast biorcą będzie ktoś inny.

taka decyzja ma silne przełożenie na treść własnego życia. Stanowi ważny moment w życiu człowieka, daje mu poczucie sprawstwa i podmiotowości. Jak wyraża to jeden z potencjalnych dawców: „Mężczyzna jest bardziej narażony na pustkę niż kobieta. Tego rodzaju uczynek mógłby nadać sens mojemu życiu." (Drakulić 2012, s. 67).

Ów „czysty dar", który biorcy wydaje się poświęceniem przerastającym jego wyobrażenia, okazuje się zatem znaczący i przynoszący pewne korzyści również dawcy.[8] Pełne konsekwencje wynikające z ofiarowania narządu do przeszczepu są zresztą możliwe do uchwycenia dopiero po dokonaniu zabiegu. Akt daru i silne dążenie do jego urzeczywistnienia nie przewidują na wstępie możliwych korzyści, chociaż to, co może z niego wynikać, kryje się po części w rzeczywistych motywacjach, jakie do niego doprowadzają. Według słów jednej z dawczyń: „Nagrodą jest świadomość, że poprawiłam komuś jakość życia. Na niczym więcej mi nie zależy. (Drakulić 2012, s. 149). Widać tu wyraźnie, że chęć poprawy czyjegoś życia, wynikająca z pobudek altruistycznych, może w wyniku jej realizacji zamienić się w nagrodę, a więc satysfakcję z faktu, że do tej poprawy życia doszło. Należy także wspomnieć o „mocy daru", a ma on miejsce wówczas, kiedy dawcy mówią wprost o tym, że mają poczucie zmieniania czyjegoś losu na lepszy, a robiąc to, co uważają, że robić należy, zyskują potwierdzenie własnej dobroci. To, w zasadzie, zachwianie darem w jego czystej postaci, ponieważ należałoby tu mówić o darze biorcy dla dawcy. Byłoby nim, przykładowo, uzmysłowienie czy wręcz uzyskanie poczucia sensu istnienia, zyskanie przez dawcę na samoocenie i wzrost szacunku do samego siebie. Zysk w postaci potwierdzenia własnej dobroci leży po stronie dawcy i może być uznany za dar biorcy dla niego, ponieważ właśnie dzięki biorcy może się urzeczywistnić.

Konsekwencje spotkania obu stron po zabiegu, a szczególnie ich emocjonalne reakcje, nie są ani jednoznaczne, ani przewidywalne. Brak tu jakiegokolwiek wzoru, a dalsze kontakty dawcy i biorcy pozostają teleologicznie otwarte. Pojawia się również tutaj, podobnie jak to ma miejsce przed zabiegiem, możliwość skorzystania z zasady anonimowości. Wyrasta ona z obaw, które najlepiej oddaje następująca wypowiedź jednej z dawczyń: „To musi być straszne, jeśli biorca nie przypadnie ci do gustu, może dlatego niektórzy dawcy wolą uniknąć spotkania." (s. 227). Jednak, co warto podkreślić, właśnie po dokonaniu przeszczepu, w przypadku, kiedy kontakty są utrzymywane, relacja może przekształcić się w stosunek wymiany, w którym może pojawić się problem wzajemności, wzbogacony ponadto o wymiar emocjonalny. [9]

[8] Dobitnie świadczy o tym jeden z przypadków opisanych przez autorkę. To niesfinalizowana transplantacja nerki od jej przyjaciela, o której była już mowa w tekście. W sytuacji, kiedy zdecydowana postawa altruistyczna i chęć oddania narządu napotkały nagłą przeszkodę medyczną u potencjalnego dawcy wywołało to poczucie niesprawiedliwości. Zablokowanie pragnienia podarowania „części siebie" nie pozwoliło mu uzyskać tego, czego po tym czynie oczekiwał – poczucia sensu życia. Sama Drakulić, zdając sobie z tego sprawę, odczuwała natomiast swoiste „poczucie winy". Uzmysłowiła sobie mianowicie, że dzięki udanej transplantacji nie tylko ona sama uzyskałaby korzyść, ale także jej przyjaciel poprzez ten dar zyskałby wiele dla siebie, potwierdzając sens swojego życia dobroczynnym aktem.

[9] To zagadnienie można byłoby analizować odrębnie jako przykład zastosowania socjologii emocji (Turner, Stets 2009) do badania motywacji zachowań ludzi wchodzących w re-

Zamiast podsumowania

Książka Slavenki Drakulić wnosi wiele ciekawych wątków do filozoficznych dyskusji dotyczących natury człowieka oraz socjologicznych analiz więzi społecznych i organizacji życia społecznego. W dużej mierze ukazuje również niejednoznaczność relacji międzyludzkich w wielowymiarowym i skomplikowanym społeczeństwie współczesnym. Szczególny charakter stosunków opisywanych przez Drakulić wyraża się tym, że nie podlegają one zasadzie wzajemności, a co więcej, zrozumienie ich natury wymaga przemyślenia na nowo pojęcia daru i zmiany pojmowania odpowiedzialności. Pojawia się tu kwestia odpowiedzialności „za" drugiego człowieka, bo czyn polegający na ofiarowaniu „części siebie" obcej osobie dla ratowania jej życia z trudem poddaje się normatywnej regulacji (odpowiedzialność „wobec").[10] Nie sposób również jednoznacznie ustalić, jakiej zasadzie etycznej miałyby podlegać tego rodzaju działania. Ponadto byłoby wysoce dyskusyjne, aby uznać, że zasada taka samodzielnie i niepodzielnie reguluje takimi zachowaniami. Motto, jedno z dwóch, jakie autorka umieściła na początku książki, to znany werset z Talmudu: „Kto ratuje jedno życie, jakby cały świat ratował." Jeśli z tej formuły daje się wyprowadzić jakaś konkretna norma (raczej religijna, a nie prawna), to jednak trudno jej przypisać powszechną obowiązywalność i imperatyw jej przestrzegania. Postępowanie zgodne z nią, odpowiedzialne „wobec" niej, niekoniecznie zaś musi odwoływać się do motywacji altruistycznych.

Trudno więc w jednoznaczny sposób odpowiedzieć na postawione przez autorkę pytanie główne – dlaczego dawcy TO robią? Nie ma wyznaczonych wzorów ani konkretnych motywów skłaniających do podarowania „części siebie" drugiej, obcej osobie. To, co jest dostępne, to historie życia ludzi, historie mniej lub bardziej umiejętnie przez nich opowiedziane, którzy decydując się na taki czyn, nie zrobili tego pod wpływem chwili, nagłego impulsu czy porywu serca. To, co niewątpliwie wiąże losy wszystkich dawców opisanych w książce, to ich decyzje, głęboko przemyślane, do których dojrzewali przez dłuższy czas, niejednokrotnie angażując się w inne formy pomocy osobom potrzebującym.[11] W swoim wcześniejszym postępowaniu dawcy działali na rzecz dobra innych, co pozwala lepiej zrozumieć tkwiące w nich przekonanie, że podejmując decyzję o darowaniu narządu, zrobili w zasadzie niewielki krok na drodze czynienia dobra. To z punktu widzenia biorcy decyzja ta jawi

lacje wymiany lub daru oraz poznania ich zachowań względem siebie po przeprowadzeniu zabiegu przeszczepu. Zasadniczo rzecz ujmując, próbuje się tu dookreślić spójność pomiędzy oczekiwaniami a doświadczeniami, co w kontekście omawianego problemu mogłoby dostarczyć interesujących spostrzeżeń.

[10] Rozróżnienie odpowiedzialności „za" i „wobec" znakomicie przedstawia Zygmunt Bauman (Bauman 1995).

[11] Dawcy nerek opisani przez autorkę, byli wcześniej dawcami osocza i płytek krwi, osobami zarejestrowanymi w banku dawców szpiku kostnego, zajmowali się opieką nad starszymi i niedołężnymi. Byli wśród nich: pracownik domu opieki, bibliotekarka z aspiracjami podjęcia studiów w zakresie pracy socjalnej, pisarka zajmująca się przewlekle chorą siostrą, multimilioner i celebryta, osoba z wykształceniem nauczycielskim podejmująca prace dorywcze i inni.

się jako wielka, nadzwyczajna i przerastająca możliwości człowieka, wręcz trudna do wytłumaczenia i tym samym, do przyjęcia. Jak mówi jeden z dawców: „(…) dobro zaskakuje nas, bo nieczęsto się przytrafia." (Drakulić 2012, s. 238). Natomiast biorcy mają jakiś wewnętrzny imperatyw poszukiwania wytłumaczenia sytuacji, w której znaleźli się dzięki decyzji drugiej osoby. Tkwią w koniecznym dla nich dążeniu do zrozumienia niezwykłości tej sytuacji i szukają za wszelką cenę odpowiedzi na nurtujące ich pytanie „dlaczego?". Wynika to z pewnością z nierównoważności i asymetryczności tej relacji. Autorka przyznaje, że jedno z opisanych przez nią spotkań (intensywne, trzydniowe), z uwagi na możliwość obserwacji jednocześnie dawcy i biorcy, w jakimś sensie zdjęło z niej „brzemię daru". (s. 229). Rozpoznawanie motywacji i rzeczywistych uzasadnień zachowania altruistycznego pozwoliło jej zrozumieć, że różnorakie powody tych decyzji wiązały się jednak z pewnymi korzyściami po stronie dawców, co z punktu widzenia biorcy nie umniejsza w żaden sposób dokonanego aktu altruizmu. Dar z siebie nie wywołuje u dawców myślenia w kategoriach „straty", lecz bez względu na to, kto nim zostaje obdarowany, akt ten ubogaca ich i utwierdza w przekonaniu dobrze podjętej decyzji, przynosząc olbrzymią satysfakcję odczuwaną jako spełnienie.

