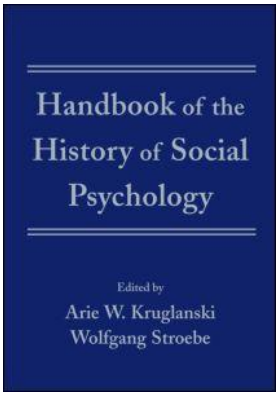

Handbook of the History of Social Psychology

Arie W. Kruglanski, Wolfgang Stroebe

The Evolution of Research Methodologies in Social Psychology: A Historical Analysis

Publication details
https://test.routledgehandbooks.com/doi/10.4324/9780203808498.ch8
William D. Crano, Andrew Lac
Published online on: 01 Dec 2011

8 The evolution of research methodologies in social psychology: A historical analysis

William D. Crano and Andrew Lac

> Science must begin with myths, and with the criticism of myths; neither with the collection of observations, nor with the invention of experiments, but with the critical discussion of myths, and of magical techniques and practices.
>
> (Karl Popper, 1963/1984, p. 50)

Depending on one's view of the proper domain of social psychology, a study of its historic methodological roots could range from the somewhat constricted to the extremely broad. To set the boundaries of this review, we adopt a broad conception of social psychology as the systematic study of the interplay of cognition, society, and behavior, and then proceed from the proposition that contemporary social psychology is an empirical science. Given these understandings, discussion of the history of research methodology in the field must begin with a consideration of the scientific method in general, whose exact beginnings are lost in the sands of time and cannot be established with certainty. More recent approaches and assumptions that undoubtedly grew from these lost beginnings are available, however, and will be considered here.

Foundations

We could begin with a commentary on proto-scientific Egyptian medical practices, whose papyritic record reaches back nearly 2,500 years (Breasted, 1930), with Thales, whom Russell (1945) anointed the father of western philosophy, or even with Plato's (380 BCE/1892) *Protagoras*, but we prefer to start with the Islamic Golden Age (8th–15th centuries AD), where important epistemological advances in the development of the scientific method, and hence, in social psychology's development as a scientific discipline, were made. Writing under house arrest in Egypt in the 11th century, Alhazen (Ibn al-Haytham: 1021/2001) produced a seven-volume work in which he discussed and promoted the experimental (or empirical) method.[1] His experimental approach produced persuasive results that flew in the face of accepted Ptolemaic and Aristotelian views of vision (Sabra, 1994). In his *Book of Optics*, Alhazen conformed to the scientific method in exploring visual perception and optical illusions, which play important roles in contemporary social psychological research on social influence (Adams, 1912; Crano & Hannula-Bral, 1994; Sabra, 1998; Sherif 1966; see also Prislin & Crano, this volume). Evidently, house arrest agreed with him, as it did with Geber, who about 200 years before Alhazen insisted on controlled variation between groups and measurement, as opposed to conjecture based on an accepted philosophical position. There was more than mere adoption of an empiricist, proto-experimental orientation in all of this. Scientists of the Arabian Peninsula at the time of Islam's Golden Age even anticipated contemporary notions of error and bias. Al-Biruni advocated replication and multiple measurements to mitigate random and systematic error (Glick, Livesey, & Wallis, 2005; Huff, 2003; Lindberg, 1992). Like contemporary thinkers, he recognized that any single measurement was "inevitably equivocal" (Campbell, 1969a, p. 49; Campbell & Fiske, 1959), and hence called for multiple measures and replication.[2]

European Renaissance

During the European Renaissance a few centuries later, the methodological insights of the early Islamic scientists gained considerable traction. However, because of the tenuous nature of measurement and research methodology of the time, knowledge gained from studies involving systematic variation between groups using a deductive, or top-down, approach was considered suspect by many. More certain for this audience was rational knowledge concerning the cause of an outcome, which was gained by commonplace observation and induction from fundamental axioms (first principles) and by definitions that could be demonstrated as true by mathematical proof. Thus, if the results of an empirical study using systematic variation were inconsistent with what was "known" to be true axiomatically, rejection of the study was more likely than rejection of the axiom. Given the dicey state of measurement at the time, this orientation was far from preposterous.

