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)