Z perspektywy biorcy to rozpoznanie pozwala z kolei oswoić się z otrzymanym dobrem. Tak jak zostało to wspomniane na wstępie, świadomość pozostawania w długu wdzięczności nie do spłacenia stanowiła dla autorki ważny impuls do naukowego wyjaśnienia i rozpoznania motywacji dawców. Napisanie książki, jak można wnioskować z jej końcowych fragmentów, niekoniecznie przyniosło jednoznaczne wnioski i konkluzje satysfakcjonujące autorkę-badacza, a więc w aspekcie rezultatów naukowych. Pozwoliło natomiast częściowo zrozumieć różne punkty widzenia dawców i biorców oraz przybliżyć to, co de facto stanowi o tej relacji i, co ją wyznacza. Drakulić nie ukrywa, że deklarowane przez dawców motywy ich decyzji nie dają możliwości stworzenia ich typologii, nawet ograniczonej do tych przypadków, które poddała badaniu. Są to decyzje jednostkowe, osobiste, których podłoże może tylko z pozoru wydawać się podobne. Jednak różniące je szczegóły (pewne zdarzenia, czasem drobne fakty, osoby z życia dawców, doświadczenia utraty lub choroby w rodzinie, wreszcie cechy psychologiczne i emocjonalne konkretnych osób) tak osobliwie są powiązane ze sobą w syndromie ich cech wspólnych, że rozpoznanie wagi tych elementów i ich przełożenia na ostateczną decyzję o podarowaniu części siebie, nie jest możliwe. Zresztą nie jest to chyba niezbędne.

Wymowa i przesłanie książki są inne. Odkrywanie jak rodzi się dobro, jakie są sposoby jego rozprzestrzeniania się, być może nawet nabywanie umiejętności jego czynienia, metody „zarażania" pozytywnym przykładem i wreszcie ukazanie, że dobro najwyższej jakości jest w gruncie rzeczy dziełem zwykłych ludzi – to najistotniejsze walory recenzowanej publikacji. Jakkolwiek książka nie zawiera moralizatorskich elementów, to socjologiczny zmysł jej autorki nie pozwala ominąć pewnych kwestii, które ze społecznego punktu widzenia mogłyby być znaczące dla zwiększenia liczby transplantacji i upowszechniania tego rodzaju pomocy ludziom, którzy

jej bezwzględnie potrzebują[12]. Nie jest to jedynie kwestia decyzji indywidualnych. Autorka zastanawia się czy w sytuacji, kiedy dawcy są w pełni świadomi swoich czynów i wynikają one z otaczającej ich atmosfery społecznej aprobaty, a nawet podziwu, altruizm taki jest być może łatwiej ujawniany. W szerszym kontekście, jest to pytanie o to, czy altruizm rodzi się na podłożu szczególnego rodzaju stosunków społecznych, a także, jakie znaczenie dla przyszłych zachowań może mieć propagowanie tego typu zachowań. Drakulić pisze wprost: „Czy altruizm jest domeną szczególnego rodzaju ludzi, czy też zależy od wartości uznanych przez społeczeństwo, w którym się wychowaliśmy?" (s. 82). Innymi słowy, czy altruizm wchodzi w zakres kapitału społecznego, który przekazywany i utrwalany, utwierdza się w jednych społeczeństwach, podczas gdy w innych społeczeństwach zjawisko to ma znaczenie marginalne, a tym samym trudne do zrozumienia i przyjęcia.

Rozpoznawanie podłoża motywacji żywych dawców tylko częściowo przybliża do odpowiedzi na to pytanie. Fakt ze przeszczep narządów od żywych dawców niespokrewnionych jest praktyką znaną i stosowaną w niektórych krajach, a w innych prawnie niedopuszczalną, bądź mocno ograniczoną, świadczy nie tyle o możliwościach technologicznych (poziomie usług medycznych) poszczególnych krajów, ale mówi coś konkretnego o tych społeczeństwach.

Próba sformułowania omawianego problemu w kategoriach socjologii teoretycznej prowadzi do konstatacji, że akt „podarowania części siebie" zbliża się do kategorii daru. Niemniej jednak, można się w tym wypadku zgodzić ze słowami Zygmunta Baumana: „Regułą życia codziennego jest właśnie wieloznaczność stosunków międzyludzkich, a nie ich modelowa czystość". (Bauman 1996, s. 113). Wynikające z daru konsekwencje dla obu stron relacji nie upoważniają do jednoznacznych ocen i opisów w rodzaju „czystej relacji". Podobnie jak tego rodzaju osądy nie mogą być arbitralnie zastosowane wobec zachwiania zasady wzajemności. Ta niejednoznaczność wymaga raczej, ponownego przemyślenia zasady wzajemności, a zwłaszcza jej stosowalności i odmiennego znaczenia przypisywanego jej w wysoko zaawansowanych społeczeństwach współczesnych.

Bibliografia

Bauman Z. (1996). *Socjologia*, tłum. J. Łoziński, Poznań: Wydawnictwo Zysk i S-ka.

Bauman Z. (1995). *Wieloznaczność nowoczesna. Nowoczesność wieloznaczna*, tłum. J. Bauman, Warszawa: PWN.

Bierówka J. (2009). *Zasada wzajemności w społeczeństwie informacyjnym.* Kraków*:* Oficyna Wydawnicza AFM.

[12] Fakty przedstawione w książce nie pozostawiają wątpliwości, co do tego, że liczba narządów pobranych u osób ze stwierdzoną śmiercią mózgową jest niewystarczająca w odniesieniu do liczby osób oczekujących na przeszczep. Co więcej, nawet łącznie, narządy od żywych i zmarłych dawców nie są w stanie zapewnić, że nie dojdzie do śmierci kogoś kto zbyt długo oczekuje na przeszczep.

Drakulić S. (2008). *Ciało z jej ciała. O banalności dobra*, tłum. Dorota Kozińska, Warszawa: Wydawnictwo W.A.B.

Drakulić S. (2006). *Oni nie skrzywdziliby nawet muchy: zbrodniarze wojenni przed Trybunałem w Hadze*, tłum, J. Szacki, Warszawa: Wydawnictwo W.A.B.

Gouldner A.W. (1992). *Norma wzajemności. Preliminaria.* W: Kempny M., Szmatka J. (red.). *Współczesne teorie wymiany społecznej. Zbiór tekstów*. Warszawa: Wydawnictwo Naukowe PWN, s. 73–106.

Kempny M., Szmatka J. (red.). (1992). *Współczesne teorie wymiany społecznej. Zbiór tekstów.* Warszawa: Wydawnictwo Naukowe PWN.

Turner J.H., Stets J.E. (2009). *Socjologia emocji*, tłumaczenie Marta Bucholc, Warszawa: Wydawnictwo Naukowe PWN.

FOLIA 202

Annales Universitatis Paedagogicae Cracoviensis

Studia Sociologica VIII (2016), vol. 1, p. 190–197
ISSN 2081–6642

Małgorzata Krywult-Albańska
Uniwersytet Pedagogiczny im. KEN w Krakowie

Cicha rewolucja. Czy rosnąca skala życia w pojedynkę zmieni oblicze społeczeństwa?

Eric Klinenberg, *Going Solo. The Extraordinary Rise and Surprising Appeal of Living Alone*, Duckworth Overlook, London 2014, 273 s.

Rodzina stanowi jeden z centralnych obszarów zainteresowania socjologii pełniąc – wśród wielu innych – funkcję sytuowania jednostek w przestrzeni społecznej i określania ich tożsamości. Dla demografów ważna jest z uwagi na fakt, że zmiany jej dotyczące wpływają z kolei na zmiany procesu odtwarzania pokoleń. Dotykające ją przeobrażenia, związane z przemianami ról rodziców (zwłaszcza ojców), nowymi jej formami (związki osób tej samej płci, rodziny zrekonstruowane itp.) czy też rozwojem medycyny reprodukcyjnej, są jednocześnie przedmiotem zainteresowania opinii publicznej i emocjonalnych debat toczonych w mass mediach. Wśród opublikowanych w ostatnich latach opracowań powiązanych z tym tematem na szczególną uwagę zasługuje książka amerykańskiego socjologa Erica Klinenberga: *Going Solo. The Extraordinary Rise and Surprising Appeal of Living Alone*[1]. Nie traktuje ona o rodzinie jako takiej, ale o życiu w pojedynkę (*solo life*), które – stając się na całym świecie coraz bardziej popularne – jak się wydaje, systematycznie zyskuje na atrakcyjności jako alternatywa dla życia w ramach rodziny i niesie z sobą rewolucyjną wręcz zmianę społeczną o trudnych dzisiaj do przewidzenia konsekwencjach.

Zawarte w opracowaniu Klinenberga analizy opierają się na obserwacji etnograficznej oraz wywiadach pogłębionych, przeprowadzonych przez autora i jego współpracowników z ponad trzystoma osobami mieszkającymi w pojedynkę. Dodatkowo przeprowadzono wywiady z osobami, które wchodzą w interakcje, wspierają w jakiś sposób lub pracują na rzecz takich osób, m.in. pracownikami socjalnymi, urzędnikami, planistami oraz architektami. Z uwagi na miejsce badań – duże ośrodki miejskie, zwłaszcza Nowy Jork (głównie dzielnice: Brooklyn, Bronx, Manhattan oraz Queens) i kilka innych dużych miast amerykańskich – zawarte w książce refleksje odnoszą się, jak zastrzega autor, do życia w pojedynkę w miastach. Obok badań własnych, w publikacji wykorzystano obszerną literaturę naukową dotyczącą omawianej problematyki, zarówno amerykańską, jak i z innych części

[1] Pierwsze wydanie tej książki ukazało się w Stanach Zjednoczonych w 2012 r., niniejszy tekst opiera się na londyńskim wydaniu z roku 2014.

świata, m.in. Wielkiej Brytanii, Francji, Australii, Chin, Japonii, Korei Południowej, Indii i Brazylii. Wielość źródeł oraz przejrzysty charakter narracji, przeplatanej licznymi studiami przypadków, czynią z tej publikacji lekturę fascynującą, nie tylko dla familiologów.