Translations of Islamic commentaries fostered an empirical orientation in the European Renaissance, and were important influences on Roger Bacon's development and refinement of the scientific method. Bacon's (1267/1859, 1267/1972) works, stimulated by his reading of Alhazen's (1021/2001) *Optics*, identified four causes of erroneous inference, including the appeal to authority. This was a courageous argument. As a Franciscan, all of Bacon's writings required the imprimatur of Clement IV, who incidentally died about a year after Bacon delivered his masterworks (we imply no necessary causal linkage between these

events). Bacon argued for the primacy of the experiment. He distinguished the experiment from any other form of knowing, which he variously termed argument or speculation. Only experimental science, he contended, could verify its suppositions, reveal heretofore undiscovered truths, and as such, learn Nature's (God's) secrets (Bacon, 1267/1859). Bacon insisted on description in exquisite detail of the ways in which experimental variables were defined, the manner in which experiments were conducted, and on replication (or verification) by independent researchers. In many ways, he anticipated Bridgeman's (1927) operationism, which was instrumental in the development of logical positivism (Carnap, 1952; Feigl, 1959), and which has been adopted wholesale in contemporary social psychology. Bridgeman (1927, p. 5) contended that abstract concepts were defined by the operations by which they were measured: "We mean by any concept nothing more than a set of operations; *the concept is synonymous with the corresponding set of operations*" (emphasis in original).[3]

Roger Bacon's work anticipated the eventual movement beyond the natural philosophy of Aristotle. Three to four centuries after Roger Bacon, Galileo and Francis Bacon (1620/1855), among others, even more forcefully rejected the Aristotelian approach (Avicenna, 1974). Galileo's work was influenced by two important events: his learning about the invention of the telescope, which led to his eventual building of a more refined instrument, and his encounter with *De revolutionibus orbium coelestium* (*On Revolutions of the Heavenly Spheres*), the heliocentric masterwork of Nicolaus Copernicus (1543/1992) that reduced the Ptolemaic cosmos to stardust. The observations Galileo made with his telescope convinced him that Copernicus was on the right track. With newly available technology, he was able to peer deeper into the cosmos than any of his predecessors, and what he found was not at all what he had been led to expect.

Although suitably circumspect in his production of revolutionary ideas—the Inquisition was in full flower—Galileo could not restrain his enthusiasm, whereas many learned men of his time refused even the opportunity he offered them to peer into the heavens. They viewed such observation as irrelevant, as it would either confirm what they knew or erroneously disconfirm the obvious. Galileo's legendary impertinence was to cause him significant problems when in 1632 he published his *Dialogue Concerning the Two Chief World Systems* (Galileo 1632/1967), for which he was called to Rome where his Inquisitors found him "vehemently suspect of heresy" and sentenced him to death (later commuted to house arrest for life). He appears not to have been convinced. It is said he muttered *"Eppur si muove"* ("And yet, it still moves") on leaving the courtroom. History's verdict on the great Galileo was considerably kinder than that of the Inquisitorial tribunal that condemned him and his ideas. Four hundred years after his death, Einstein (2002, p. 398) wrote:

> before mankind could be ripe for a science which takes in the whole of reality, a . . . fundamental truth was needed, which only became common property among philosophers with the advent of Kepler and Galileo. Pure logical thinking cannot yield us any knowledge of the empirical world; all knowledge of reality starts from experience and ends in it. Propositions arrived at by purely logical means are completely empty as regards reality. Because Galileo saw this, and particularly because he drummed it into the scientific world, he is the father of modern physics – indeed, of modern science altogether.

On the other side of the Channel, Sir Francis Bacon, Galileo's contemporary, had come to similar conclusions regarding the limitations of classic (Aristotelian) approaches to epistemology. His adoption of experimentation was far more enthusiastic than Galileo's, and his *Novum Organum* outlined a proto-experimental approach that was based heavily on the collection of facts, rather than the recitation of known truths. These facts were compared to help eliminate alternate hypotheses and ultimately lead to scientific truth, much like Karl Popper's (1959, 1962) more skeptical falsification approach that followed a few centuries later.

Bacon's deductive approach contrasted with the Aristotelian inductive alternative of drawing conclusions logically from a set of axioms whose validity was, in reality, itself a matter of conjecture. The deductive orientation, evident from the time of the early Islamic scholars, began to gain steam at this point in the history of science and, by extension, in the history of social psychology.

