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: status, 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.

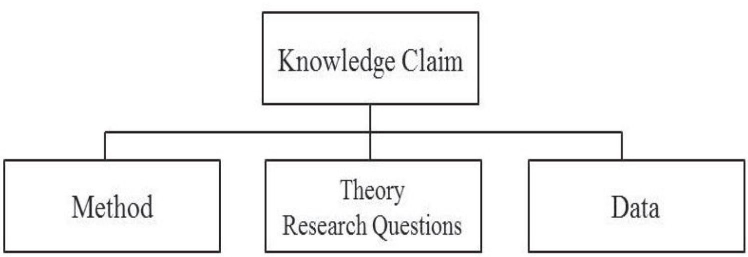

Figure 1. Assessing epistemic claims from experiments

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		
Method 1	**Traits**	A(1)	B(1)		A(2)	B(2)
	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

<table>
<tr><th>Predictions</th><th colspan="2">Target: Adults</th><th>Target: Children</th></tr>
<tr><td>1. Appearance</td><td colspan="2">+</td><td>+</td></tr>
<tr><td>2. Gender</td><td colspan="2">+</td><td>No data</td></tr>
<tr><td>3. No Intellectual Data</td><td colspan="2">+</td><td>+ (Weak Overall, Low Males)</td></tr>
<tr><td>4. Actual Competence</td><td colspan="2">Mixed</td><td>Weak</td></tr>
<tr><td>5. Perceived Competence (Target)</td><td>+</td><td>Judge+</td><td>No data</td></tr>
</table>

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*. 2^{n}nd 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