Swoje analizy autor zaczyna od spostrzeżenia, iż życie w pojedynkę, które wydaje się przeczyć społecznej naturze człowieka (jak również większości zwierząt), stało się w ostatnich latach zjawiskiem zaskakująco powszechnym. Statystyki dotyczące Stanów Zjednoczonych dobitnie o tym świadczą. W 1950 roku 22% dorosłych Amerykanów było singlami. Cztery miliony ludzi żyło samotnie, co stanowiło 9% wszystkich gospodarstw domowych. Życie w pojedynkę było powszechne wówczas zwłaszcza w stanach zachodnich, przyciągających migrujących pracowników, i stanowiło z reguły przejściowy etap na drodze do „konwencjonalnego" życia domowego. Obecnie już ponad 50% Amerykanów to single, 31 milionów (mniej więcej jedna na siedem osób dorosłych) żyje w pojedynkę (pomijając 8 milionów osób zamieszkujących w instytucjach zbiorowego zakwaterowania, takich jak domy pomocy społecznej czy więzienia), stanowiąc 28% wszystkich gospodarstw domowych. Przeważają wśród nich kobiety, największą grupę stanowią osoby w wieku 35–64 lat. Sporą grupę stanowią osoby starsze, natomiast najszybciej rośnie udział młodych dorosłych w wieku 18-34 lata. Żyjący w pojedynkę zamieszkują głównie obszary metropolitarne we wszystkich regionach kraju. Na Manhattanie ich gospodarstwa domowe stanowią ponad połowę ogółu. Jak zauważa Klinenberg (s. 5), łącznie z bezdzietnymi parami, gospodarstwa te stanowią dominującą formę gospodarstw domowych, bardziej powszechną niż rodzina nuklearna, rodzina wielopokoleniowa, czy też zbiorowe gospodarstwa tworzone przez współlokatorów (*roommate or group home*).

Porównanie międzynarodowe pokazuje, że w innych krajach życie w pojedynkę jest jeszcze bardziej powszechne – najbardziej w Szwecji, Norwegii, Finlandii i Danii, gdzie około 40–45% wszystkich gospodarstw domowych to gospodarstwa jednoosobowe (s. 10). Większy niż w USA odsetek jednoosobowych gospodarstw domowych mają także Japonia, Niemcy, Francja, Wielka Brytania, Australia, Kanada. Najszybciej odsetek ten rośnie w Chinach, Indiach i Brazylii.

Klinenberg wskazuje na dwa główne czynniki odpowiedzialne za ten ogromny przyrost. Pierwszy to wzrost poziomu zamożności oraz zabezpieczeń społecznych, które zapewniają współczesne *welfare states*. Mówiąc w uproszczeniu, obecnie więcej ludzi może sobie pozwolić na życie w samotności, niż miało to miejsce dawniej. Drugim czynnikiem jest zmiana kulturowa, którą Durkheim określał mianem „kultu jednostki", a zatem szeroko rozumiane procesy indywidualizacji.

Bardziej szczegółowa analiza pokazuje, że warunki, w których „jednostka może rozkwitać" (s. 13) powstały w społeczeństwach zachodnich (i nie tylko) dopiero w drugiej połowie dwudziestego wieku. Na owe warunki składają się:

1. Wzrost pozycji społecznej kobiet, polegający na dostępie do wykształcenia na podobnym poziomie co mężczyźni, masowym wejściu na rynek pracy, jak również zyskaniu większej kontroli nad własnym ciałem, życiem seksualnym

i reprodukcją, z czym wiąże się zmiana charakteru związków, opóźnianie małżeństwa, dłuższy okres przejścia do dorosłości i częstszy rozpad związków;
2. Rewolucja komunikacyjna. Nowe technologie, począwszy od telefonu, poprzez telewizję, po Internet, umożliwiają uczestniczenie w życiu społecznym także osobom mieszkającym samotnie;
3. Masowa urbanizacja, która umożliwiła powstanie subkultury singli, podzielających podobne systemy wartości oraz style życia, podtrzymywane przez rozbudowaną miejską infrastrukturę uwzględniającą potrzeby osób żyjących w pojedynkę[2];
4. Rewolucja długowieczności, w wyniku której doświadczenie długiego życia w pojedynkę stało się w szczególności udziałem kobiet żyjących dłużej od mężczyzn.

Przyjrzawszy się uwarunkowaniom zjawiska, Klinenberg analizuje różne oblicza życia w pojedynkę, przyznając jednocześnie, że jego badania ograniczone są w dużym stopniu do doświadczeń klasy średniej. Wśród osób, z którymi przeprowadzono wywiady, znaleźli się przedstawiciele różnych grup społecznych: młodzież i osoby w fazie wczesnej dorosłości, osoby żyjące w pojedynkę po rozwodzie lub separacji, samotni mężczyźni w średnim wieku, o niskich kwalifikacjach, zamieszkujący budynki zbiorowego zakwaterowania[3], osoby starsze. Wśród doświadczeń związanych z życiem w pojedynkę można wymienić poczucie osamotnienia, dyskryminację w pracy, stygmatyzację, skłonność do izolacji społecznej oraz „koszmar samotnego chorowania i umierania". Na podstawie przeprowadzonych badań autor dochodzi jednak do wniosku, że tej formy życia nie można sprowadzać do takich problemów, ponieważ doświadczenia ogromnej większości osób są zdecydowanie bogatsze i bardzo zróżnicowane, a dylematy podobne do tych związanych z życiem w pojedynkę (czy byłoby mi lepiej w związku, ze współlokatorem/ką? itp.) nie są też obce osobom mieszkającym z partnerami lub innymi osobami (czy byłoby mi lepiej samej/samemu, z innymi?). Osoby należące to tej ostatniej kategorii także mogą doświadczać samotności, zatem życia w pojedynkę nie należy utożsamiać z poczuciem osamotnienia. Cytowane przez autora osoby rozwiedzione lub separowane wspominają o tym, że życie z niewłaściwą osobą też powoduje poczucie osamotnienia nieporównywalne z niczym innym (s. 20).

W kontekście prowadzonych na ten temat debat istotna wydaje się odpowiedź na pytanie, czy rosnącą skalę życia w pojedynkę traktować należy – jak czyni to wielu komentatorów, naukowców i polityków – jako problem społeczny, oznakę narcyzmu, fragmentacji i atrofii życia społecznego? Klinenberg odrzuca tego typu alarmistyczne w swojej wymowie poglądy, podkreślając w oparciu o przeprowadzone przez siebie badania, iż życie w pojedynkę jest jednostkowym wyborem,

[2] Jak podkreśla Klinenberg, życie w pojedynkę wyrasta z kultury współczesnych miast, a nie na przykład – jak można by sądzić – z tradycji monastycznych (s. 47).

[3] Tzw. *single-room occupancy dwellings*, zamieszkiwane przez osoby „z przeszłością" – po doświadczeniach nadużywania alkoholu i narkotyków, byłych więźniów, ubogich, bezrobotnych.

nie mniej uprawnionym niż wybór małżeństwa lub zamieszkania z partnerem. Upowszechnienie tej formy życia jest zbiorowym osiągnięciem, czego potwierdzeniem jest fakt częstszego występowania w krajach zamożniejszych w przeciwieństwie do krajów biedniejszych, w których zjawisko pojawia się na znacznie mniejszą skalę. Jednocześnie zamieszkiwanie w pojedynkę nigdzie nie jest traktowane jako cel sam w sobie i dlatego próby organizowania ruchów społecznych działających na rzecz interesów osób żyjących samotnie napotykają na duże trudności. Nigdzie także zjawisku temu nie towarzyszył zanik wspólnotowych form życia. Wręcz przeciwnie – osoby mieszkające samotnie (zwłaszcza kobiety) utrzymują niejednokrotnie intensywniejsze przyjaźnie i znajomości, oraz są bardziej społecznie zaangażowane niż osoby mieszkające z partnerami.

Życie w pojedynkę pozostanie trwałą cechą krajów wysoko rozwiniętych, ponieważ procesy, które prowadzą do pojawienia się tego zjawiska (indywidualizacja, wzrost statusu kobiet, urbanizacja, rozwój technologii komunikacyjnych), wydają się mieć nieodwracalny charakter. Przytoczony w końcowych partiach książki przykład Szwecji pokazuje jednocześnie, że rozpowszechnienie tej formy życia nie prowadzi bynajmniej – jak wspomniano wcześniej – do rozpadu życia wspólnotowego, ani zaniku znaczących więzi społecznych (s. 213). Odsetek jednoosobowych gospodarstw domowych sięga tutaj 47% (w porównaniu do 28% w USA), a w Sztokholmie nawet 60%. Umożliwia to duża dostępność małych mieszkań, projektowanych z myślą o osobach żyjących samotnie, dynamiczna gospodarka rynkowa oraz rozwinięte instytucje państwa opiekuńczego (przywołując cytowane w książce słowa, tak wielu Szwedów mieszka samotnie, ponieważ może sobie na to pozwolić). Dla młodzieży zamieszkanie we własnym domu stanowi niezbędny etap w procesie stawania się dorosłym – młodzi ludzie zyskują czas, aby zainwestować w siebie w czasach, gdy związki stały się bardzo kruche, a rynek pracy jest niepewny i wymagający. Osobom rozwiedzionym lub separowanym pozwala na odzyskanie niezależności i samokontroli, a starszym na zachowanie godności, niezależności i możliwości decydowania o własnym losie.

Autor zdecydowanie odrzuca powszechne obecnie w literaturze socjologicznej (określa ją mianem *pop sociology*) utożsamianie życia w pojedynkę ze wzrostem poczucia osamotnienia, upadkiem społeczeństwa obywatelskiego oraz zanikiem troski o dobro wspólne. Zdaniem autora, takie uogólnienia są szkodliwe przede wszystkim dlatego, że odwracają uwagę od osób faktycznie potrzebujących pomocy, żyjących w całkowitej izolacji, oraz pociągają za sobą ignorowanie problemów wymagających pilnego rozwiązania (s. 230). Znowu daje znać o sobie pragmatyczne podejście autora. Skoro upowszechnianie się życia w pojedynkę jest procesem nieodwracalnym, nieodłączną cechą współczesnych społeczeństw wysoko rozwiniętych, zamiast przekonywać, że lepszy byłby powrót do życia wspólnotowego, powinniśmy raczej myśleć o tym, w jaki sposób zaspokoić potrzeby osób żyjących w pojedynkę.