For many, the deductive/inductive clash was decided by Isaac Newton, whose four rules of scientific reasoning in the *Principia* include the principle that "arguments based on induction may not be nullified by hypotheses" (Newton, 1687/1999, p. 796). The Royal Society, chartered about the time of Newton's most productive period and often involved in deep discussion of his work, adopted the experiment and experimental evidence as the ultimate arbiters of truth, providing a powerful, state-approved imprimatur of the approach over the metaphysical orientation with which the experimental method had contended from the time of the Islamic Golden Age. The battle would continue over the years, with empirical deductive methods gradually but inexorably overhauling inductive approaches. In mainstream social psychology today, the contest is clearly settled; the hypothetico-deductive method has been adopted almost uniformly.

Methods leading to contemporary social psychology

Psychophysics

A direct approach to understanding the workings of the mind and its interaction with the world, social psychology's fundamental focus, is found in the work of Gustav Fechner, a physician turned physicist turned psychologist. His unique combination of abilities led to the development of psychophysics, a foundational feature of research in social psychology. Fechner performed "with scientific rigor those

first experiments which laid the foundations for the new psychology and still lie at the basis of its methodology" (Boring, 1950, p. 275). Fechner's motivation was to understand the relation between mind and body, between physical stimuli and their sensation, which he discussed in two influential books (Fechner, 1899, 1901). In coming to grips with the mind–body problem, he developed highly replicable methods of quantifying the relation between physical stimulation and perception, showing that mental events could be quantified via measurable stimuli.

The methodology was developed and extended in his creation of psychophysical methods, which helped lay the foundation of modern scale construction (Fechner 1860/1966), a *sine qua non* of almost all social psychological research. In *Elements of Psychophysics*, he demonstrated the benefits of careful experimentation and attention to the link between physical and mental events, and established the place of psychophysics in modern social psychology. Contrary to Kant's (1781/1887) observation that psychology could never reach the status of true science, as its principal objects of study were beyond measurement, Fechner's research showed that mental events could indeed be measured, and that they were linked systematically with measurable variations in stimulus intensity.

More or less contemporary with Fechner was Charles Sanders Peirce, whose contributions are woefully underappreciated in psychology, if not philosophy. Peirce (1878) argued for the necessity of both deduction and induction (vs. Hume, 1748/1963; 1772), and with his contributions to symbolic logic, outlined three fundamental and complementary modes of reasoning—abduction, deduction, and induction—that today are acknowledged in practice as necessary components of the scientific enterprise. In Peirce's view, scientific investigation is initiated by abduction, a guess or hypothesis, followed by deductive inferences drawn from hypothesis about other relations that must exist if the hypothesis is true. Developing this nomological theoretical net (anticipating Cronbach & Meehl's (1955) discussion of construct validity by 80 years) enables the scientist to deduce other testable hypothetical relations. One finds echoes of Peirce's notions of hypothesis generation many years later in McGuire's (1973) lucid discussion of the desirability of formal instruction on methods of hypothesis development in the training of social psychologists.

In his social research, Peirce was one of the developers (before Sir Ronald Fisher) and early practitioners of the randomized experiment and a major contributor to modern statistics (Peirce, 1877). With Jastrow, he published possibly the first randomized experiment in social psychology, using a shuffled deck of playing cards to randomly determine the presentation order of stimuli used in a psychometric study (Peirce & Jastrow, 1885). Using each other as participants, the researchers made a series of weight judgments comparing extremely similar weights and stated their confidence on each judgment. On most judgments the weights were so similar that there was "an absence of any preference for one answer over its opposite, so that it seemed nonsensical to answer at all" (Peirce & Jastrow, 1885, p. 74). For both researchers, most judgments seemed next to impossible. Jastrow, for example, expressed absolutely no confidence on 1,100 of 1,200 judgments, but guessed correctly on more than 70% of them, suggesting his judgments were substantially influenced by sensations of which he was consciously unaware. These results anticipated the implicit associations movement by more than a century. In distinction to Nisbett and Wilson's (1977) "Telling more than we can know," Peirce and Jastrow (1885) might have titled their paper "Knowing more than we can tell."