Z samotnym zamieszkaniem mogą się wiązać rozmaite problemy, takie jak: izolacja i niewystarczająca opieka nad chorującymi i ubogimi starszymi osobami; stres

i niepokój samotnych kobiet, które chciałyby mieć dziecko a zbliżają się do granic wieku rozrodczego; czy też niestabilna sytuacja ekonomiczna osób, tracących pracę, a nie mających partnera, na którego materialną pomoc mogłyby liczyć do czasu znalezienia nowej pracy. Dla takich problemów można jednak znaleźć dobre rozwiązania. Wydają się nie na miejscu ponure diagnozy w rodzaju: „śmierć wspólnoty" czy też „upadek społeczeństwa obywatelskiego", potrzebne są natomiast konkretne działania, takie jak wspieranie kobiet samotnie wychowujących dzieci. W Szwecji obejmuje ono opłacanie przez państwo i pracodawcę trwającego szesnaście miesięcy urlopu rodzicielskiego, jak również w wysokim stopniu subsydiowaną przez państwo opiekę nad dzieckiem oraz doskonale funkcjonującą służbę zdrowia (s. 222). Inna propozycja to różne formy wspieranego przez państwo budownictwa dla osób samotnych znajdujących się w trudnej sytuacji życiowej – jak wykazuje autor – rynek okazał się pod tym względem niewydolny.

Nie można ignorować także korzyści (zarówno dla jednostek, jak i społeczeństwa), jakie przyniosło pojawienie się życia w pojedynkę na dużą skalę. Można tu wskazać szereg przykładów. Dzięki samotnym młodym osobom oraz tym w średnim wieku doszło do rewitalizacji życia publicznego w miastach. Osoby takie częściej niż osoby mieszkające z innymi spędzają czas z przyjaciółmi i sąsiadami, a także chodzą do barów, restauracji, uczestniczą w różnego typu nieformalnych grupach (*civic groups*) i formach aktywności. Z kolei kulturowa akceptacja życia w pojedynkę uwolniła kobiety od „złych małżeństw i opresyjnych rodzin", przez co uzyskały pełniejszą kontrolę nad własnym życiem, a także zwiększyło się ich uczestnictwo w życiu obywatelskim. Wbrew obawom o zgubne skutki, jakie przyniesie dla środowiska rosnący odsetek osób żyjących w pojedynkę, wiele wskazuje na to, że życie w małych mieszkaniach i relatywnie „zielonych" miastach (a nie w domach na przedmieściach tworzących autonomiczne enklawy) oznacza mniejsze zużycie energii. Życie w pojedynkę może wreszcie zapewnić czas i przestrzeń konieczne, aby móc zaangażować się w znaczące relacje z innymi lub na rzecz jakiejś sprawy[4].

Podsumowując, opracowanie Klinenberga stanowi doskonałe studium jednego z najbardziej interesujących współcześnie zjawisk, wartościowe również pod względem metodologicznym. Opierając się zasadniczo na danych jakościowych, autor wykorzystuje także dane z badań o charakterze ilościowym, budując spójną narrację fascynującą nie tylko dla socjologów, ale dla wszystkich zainteresowanych procesami zachodzącymi we współczesnych społeczeństwach. Klinenberg unika przy tym sensacyjnych wątków, z którymi – jak pisze w aneksie opisującym metodologię badań – zetknął się w trakcie badania świata osób żyjących samotnie. Jako socjolog analizuje wspólne doświadczenia i orientacje ludzi żyjących w pojedynkę, starając się odkryć i opisać podstawowe cechy tego sposobu społecznego

[4] Warto w tym miejscu przytoczyć konkluzję, która pojawia się pod koniec książki Klinenberga. Autor zauważa – przytaczając wypowiedzi socjologów i psychologów, ale także poetów i pisarzy – że samotność przeżywana w odpowiedni sposób nie tylko umożliwia wyzwolenie tkwiącej w nas energii, ale podpowiada nam pomysły jak ulepszyć nasze wspólne życie.

funkcjonowania.[5] Wnioski, które wyciąga, mają implikacje praktyczne – wskazują na potrzebę rozwiązania niektórych problemów, z jakimi wiązać się może życie w pojedynkę i zawierają propozycje konkretnych rozwiązań. *Going Solo* to zaangażowana, oparta na faktach i świetnie napisana praca, analizująca zjawisko, które w Polsce prawdopodobnie także będzie poszerzało zasięg swego występowania. W jakim stopniu zmieni ono życie nas samych, naszych rodzin, wspólnot, państw? Jak zauważa Klinenberg, mamy tu do czynienia z ogromną transformacją społeczną. Obserwujemy dopiero jej początkową fazę i nie jesteśmy jeszcze w stanie przewidzieć jej konsekwencji.

Polski kontekst i badania dotyczące życia w pojedynkę

Na marginesie refleksji nad książką Klinenberga warto odnieść się krótko do polskiego kontekstu i badań poświęconych życiu w pojedynkę w Polsce. Powstało na ten temat już kilkanaście opracowań opartych głównie o badania jakościowe i opisujących różne aspekty życia singli (Bujała 2013; Czernecka 2008, 2010, 2011; Grzeszczyk 2005; Paprzycka 2008; Szpakowska 2006; Żurek 2004, 2005, 2008a, 2008b, 2011).

Zakrojone na stosunkowo dużą skalę badania przeprowadziła Aldona Żurek (2008a), która przebadała 600 osób w wieku 20–50 lat, stanu wolnego, tworzących jednoosobowe gospodarstwa domowe, mieszkających w Poznaniu oraz mniejszych miastach województwa wielkopolskiego. Trudniej jest oszacować, jak bardzo rozpowszechnione w Polsce jest życie w pojedynkę. Dostępne są dane Głównego Urzędu Statystycznego na temat jednoosobowych gospodarstw domowych, jednak z definicji gospodarstwa domowego wynika, że głównym kryterium jego wyodrębnienia jest odrębne źródło utrzymania, a nie osobne zamieszkiwanie. Teoretycznie zatem możliwe jest wspólne zamieszkanie kilku osób, które tworzą jednocześnie odrębne gospodarstwa domowe. Zatem przytoczone niżej dane dotyczące tych ostatnich stanowią jedynie pewne przybliżenie skali i trendów dotyczących życia w pojedynkę.

Podobnie jak w innych krajach, także w Polsce zauważalna jest tendencja spadkowa, jeśli chodzi o wielkość gospodarstwa domowego. W roku 2011 średnia wielkość gospodarstwa domowego w Polsce (mierzona przeciętną liczbą osób) wynosiła 2,82 w stosunku do 3,1 z roku 1988. Jest to wynik znaczącego wzrostu odsetka gospodarstw jednoosobowych (GUS 2003, 2013). Analiza zmian w strukturze ludności według sytuacji rodzinnej na przestrzeni ostatnich dekad wskazuje jednocześnie na spadek skłonności do tworzenia gospodarstw wielorodzinnych oraz do współzamieszkiwania z osobami spoza rodziny (Kotowska 2014, s. 30).

Odsetek osób tworzących jednoosobowe gospodarstwo domowe wzrósł wśród osób w wieku 20–64 lat (zarówno kobiet, jak i mężczyzn). Jednak także osoby

[5] Klinenberg zainteresował się problematyką życia w pojedynkę w połowie lat 90., gdy w następstwie śmiertelnej fali upałów w Chicago w roku 1995 setki osób zmarło samotnie w swoich domach. Sprawie tej poświęcił całą książkę (Heat Wave. A Social Autopsy of Disaster in Chicago. Chicago-London 2002: The University of Chicago Press).

w wieku 65 lat i więcej coraz częściej tworzą jednoosobowe gospodarstwa domowe i coraz rzadziej mieszkają z rodzinami dorosłych dzieci. Odsetek starszych mężczyzn tworzących takie gospodarstwa wzrósł z 10% do 14% w latach 1988–2002, natomiast kobiet w tym samym wieku z 28% do 34% (według danych ze spisów powszechnych). Dane Diagnozy Społecznej 2013 potwierdzają kontynuację tego trendu.

Tabela 1. Struktura gospodarstw domowych w Polsce według wielkości (w %) i średnia wielkość gospodarstwa

Gospodarstwa	1988	2002	2011
Jednoosobowe	18,3	24,8	24,0
Dwuosobowe	22,3	23,2	25,7
Trzy- i czteroosobowe	42,3	37,9	36,4
Pięcioosobowe i większe	17,1	14,1	13,9
Średnia wielkość gospodarstwa	3,1	2,84	2,82

Źródło: dane GUS pochodzące ze spisów powszechnych

Zasadniczo wydaje się, że w Polsce występują i będą nadal występowały przesłanki do wzrostu liczby osób żyjących w pojedynkę. Należą do nich wskazane przez Klinenberga uwarunkowania: procesy indywidualizacji, pociągające za sobą w sferze relacji rodzinnych wzrost akceptacji dla sposobów życia alternatywnych w stosunku do rodziny nuklearnej; wzrost poziomu wykształcenia, zwłaszcza wśród kobiet, których coraz większy odsetek posiada wykształcenie wyższe[6]; rewolucja komunikacyjna oraz wydłużanie się przeciętnego dalszego trwania życia (przy czym kobiety, według danych GUS z roku 2014, żyją przeciętnie prawie osiem lat dłużej od mężczyzn). Mniej jednoznaczny charakter mają jedynie postępy w zakresie procesów urbanizacji (Polska należy do średnio zurbanizowanych krajów, a tempo procesów urbanizacji uległo w ostatnich latach zahamowaniu).

Bibliografia

Bujała A. (2013). *Więzi społeczne polskich singli*. Uniwersytet Łódzki, Wydział Ekonomiczno-Socjologiczny. Praca doktorska dostępna w Repozytorium UŁ: http://dspace.uni.lodz.pl:8080/xmlui/.

Czernecka J. (2008). *Polski singiel: obraz w mediach a autowizerunek* W: Malinowska E. (red.), *Stereotypy a rzeczywistość*, Łódź: Oficyna Wydawnicza „Tercja".

[6] Pomimo utrzymującej się różnicy w poziomie wynagrodzeń (przeciętna płaca kobiet stanowi średnio 80-82% wynagrodzenia mężczyzny), zróżnicowanie płac ze względu na płeć maleje. Biorąc pod uwagę fakt, że kobiety w Polsce dążą do wiązania się z mężczyznami o wyższym lub podobnym statusie społecznym (tzw. hipergamia lub homogamia małżeńska), a mężczyźni nie są skłonni do wiązania się z kobietami znajdującymi się wyżej w hierarchii społecznej, czynniki te prowadzą do nierównowagi na rynku małżeńskim (Czernecka 2010; Grzeszczyk 2005; Paprzycka 2008).

Czernecka J. (2010). *Społeczne przyczyny egzystencji w pojedynkę oraz styl życia łódzkich i warszawskich singli*. Praca doktorska. Uniwersytet Łódzki.

Czernecka J. (2011). *Wielkomiejscy single.* Warszawa: Poltext.

Domagalska K. (2015). *Nie przeproszę, że urodziłam. Historie rodzin z „in vitro"*. Wołowiec: Wydawnictwo Czarne.

Grzeszczyk E. (2005). *Pojedyncze, profesjonalistki… 'Single Professional Women' w Polsce i na świecie.* Kultura i Społeczeństwo, XLIX (2), s. 199–225.

GUS 2003. *Gospodarstwa domowe i rodziny* 2002. Narodowy Spis Powszechny Ludności i Mieszkań, Powszechny Spis Rolny. Warszawa.

GUS 2013. *Gospodarstwa domowe w* 2011 *roku – wyniki spisu ludności i mieszkań* 2011. Główny Urząd Statystyczny, Warszawa.

Kotowska I. (red.) (2014). *Niska dzietność w Polsce w kontekście percepcji Polaków*. Diagnoza Społeczna 2013. Raport tematyczny, http://www.diagnoza.com/pliki/raporty_tematyczne/Niska_dzietnosc_w_Polsce.pdf

Paprzycka E. (2008). *Kobiety żyjące w pojedynkę. Między wyborem a przymusem*. Warszawa: Wydawnictwo Akademickie „Żak".

Szpakowska M. (2006). *Antropologiczne przesłanki samotności.* W: *Kobieta i rewolucja obyczajowa. Społeczno-kulturowe aspekty seksualności. Wiek* XIX *i* XX. Zbiór studiów pod redakcją Anny Żarnowskiej i Andrzeja Szwarca. Tom IX, Warszawa: Wydawnictwo DIG, s. 439–446.

Ubóstwo w Polsce w latach 2013 i 2014. Warszawa: Główny Urząd Statystyczny.

Żurek A. (2011). *Czas wolny singli – sfera życia codziennego podlegająca procesom upłciowienia.* W: Mroczkowska D. (red.), *Czas wolny. Refleksje, dylematy, perspektywy.* Warszawa: Difin, s. 224–235.

Żurek A. (2008a). *Single. Żyjąc w pojedynkę.* Poznań: Wydawnictwo Naukowe UAM.

Żurek A. (2008b). *Czas wolny singli – pomiędzy rodzina a przyjaciółmi.* W: W. Muszyński W., Sikora E. (red.), *Miłość, wierność i uczciwość na rozstajach współczesności. Kształty rodziny współczesnej.* Toruń: Wydawnictwo Adam Marszałek, s. 427–441.

Żurek A. (2005). *Relacje łączące młodych singli z ich rodzicami*. W: Wrzesień W., Żurek A., Przybył L. (red.), *Rodzice i ich dzieci w tańcu pokoleń.* Poznań: Wydawnictwo Naukowe UAM, s. 71–122.

Żurek A. (2004). *Warunki udanego życia w oczach osób samotnych życiowo.* W: Golka M. (red.), *W cywilizacji konsumpcyjnej*. Poznań: Wydawnictwo Naukowe UAM, s. 109–127.

FOLIA 202

Annales Universitatis Paedagogicae Cracoviensis

Studia Sociologica VIII (2016), vol. 1, p. 198–203
ISSN 2081–6642

Justyna Tomczyk
Uniwersytet Pedagogiczny im. KEN w Krakowie

Gender w PRL: droga ku (nie)udanej emancypacji kobiet

Małgorzata Fidelis, *Kobiety, komunizm i industrializacja w powojennej Polsce*, tłum. M. Jaszczurowska, Wydawnictwo W.A.B., Warszawa 2015, 366 s.

W 2015 roku na krajowym rynku wydawniczym pojawiła się książka Małgorzaty Fidelis, zatytułowana *Kobiety, komunizm i industrializacja w powojennej Polsce.* Jest to przekład na język polski opublikowanej przez nią wcześniej monografii *Women, Communism, and Industrialization In Postwar Poland* (New York 2010: Cambridge University Press), stanowiącej zmienioną wersję jej rozprawy doktorskiej, obronionej w Stanford University. Autorka należy do znawców społecznych dziejów polskich kobiet w czasach komunizmu. Jej pionierskie opracowanie bardzo wzbogaciło historiografię tego okresu. Obecnie, jako profesor historii i wykładowczyni na Uniwersytecie Illinois, kontynuuje badania w obszarze feminizmu oraz *gender studies*, koncentrując uwagę na sytuacji społecznej kobiet w Europie Środowo-Wschodniej w okresie powojennym.

Omawiana książka to opracowanie potrzebne i ważne. Zasługuje na uznanie i wyróżnienie przede wszystkim dlatego, że wnosi istotny wkład do dotychczasowych ustaleń na temat komunizmu – jako projektu zmiany dziejowej. Istnieje już bogata literatura poświęcona minionemu systemowi politycznemu i aktorom, którzy ten system budowali i przebudowywali, dostrzegając w nim podwaliny nowego porządku społecznego w skali globalnej. Fidelis zajęła się w swojej pracy kobietami jako szczególną kategorią społeczno-kulturową, której rola w tamtych czasach, z punktu widzenia ogólnej analizy systemu komunistycznego, może się wydawać marginalna, ale jak przekonuje autorka, przeczy to faktom. Podejmując wątki standardowo uwzględniane w pracach dotyczących doktryny, ideologii czy propagandy komunistycznej, autorka znacznie wykroczyła poza nie lub uzupełniła je o nieznane dotąd treści. Wartością, jaką ta książka dodaje do dotychczasowych ujęć tej problematyki, jest udana próba wyeksponowania i szerszego omówienia tematu równouprawnienia i równości płci, ujawnienia stereotypów, uproszczeń i mitów, jakie zostały wykreowane i funkcjonowały na gruncie tamtego ustroju. Rzecz idzie nie tyle o propagowane przez socjalizm hasła emancypacyjne, co o niebezpieczną metodę oddziaływania na społeczeństwo za pośrednictwem polityki płci, mającej przyczynić się do uzyskania bezwzględnego poparcia dla władzy.

Niekwestionowaną zaletą książki jest wybór ambitnego i oryginalnego tematu badań. Opracowanie dotyczy rzadko podejmowanej, by nie powiedzieć lekceważonej, problematyki losów polskich kobiet w latach 1945–1989. Kobiety w dobie komunizmu nie były popularnym ani szczególnie zajmującym obiektem badań, stąd niewiele wiadomo o pełnionych przez nie podówczas rolach społecznych, zajmowanych pozycjach, aktywności publicznej, podejmowanych interakcjach i stosunkach społecznych, uczestnictwie w ruchach obywatelskich. Przyczyn tego stanu rzeczy należy upatrywać w niechęci wobec feminizmu i powszechnie szerzonej wówczas opinii, iż jest to ideologia niepoważna i niepotrzebna, a nawet szkodliwa, opozycyjna, wrogo nastawiona wobec systemu. Tym samym problematyka kobiet w PRL została przemilczana – można nawet odnieść wrażenie, że kobiety nie brały udziału w tworzeniu struktury i kultury społeczeństwa socjalistycznego. Pomijanie, marginalizowanie i deprecjonowanie znaczenia kobiet w historii powojennej Polski przejawia się przede wszystkim w nieuwzględnianiu kobiecych sylwetek, biografii, dokonań i sukcesów w diagnozach tamtej rzeczywistości. Prostą tego konsekwencją jest fakt, że kobiety niemal w ogóle nie pojawiały się na kartach podręczników, jakby były nieobecne. Dopiero okres III RP rozbudził potrzebę dyskutowania spraw kobiet oraz problemów, które ich dotyczą.

Mając świadomość tych braków i zafałszowań, autorka stara się przedstawić kobiety jako bohaterki dnia codziennego, widzi w nich samodzielny, autonomiczny podmiot działań, obdarzony pewną choć ograniczoną mocą sprawczą, nie zaś tylko przedmiot poddający się biernie władzy partii komunistycznej. Owa kobieca podmiotowość i sprawstwo, wyrażające się w wywieraniu nacisku na struktury państwowe, nieposłuszeństwie obywatelskim lub biernym oporze, stanowiły realną przeciwwagę dla odgórnych postanowień i decyzji. Autorka w klarowny i skrupulatny sposób, posługując się potoczystym stylem oraz przystępnym dla czytelnika językiem, przedstawia dość skomplikowane losy robotnic pochodzących z różnych terenów Polski (włókniarek z Żyrardowa, pracownic zambrowskiej fabryki bawełny, górniczek początkowo pracujących tylko na powierzchni, a od 1951 r. także pod ziemią). Na takich przykładach pokazuje, jak industrializacja kraju wpływała na życie codzienne kobiet, zwłaszcza na przestrzeń intymną, stosunki rodzinne, ewentualne decyzje o zamążpójściu czy macierzyństwie.

Interpretacja jest prowadzona z perspektywy genderowej, co oznacza, że pojęcie płci kulturowej stanowi dla autorki jako układ odniesienia, gdy opisuje przeżycia kobiet, ich doświadczenia i ścieżki życiowe. Aby kompleksowo przedstawić tzw. kwestię kobiecą, Fidelis dość precyzyjnie sformułowała dwa zasadnicze problemy badawcze. Po pierwsze, postawiła pytanie o to jak rozumiano płeć kulturową w komunizmie jako systemie ideologicznym afirmującym i postulującym równość płci. Po drugie, chciała poznać jaki wpływ na życie codzienne kobiet miała komunistyczna ideologia płci. Odpowiadając na te pytania, autorka wykorzystała teoretyczną podbudowę *gender studies,* co okazało się ciekawym zabiegiem eksplikacyjnym, przynoszącym odkrywcze odpowiedzi.

Objętościowo książka wydaje się typowa dla prac z pogranicza socjologii, politologii i historii, liczy 366 stron. Składa się z siedmiu odrębnych, choć kontekstowo powiązanych ze sobą rozdziałów tematycznych, obszernej bibliografii, indeksu nazwisk i wykazu ilustracji. Taka struktura sprzyja kierunkowemu rozwijaniu narracji – od podania ogólnych informacji na temat miejsca i roli kobiet w budowaniu państwa socjalistycznego (ze szczególnym uwzględnieniem „produktywizacji kobiet"), poprzez przedstawienie podejścia władz do równouprawnienia płci, aż do analizy dominującego w okresie postalinowskim dyskursu, w którym na czoło wysunęły się kategorie seksualności, cielesności, reprodukcji i macierzyństwa. Chociaż w warstwie krytycznej dyskurs ten cechowała pewna zmienność, to jego przesłanie było dość jednoznaczne. Kobiety były adresatkami propagandy komunistycznej nastawionej na wspomaganie procesu przejmowania radzieckich wzorców kulturowych.