Correlational methods

Along with many of his time, Peirce was influenced by Darwin's (1859) *Origins* and the evolutionary perspective that his work brought to science. Darwin's startling observations poured fuel on the empiricists' fire, one of whose most distinguished arsonists was Francis Galton, his half-cousin. Galton's pioneering work on correlation continues to play a central role in contemporary social psychology. His influence is remarkable given the almost continual changes in technique, method, and issues that have drawn the attention of social researchers over the years. Galton's interests ranged from the inheritance of physical and mental traits in *Hereditary Genius* (Galton, 1869, 1874) to fingerprinting (Galton, 1895) to eugenics (Galton, 1909). His fascination with heredity brought him to a clear understanding of the regression artifact, the tendency of extreme values on fallible tests to regress to the mean of their distribution upon retesting. This statistical inevitability owes its existence to random error, and continues to bedevil researchers who assign participants to conditions on the basis of fallible measures (Campbell & Erlebacher, 1970; Crano & Brewer, 2002).

Galton developed the correlational technique to further his research on kinship (Galton, 1888, 1890; Stigler, 1989). The technique was refined and extended by Pearson (1896, 1920) and Yule (1897) to include multiple predictors of a criterion variable in what they termed multiple correlation (e.g., Pearson, 1914). Although some might argue that correlational techniques are, strictly speaking, statistical, and thus do not belong in a discussion of research methodology, the fundamental research design considerations that are brought to bear in development of correlational studies reasonably belong in a review of this type. Karl Pearson is credited with developing Galton's correlational insights with his product–moment correlation coefficient, the most widely used correlational test statistic today, but Pearson's *r*, as it has come to be called, stemmed from mathematical concepts conceived by Galton in 1888 (Bulmer, 2003; Pearson, 1920; Stigler, 1989). Galton's seminal ideas for correlation sprang from his desire to appraise the strength of the relationship between a pair of variables that might be drawn from entirely different distributions, and operationalized with different measurement scales. His

solution rested in part on standardizing the variable distributions prior to computing their degree of dependence. Initial applications of the concepts of correlation focused on heredity (Galton, 1889, 1890). Galton also was a trailblazer in the scientific development of questionnaires or surveys (Bulmer, 2003), arguably the most widely used response modality in social psychology (Campbell & Stanley, 1963; Sudman, Bradburn, & Schwarz, 1996).

While the correlational technique is an invaluable feature of today's social research, the method does require some constraint. Practical reasons and ethical concerns often preclude manipulation of theoretically relevant variables, so rather than manipulate, researchers correlate to understand the relatedness of naturally occurring variables. In consequence, correlational results are subject to contrasting causal interpretations. Examining the association between two variables is the most common form of correlational design. In these simple bivariate analyses, predictor and criterion are not distinguished. Although a strong correlation suggests a strong association between two variables, the causal priority of these variables cannot be determined via simple correlation. And even if it can be established that variation in one variable temporally precedes that of another, it is possible that an unspecified third factor may spuriously affect the relation between the two variables, further muddying the correlational waters. For this reason, demonstrating cause–effect relationships is generally unfeasible in correlational designs, but more recent longitudinal approaches that build time into their designs help the researcher identify likely causal linkages (e.g., see Huesmann et al., 2006).

Mediation and moderation

The analysis of mediation and moderation in correlational relationships expanded the scope of Galton's and Pearson's simple correlational model. Mediation and moderation models, popularized by Baron and Kenny (1986), provide means of assessing an intervening variable's influence on an outcome measure. Baron and Kenny's work has had great impact on the way social scientists view and understand correlational methods. A mediating variable elucidates the link between a predictor and an outcome. Analysis of mediation indicates the extent to which a putative causal variable operates through another (the mediator) in affecting an outcome. The quest for mediation is longstanding in psychology. It represents a felt need to delve deeper into the causal processes underlying a relationship in contexts in which the necessary causal methodologies are unavailable. This approach is reasonable, so long as researchers understand that causal castles based on meditational results are built on correlational sand.

An interesting example of using meditational thinking to advance a theoretical position is seen in the work of Tolman (1938), who theorized that appetite or biases might be intervening variables (or mediators, in our terms) of stimulus–response relations. In Clark Hull's (1943, 1951) learning theory, an approach that dominated much of psychology, including social psychology, in the 1940s and 1950s (Hovland, Janis, & Kelley, 1953; Hull, 1916), intervening variables (mediators) played a major explanatory role in his analyses of behavior. In practice, the strength of the meditational process is demonstrated by the extent to which the relationship from predictor to criterion is attenuated once the mediator has been taken into account. Alternative procedural variations to the Baron and Kenny (1986) model for testing meditational processes have been proposed recently (e.g., MacKinnon, Fairchild, & Fritz, 2007) and appear promising.