Z książki polsko-amerykańskiej badaczki czytelnik dowiaduje się także, jak kobiety w Polsce odbierały ten przekaz, jak postrzegały politykę państwa, oraz jak reagowały na zarządzenia i posunięcia polityczne, mające wpływ na ich życie. Chodzi tu przede wszystkim o subiektywne przeżycia oraz osobiste oceny otaczającej je rzeczywistości, pełnej paradoksów, niejasności czy wręcz sprzeczności pomiędzy głoszonymi hasłami a ich praktyczną realizacją. Wywód jest spójny, logiczny i przejrzysty – jak można przypuszczać, dobrze oddaje tok myślenia autorki i zamysł kompozycyjny całości dzieła.

Rzetelnie podchodząc do tematu i konsekwentnie realizując swe cele badawcze, autorka dowiodła swoją książką, że w stopniu najwyższym opanowała warsztat badawczy, jak również posiada szeroką wiedzę w danym obszarze oraz nieprzeciętne umiejętności pisarskie. Dodatkowym walorem książki jest jej zaplecze faktograficzne z uwzględnieniem wielu nietypowych faktów i okoliczności.

Praca opiera się na pogłębionych wywiadach ze świadkami ówczesnych wydarzeń oraz istotnych i oryginalnych źródłach zastanych: dokumentach pochodzących z archiwów partii, tajnych i wcześniej niedostępnych dla opinii publicznej kartotekach policyjnych, listach czytelników i czytelniczek do gazet.

Przyjęty przez autorkę azymut metodologiczny odwołuje się do zainicjowanej przez Wiliama Thomasa i Floriana Znanieckiego metody biograficznej, niezwykle użytecznej w analizie ludzkich działań, ułatwiającej poznanie tzw. miękkich uwarunkowań postępowania jednostek (emocji, motywacji, przekonań, doznań i przeżyć wewnętrznych). Uzyskujemy dzięki temu wiarygodny obraz powojennej rzeczywistości – widzianej oczami jej uczestników i przez nich samych odbieranej i interpretowanej. Dotarcie do materiałów archiwalnych, systematyczny przegląd zasobów źródłowych oraz wyważony ich dobór był zadaniem niełatwym, żmudnym i czasochłonnym, co tym bardziej podnosi rangę opracowania. Zebrany przez autorkę materiał empiryczny zainspirował ją do konstruowania ogólnych modeli wyjaśniających, mogących służyć dla dalszej eksploracji badanego zjawiska lub zjawisk pokrewnych.

Książka obfituje w istotne spostrzeżenia i rozstrzygnięcia. Najważniejsze z nich, moim zdaniem, dotyczy wzajemnej, bardzo silnej interferencji sfer polityki i płci

w powojennej Polsce. Oba te dyskursy pozostawały w dialektycznym związku i intensywnie oddziaływały na siebie, co uwidaczniało się w sposobie, w jaki komunizm definiował kobiecość, męskość, emancypację i równouprawnienie. Wiązanie polityki z płcią (i odwrotnie) było ze strony decydentów zabiegiem przemyślanym, celowym, strategicznym – stanowiło skuteczną taktykę kontroli normatywnej. Jak słusznie zauważa autorka, ustawodawstwo dotyczące płci stało się narzędziem manipulacji i wpływu, wykorzystywanym do wyznaczania politycznego kursu oraz legitymizowania systemu władzy. Koncepcja równości płci była ustawicznie redefiniowana w zależności od nastrojów społecznych, potrzeb ekonomicznych, obrazów rzeczywistości funkcjonujących w potocznej świadomości (puste półki sklepowe, deficyt towarów, siermiężność warunków życia). Definicje płci opierały się więc na antynomiach – z jednej strony w dyskursie oficjalnej propagandy głoszono równość między dwiema płciami (równy dostęp do dóbr, stanowisk, tytułów, zawodów, awansów), a z drugiej wzmacniano – na potrzeby utrzymania dotychczasowego ładu społecznego – stereotypowe i tradycjonalistyczne wizerunki kobiety i mężczyzny. Można w tym kontekście mówić o paradoksie równouprawnienia, który polegał na tym, że wraz z postępującym zanikiem równości jako powszechnie uznanej i obowiązującej praktyki, rosło poparcie dla działań pozorowanych mających tę równość ustanawiać.

Jako egzemplifikację tej tezy, Fidelis przypomina główne założenie komunistów, głoszące iż praca zawodowa to najskuteczniejsza metoda emancypacji i równouprawnienia. Skrajnie upolitycznionа strategia aktywizacji zawodowej kobiet, prowadzonej na masową skalę (bez względu na wykształcenie, status majątkowy, stan cywilny, pochodzenie klasowe itd.), pojawiła się już w latach 1948-56. U jej podstaw stała ideologia głosząca równość wszystkich obywateli w dostępie do pracy i zatrudnienia, nawet jeśli miałyby to być obszary do tej pory zmaskulinizowane, jak np. fabryki, huty, kopalnie. Zjawisko to sugestywnie przedstawione zostało w podrozdziale zatytułowanym *Praca pod ziemią*, w którym autorka przedstawia zależności między pracą kobiet pod ziemią a kulturą i etosem górniczym. Dobrze, że ten wątek został w książce podjęty, bowiem rzadko jest on rozwijany w pracach naukowych, co z kolei powoduje wyparcie ze świadomości zbiorowej sylwetki górniczki – kobiety urzeczonej atmosferą kopalni i w pełni oddanej wykonywanej pracy. Z opisu pracy górniczek, stanowiącego apologię pracy kobiet w zawodach typowo „męskich", przebija jakiś nierealistyczny entuzjazm; wiadomo, że praca pod ziemią była trudna i wycieńczająca (zwłaszcza dla słabszych fizycznie kobiet), zaś z jej wykonywaniem wiązało się ogromne niebezpieczeństwo i ryzyko. W okresie odwilży, wraz z pojawieniem się nowego politycznego kursu, praca produkcyjna kobiet była w coraz mniejszym stopniu apoteozowana – jej miejsce zajęły małżeństwo i rodzina, promowane jako integralne elementy kobiecej tożsamości.

Wśród licznych interesujących wniosków, diagnoz i zestawień przyciąga uwagę analiza porównawcza definicji równości płci propagowanych przez trzech zinstytucjonalizowanych aktorów społecznych: państwo, Kościół katolicki, organizacje kobiece. W tym kontekście Fidelis zasadnie odwołuje się do teorii „nowego

matriarchatu" zbudowanej na traumatycznym doświadczeniu wojny, która zdezorganizowała stosunki społeczne. W efekcie zmieniła się struktura demograficzna społeczeństwa – mężczyźni walczyli i ginęli na frontach, zaś kobiety przejęły ich funkcje w gospodarstwie domowym, a z czasem z powodzeniem zajęły domeny i pola instytucjonalne dotychczas zarezerwowane dla mężczyzn. Kobiety, podejmując pracę zarobkową, zakwestionowały (być może w części obaliły) stereotyp płci. Ich obecność i coraz większa widoczność w sferze publicznej skutkowała uzyskaniem dodatkowych praw obywatelskich oraz zwiększoną możnością aktywnego uczestnictwa w życiu politycznym.

Rewolucja płci, niejako samoczynnie dokonująca się na skutek wojny, stała się odniesieniem dla socjalistycznej idei równości. Idea ta opierała się na utopiach Marksa, Engelsa, Lenina, w świetle których ucieleśnieniem równych praw stała się robotnica zatrudniona w zakładzie przemysłowym. Na ten podstawie powstał spopularyzowany wizerunek „nowej kobiety" – realizującej się w pracy zawodowej podejmowanej na rzecz społeczeństwa socjalistycznego. Sugestywnym, alegorycznym, osobowym wypełnieniem tego wizerunku stały się przodownice pracy, m.in. znana z propagandowych obrazów Aleksandra Kobzdeja ceglarka i Magdalena Figur, pierwsza peerelowska traktorzystka.

Fidelis słusznie zauważa, iż proponowana przez władze komunistyczne reorganizacja ról płciowych nie podważyła tradycyjnego podziału na role męskie i żeńskie. Jak argumentuje, główną tego przyczyną był mocny i jednoznaczny głos Kościoła katolickiego, który stał na straży wartości chrześcijańskich oraz odwoływał się do kultywowanej przed wojną tradycji narodowej opartej na mitologii wyzwoleńczej bądź martyrologicznej. Widoczne było przywiązanie do postaci Matki-Polki i instytucji domu rodzinnego, który miał uczyć miłości ojczyzny. Choć teza o przemożnej roli kościoła wydaje się przekonywająca, to nie została przez autorkę wystarczająco uzasadniona. Warto byłoby ją uprawomocnić przez powołanie się na dodatkowe źródła lub wyniki badań empirycznych.

Opisując wpływ Kościoła katolickiego na sytuację kobiet w Polsce, Fidelis wskazuje na dość intrygującą zależność. Otóż zauważa, iż początkowo Kościół nie popierał idei podejmowania przez kobiety pracy zawodowej, stopniowo jednak zaaprobował ten pomysł, a nawet powołał specjalne stowarzyszenia, których celem była ochrona moralności robotnic. Zdaniem autorki była to planowa strategia mająca na celu uzyskanie od władz obietnicy zaostrzenia przepisów antyaborcyjnych lub nawet całkowitego zakazu i kryminalizacji przerywania ciąży. Teza ta, chociaż w książce nie została poparta odesłaniem do źródeł lub opracowań, wydaje się interesująca i w moim przekonaniu zasługuje na dokładniejsze zbadanie. Sygnalizowany alians państwa i Kościoła również wydaje się twierdzeniem dość nowatorskim, jako że przeczy obiegowej opinii o braku porozumienia lub otwartym konflikcie między tymi stronami.

Za interesujące należy także uznać porównanie wizerunków płci lansowanych przez aktywistki PPS i PPR. Fidelis pokazuje, że obóz socjalistów niemal od momentu

powstania (w roku 1919) w jego ramach frakcji kobiecej popierał zatrudnianie kobiet w różnych zawodach, ochronę praw pracowniczych robotnic, wsparcie dla rodzin pracujących kobiet, swobodę obyczajową, wreszcie powszechny dostęp do antykoncepcji. Natomiast PPR, jako partia nowa bez dłuższej tradycji politycznej, bazowała na wzorach radzieckich i głosiła konieczność „ideologicznego uświadomienia" kobiet i wprowadzaniu ich w szeregi partii. Liderki obu frakcji (tworzących później scentralizowany Wydział Kobiecy PZPR) usiłowały zaszczepić w społeczeństwie myśl feministyczną i emancypacyjną, organizowały strajki i manifestacje, pisały petycje do partii, wysuwając konkretne żądania dotyczące dostępności dóbr konsumpcyjnych i przeciwstawiając się stalinowskiemu modelowi uprzemysłowienia. Ich głos w sferze publicznej był na tyle słyszalny, że władza obawiała się dalszych reperkusji. Negowała więc znaczenie tego głosu, twierdząc, iż kobiety nie są poważnym partnerem do rozmów – zachowują się histerycznie, agresywnie, terroryzują otoczenie. W zamierzeniu ów osąd miał pozbawić kobiety wiarygodności i osłabić ich aktywność. O istnieniu tego rodzaju napięć wewnątrz szeroko rozumianego obozu opowiadającego się za nowym ustrojem mało było wiadomo, dlatego raz jeszcze wypada podkreślić innowacyjność monografii, w której kobiety zostały przedstawione jako zbiorowy adresat polityki państwa i partii, ale także jako podmiot i kreator doniosłych przemian społecznych.

Recenzowana książka, by użyć znanego określenia Maxa Webera, „odczarowuje świat", przede wszystkim przez to, że w oparciu o racjonalne argumenty rozprawia się z szeregiem mitów i ideologii stworzonych przez decydentów na potrzeby utrzymania władzy. Mitologie te brały swój początek w ZSRR, ich przenoszenie na grunt polski było odbierane było często jako zabieg dość groteskowy. Książka Małgorzaty Fidelis pokazuje, jak chaotyczne działania władz skutkowały dezorganizacją i w sumie odbiły się rykoszetem i zaszkodziły inicjatorom „polityki płci". Jest to zatem jedna z tych prac badawczych, które skłaniają do poważnej debaty na temat roli kobiet w historii i rewizji pewnych sądów, często stereotypowych i nader uproszczonych, zakotwiczonych w świadomości społecznej. Książka jest ważna, nie tylko dlatego, że stanowi wiarygodny opis sytuacji, w jakiej znajdowały się kobiety w powojennej Polsce, lecz i z tego względu, że wskazuje także czynniki strukturalne i kulturowe, które decydowały o trwałości lub zmienności wzorów płci. Wiele z tych czynników nadal działa w polskim, ponowoczesnym i zglobalizowanym społeczeństwie, o którym nieco na wyrost mówi się, że jest wyemancypowane.

FOLIA 202

Annales Universitatis Paedagogicae Cracoviensis

Studia Sociologica VIII (2016), vol. 1, p. 204–207
ISSN 2081–6642

SPRAWOZDANIA/REPORTS

Grzegorz Kubiński
Uniwersytet Pedagogiczny im. KEN w Krakowie

Sprawozdanie z konferencji naukowej „Islam w badaniach i praktyce kontaktów międzykulturowych"

Obecna sytuacja geopolityczna w regionie Bliskiego Wschodu bardzo silnie wpływa na zachowania społeczne na świecie, zwłaszcza w Europie. Odmienność kulturowa, społeczna i religijna krajów arabskich postrzegana z perspektywy europejskiej od dawna jest przedmiotem dyskusji w ramach szeroko pojętych nauk humanistycznych i społecznych. Daje się zauważyć wzrost zainteresowania tym tematem także w codziennym dyskursie publicznym. Nierzadko kwestie kojarzone z islamem oraz światem arabskim prezentowane są powierzchownie, a bywa że i opacznie. Rolą nauki jest nie tylko badanie, analizowanie i opis zachowań społecznych, ale także proponowanie racjonalnych argumentów w sporach angażujących duże grupy ludzkie. Dyskusja naukowa między badaczami, podejmującymi rzetelny i racjonalny namysł nad problemami żywo dyskutowanymi w przestrzeni publicznej, daje każdemu z nich możność poznania opinii kolegów i przemyślenia na nowo własnych poglądów, co powinno prowadzić wszystkich do wyartykułowania swoich sądów w sposób bardziej wyważony.

Taką właśnie próbą zmierzenia się z niełatwym tematem obecności i roli islamu we współczesnym świecie była konferencja naukowa *Islam w badaniach i praktyce kontaktów międzykulturowych*, która odbyła się w dniach 9–10 marca 2016 roku w murach Uniwersytetu Pedagogicznego im. KEN w Krakowie. Inicjowała i organizowała ją Katedra Socjologii Religii Instytutu Filozofii i Socjologii UP. Referaty wygłosili prelegenci reprezentujący takie ośrodki akademickie, jak: Uniwersytet w Poczdamie, Ukraińska Akademia Nauk, Uniwersytet Kardynała Stefana Wyszyńskiego w Warszawie, Uniwersytet Ekonomiczny w Krakowie, Uniwersytet Jana Pawła II w Krakowie, Uniwersytet Jagielloński oraz Uniwersytet Pedagogiczny w Krakowie.

Konferencja ta stała się okazją do dyskusji nad wieloma aspektami relacji świata zachodniego z cywilizacją muzułmańską. Obecna sytuacja międzynarodowa, to temat narzucającym się w oczywisty sposób, z jej dwoma głównymi elementami, bardzo silnie wpływającymi na masową wyobraźnię – terroryzmem związanym z działaniami Państwa Islamskiego i innych organizacji oraz migracjami ludności z krajów Bliskiego Wschodu objętych konfliktami zbrojnymi. Organizatorom zależało jednak na tym, by debata toczyła się na szerszej płaszczyźnie analizy kontaktów

międzykulturowych, by społeczność akademicka, biorąc na siebie intelektualną odpowiedzialność za wypowiedzi także w sprawach budzących wiele emocji, nie uległa presji formułowania doraźnych, uproszczonych ocen. Istotnym zamierzeniem uczestników konferencji – a także licznej i biorącej udział w dyskusji publiczności – było uniknięcie postrzegania islamu oraz społeczności muzułmańskiej przez pryzmat zbrodni, niepokojów społecznych i destabilizacji porządku społecznego.

Uczestnicy dwudniowego spotkania w swoich wystąpieniach oraz dyskusjach starali się pokazać świat islamu jako obszar funkcjonowania potężnej kultury, oddziałującej na całe społeczeństwo światowe, również pozytywnie oraz dynamizująco. Badacze biorący udział w spotkaniu skoncentrowali się na analizie islamu jako systemu religijnego odgrywającego istotną rolę we współczesnym świecie. Inaczej niż to się z reguły dzieje w zubożonym dyskursie medialnym, ukazywali te aspekty islamu, które pozwalają lepiej zrozumieć kulturę świata muzułmańskiego. Doskonale pokazały to wystąpienia dwóch prelegentów, imama Kadira Sanci z Uniwersytetu w Poczdamie oraz prof. Eugeniusza Sakowicza z Uniwersytetu Kardynała Stefana Wyszyńskiego w Warszawie. Pierwszy z nich przedstawił interesujący projekt świątyni ekumenicznej w Berlinie, nazwanej „House of One", pomyślanej jako ośrodek kultury – miejsce, gdzie wyznawcy trzech religii (judaizmu, chrześcijaństwa, islamu) mogliby się spotykać w duchu wspólnoty, wzajemnego zrozumienia i współpracy. O tych wartościach mówił także prof. Eugeniusz Sakowicz, analizując szanse dialogu Kościoła katolickiego ze światem islamu. W obu tych wystąpieniach dostrzeżona została możliwość oraz konieczność współdziałania między obiema religiami.

Referaty kolejnych prelegentów, dr. Olega Yarosha z Ukraińskiej Akademii Nauk, prof. Katarzyny Warmińskiej z Uniwersytetu Ekonomicznego w Krakowie, prof. Piotra Stawińskiego z Uniwersytetu Pedagogicznego w Krakowie, dr. Sławomira Cebuli z Uniwersytetu Jana Pawła II w Krakowie oraz mgr. Pawła Szuppego z Uniwersytetu Kardynała Stefana Wyszyńskiego w Warszawie, dotyczyły bardziej szczegółowych zagadnień nadających się do debaty na płaszczyźnie dialogu międzykulturowego. Przedmiotem rozważań była możliwość i konieczność integracji, ale także oczywiste trudności, na jakie napotyka koegzystencja obu społeczności i wyznań, również na terenie Polski. Przedstawiono także komunikaty z badań terenowych (socjologiczno-etnograficznych), zawierające interesujące, ożywcze analizy zachowań obu społeczności w tak trudnej geopolitycznie sytuacji. Bardziej teoretyczny charakter miały referaty dr Marii Rogińskiej, dr Justyny Tomczyk, dr. Michała Warchali (wszyscy troje reprezentowali Uniwersytet Pedagogiczny w Krakowie) oraz dr Elżbiety Wnuk-Lisowskiej z Uniwersytetu Jagiellońskiego. Tematy ich wystąpień ogniskowały się wokół elementów intelektualnej kultury islamu, stanowiących wkład tej kultury do nauki i filozofię w świecie zachodnim. Nie unikano podczas dyskusji także zagadnień drażliwych, dotyczących problematyki „islamskiego terroryzmu". Temat zagrożeń płynących ze strony fundamentalistycznych ruchów islamskich, żywo dyskutowany, pojawił się w referacie dr Doroty Czakon-Tralski z Wyższej Szkoły

Ekonomiczno-Humanistycznej w Bielsku-Białej oraz dr. Bogdana Pliszki, reprezentującego Politechnikę Śląską w Gliwicach.

Dzięki zaangażowaniu zaproszonych gości, prelegentów oraz publiczności, krakowska konferencja okazała się wydarzeniem niezwykle produktywnym, różnorodnym, a przede wszystkim intelektualnie stymulującym. Owocem tego spotkania badaczy, których połączył także udział w twórczej, rzetelnej, naukowej dyskusji, będzie monografia wieloautorska przedstawiającą omawianą problematykę, która ukaże się jesienią roku 2016. Jej zróżnicowana tematyka będzie odzwierciedlać zainteresowania i osiągnięcia naukowe autorów, reprezentujących różne dyscypliny – religioznawstwo, socjologię, politologię, prawo i teologie.

Prace zebrane w tym tomie mieścić się będą się w dwu blokach. W części pierwszej, „teoretycznej", znajdą się studia nad religią, kulturą i cywilizacją islamu, w przeszłości i współcześnie, prace o charakterze bardziej akademickim; w części drugiej – prace poświęcone „praktycznym" aspektom kontaktów międzykulturowych. W części pierwszej przedstawione zostaną rozważania dotyczące aktualności wcześniejszych ustaleń i propozycji interpretacyjnych, przy czym tematem głównym będzie metodologiczna refleksja i namysł nad przeszłym i obecnym miejscem nauki w świecie islamu. Ta część ukaże także zróżnicowanie instytucjonalne, dorobek i bogactwo zainteresowań współczesnych ośrodków uniwersyteckich w Polsce, w których prowadzone są badania nad Orientem.

W drugiej części pojawi się całe spektrum zagadnień związanych z cywilizacją muzułmańską oraz relacjami Wschód–Zachód, zarówno w rodzimym wymiarze kultury polskich Tatarów, jak i w ogólniejszej perspektywie islamu jako światowego systemu religijnego. Znajdą się tu też prace poświęcone edukacji w islamie, tradycji mistycznej (sufizm), dialogowi międzyreligijnemu, kulturze popularnej, feminizmowi oraz – dramatycznie aktualnemu – zagrożeniu terroryzmem.

•

Na koniec, dla celów dokumentacyjnych, przedstawiam zestawienie wszystkich referatów wygłoszonych 9–10 marca 2016 w Krakowie na konferencji naukowej, której tematem – przypomnijmy – był *Islam w badaniach i praktyce kontaktów międzykulturowych.*

1. Sławomir Cebula (UJPII), *Tatarzy z Podlasia. Sprawozdanie z badań terenowych.*
2. Dorota Czakon-Tralski (WSEH), Bogdan Pliszka (PŚ), *W cieniu paryskich zamachów czyli studenci o poczuciu zagrożenia zamachami.*
3. Maria Rogińska (UP), *Czy możliwa jest 'islamska nauka'? Kolejna odsłona sporu Wschodu z Zachodem.*
4. Eugeniusz Sakowicz (UKSW) *Dialog teologiczny z islamem? Perspektywa Kościoła katolickiego.*
5. Piotr Stawiński (UP), *Społeczny wymiar islamu.*
6. Paweł Szuppe (UKSW), *Islam w badaniach współczesnych ośrodków uniwersyteckich w Polsce.*

7. Imam Kadir Sanci (Universität Potsdam), *Islam and Education* oraz *'The House of One' – the Idea and Main Goals of the Project.*
8. Justyna Tomczyk (UP), *Feminizm w islamie – nowy paradygmat ruchu społecznego.*
9. Michał Warchala (UP), *Max Weber o islamie – próba rekonstrukcji.*
10. Katarzyna Warmińska (UE w Krakowie), *Islam a etniczność na przykładzie Tatarów polskich.*
11. Elżbieta Wnuk-Lisowska (UJ), *Mistrzowie wyobraźni. Toshihiko Izutsu i Henry Corbin o sufizmie.*
12. Oleg Yarosh (Ukraińska Akademia Nauk), *Transnarodowe ruchy muzułmańskie a globalizacja: metodologiczne zasady badań.*

Authors/Autorzy

Contributors to *Selected Topics in Experimental Social Science*

Ann Converse Shelly (PhD from Michigan State University, 1973) is Professor Emerita of Educational Foundations at College of Education, Ashland University, Ohio. She is a specialist in teacher education, educational curricula, assessment, and status in the classroom.

Szymon Czarnik received his PhD from the Jagiellonian University (2007). He has since then been working as an assistant professor for the Institute of Sociology at that university. His PhD thesis, which was based on an experimental study of redistributive behaviour, as well as earlier publications reveal the author's interest in game theory. The areas of his later research activity include methods of social research, statistical data analysis, and social aspects of the labour market. In 2010–2015 he was a core member of the research team conducting the Study of Human Capital in Poland.

Iza Desperak (PhD from the University of Lodz, 2000) is a sociologist specializing in qualitative research. She works for the Chair of Sociology of Politics and Morality at the Institute of Sociology at the University of Lodz. Her interests focus on social inequalities, the role of stereotypes, and various dilemmas met in research work, from gender identification of interviewees to ethical responsibility of social auditioning.

Martha Foschi is Professor Emerita at the University of British Columbia, Canada. After majoring in philosophy at the University of Buenos Aires and completing a PhD in sociology at Stanford University, she joined the faculty of UBC's Department of Sociology in 1967. Her achievements in the areas of group processes, theory construction, and experimental research have been recognized by her Department, which in 2014 established the Martha Foschi Award for Excellence in Research and Teaching, to be conferred bi-annually to a UBC faculty member. Her most recent publication reporting on experimental work is: Foschi M., Valenzuela J. (2015) *Choosing Between Two Semi-Finalists: On Academic Performance Gap, Sex Category, and Decision Question.* Social Science Research.

Zbigniew Karpiński is an assistant professor at the Institute of Philosophy and Sociology at the Polish Academy of Sciences where he completed postgraduate studies (following his MA in sociology from the Jagiellonian University) with a PhD in sociology (2008). His most recent publications focus on trust and cooperation, mathematical and statistical modelling of tie formation processes, perceptions of distributions of earnings, and status and legitimacy processes. His research papers appeared in Sociological Theory, International Journal of Sociology, ASK (Polish journal specialising in publishing articles in methods of social research), and other journals.

Marcel Kotkowski graduated from the Jagiellonian University with an MA in sociology (2012). He is currently a PhD student affiliated with the Institute of Sociology at that university. His interests include human-technology interaction, kinetic typography, and design.

Jane Sell (PhD from Washington State University, 1979) is a Professor and Department Head at the Department of Sociology, Texas A&M University. In her recent publications, she has combined her competence in experimental method and the perspective of structural social psychology with an interest in studying how gender differences affect social interaction.

Robert K. Shelly (PhD from Michigan State University, 1972) is Professor Emeritus at the Department of Sociology and Anthropology, Ohio University. His areas of expertise include social psychology, group processes, and mathematical models.

Tadeusz Sozański received his PhD in sociology from the Jagiellonian University in 1982 and since then he has worked for that university until 2006, when he became affiliated with the Institute of Philosophy and Sociology at Pedagogical University of Cracow. Throughout his academic career, his research interests focused on the use of mathematics in the social sciences (his two most recent articles of importance in this area are: *The Conception of Blocking Power as a Key to the Understanding of the History of Designing Voting Systems for the EU Council. Decyzje*, 2014; *On the Core of Characteristic Function Games Associated with Exchange Networks. Social Networks*, 2006). However, his habilitation book (*Społeczne i wspólne. Studium socjologiczno-filologiczne*, 2013, in Polish; English title: 'The Social and the Common. A Sociologico-Philological Study') is the product of the author's excursion into pure humanities.

Murray Webster, Jr. (PhD from Stanford University, 1968) is a Professor of Sociology at the Department of Sociology, University of North Carolina at Charlotte. He is known for his expertise in mathematical sociology, quantitative methods, social psychology, social theory, and methodology of experimenting. He co-edited (with Jane Sell) *Laboratory Experiments in the Social Sciences* (2007, 2nd edition 2014). Webster is the 2015 winner of the Cooley-Mead Award for Distinguished Scholarship given annually by the Social Psychology Section of the American Sociological Association.

Kinga Wysieńska-Di Carlo holds MAs in sociology and political science, and a PhD in sociology (2005), all from the Jagiellonian University. She is currently an assistant professor at the Institute of Philosophy and Sociology at the Polish Academy of Sciences, and a research fellow at the Albert Shanker Institute, Washington, D.C., USA. In 2012–2015 she served as a member of the Migration Policy Expert Team at the Chancellery of the President of Poland. Her research interests include status and legitimacy processes, social stratification, international migration, and discrimination. She is the author of a number of discrimination studies using situation testing and laboratory experiments in Poland, as well as studies on immigrant flows to Poland and the immigrant situation therein.

Authors of book reviews and reports/Autorzy recenzji i sprawozdań

Anna Karnat-Napieracz (PhD in sociology from the Jagiellonian University, 2007) is an assistant professor at the Institute of Philosophy and Sociology at the Pedagogical University of Cracow. She specializes in analysing contemporary sociological theories, her recent research interests being focused on the role of individual and collective identity in the postmodern era. These issues are dealt with in her monograph (*Tożsamość, czyli świadomość redivivus*, 2009, in Polish) on the revival of the concept of 'consciousness' in contemporary social thought.

Małgorzata Krywult-Albańska graduated (2001) from the Jagiellonian University with an MA in sociology and received her PhD in 2009 from the same university. She is an assistant professor at the Institute of Philosophy and Sociology at the Pedagogical University

of Cracow. Her scholarly interests include population studies, migrations, environmental studies, and methods of social research.

Grzegorz Kubiński (PhD in sociology from the Jagiellonian University, 2005) is an assistant professor at the Institute of Philosophy and Sociology at the Pedagogical University of Cracow. His recent scholarly interests focus on sociology of religion, sociology of body and cultural studies. He published the following monographs (in Polish): *Narodziny podmiotu wirtualnego* (2008), *Alain Badiou. Ontologia mnogości* (2010), *Figury i wydarzenia. Agamben, Badiou, Negri* (2011).

Justyna Tomczyk graduated from the University of Silesia in Katowice with MAs in sociology (2009) and Polish philology (2007). She received her PhD in sociology from the University of Wrocław in 2013 and joined the faculty of the Institute of Philosophy and Sociology at the Pedagogical University of Cracow in 2015. She is interested in the sociology of social problems, sociology of politics, and social communication.