Although commonly confused with mediators, moderating variables interact with the predictor to modulate influence on the criterion variable. If a significant moderating effect is determined, the strength of the relationship from the predictor to the criterion systematically varies as a function of the moderator. Researchers are not limited to assessing a single intervening variable, as recent developments have demonstrated the feasibility of specifying complex multiple-mediator and multiple-moderator models (Aiken & West, 1991; Preacher & Hayes, 2008), as well as mediated moderation and moderated mediation models (Muller, Judd, & Yzerbyt, 2005; Fairchild & MacKinnon, 2009).

Meta-analysis

Meta-analysis is a useful and increasingly common technique that integrates statistical moderation with fundamental correlational methods (Glass, McGaw, & Smith, 1981; Hunter, Schmidt, & Jackson, 1982). The general idea for the technique dates to Karl Pearson's (1904) approach to enhance power of small-scale studies by combining them into a single analysis (O'Rourke, 2007). Meta-analysis is a family of techniques used to combine results of multiple studies on the same phenomenon by calculating the size of the treatment effect found in each, and combining these estimates to estimate the average effect of a given treatment or intervention on a specified outcome. The approach has been used not only to estimate the effect size of treatments on outcomes, but also to study factors that moderate these effects. By combining studies, meta-analysis provides greater power than any study of which it makes use, and thus a better estimate of the overall effect of an intervention or treatment on a dependent measure. It can provide insights into factors that enhance or attenuate the critical relationship by identifying factors that systematically moderate effect sizes across studies. Issues in the proper estimate of effect sizes remain, and in dealing with distortions that might be introduced through the common editorial practice of publishing only statistically significant results (thus, the standard meta-analysis may overestimate effects if this possibility is not recognized and assessed—but see Rosenthal, 1991). The empirical literature of social psychology is sufficiently developed that meta-analysis is both possible and sensible, and will help accelerate progress. In addition, the rapid accumulation of reliable information presents a strong

counter to critics who decry the noncumulative nature of social psychological research.

Multiple regression and matching

Mediating and moderating mechanisms may be evaluated using multiple regression, an extension of the bivariate correlation framework that emerged from the work of Galton, Pearson, and Spearman. Multiple regression analysis enables researchers to assess and forecast the contribution of more than one predictor on a single criterion (Cohen, Cohen, West & Aiken, 2003). An advantage of this approach over correlation, aside from facilitating examination of the concurrent contribution of multiple predictors, is its capability to remove contaminating covariates that might affect an outcome. Multiple regression facilitates assessment of spurious causes and statistically equalizes sampled participants on extraneous factors. The challenge for investigators attempting to control for confounds statistically is to anticipate such artifacts and measure them in advance of the analysis. In practice, it is impossible to measure participants on every plausible intervening variable, and therefore regression models more often than not are underspecified. Even so, though regression designs often are not as rigorous, or controlled, as randomized experimental designs, they are assuredly a step above designs that ignore measurable and obvious sources of error.

The problem of underspecification also arises when using matching participants to offset the inability to them randomly to conditions. Matching, common in quasi-experimental contexts, suffers from the problem of the missing critical variable (Crano & Brewer, 2002); no matter how hard one tries, it is inevitable that a critically important matching variable is left out of the analysis. Simple matching always involves underspecification of factors that might have an important effect on outcomes.

In addition to underspecification, matched samples often are drawn from populations with decidedly different (mean) values on the critical matching variable. In this case, the matches perforce occur in the overlap of the matched samples' distributions. In these instances, the positive outliers from one group are matched with the negative outliers from the other. This overlap area is an ideal breeding ground for regression artifacts, the tendency of extreme scores to regress to the mean of their distributions, as Galton (1886) pointed out in his study of hereditary stature. Differential regression tendencies can hide real treatment differences or exaggerate them, misleading researchers in either case.

Propensity score analysis, an improvement and possible corrective to simple matching, can make use of a multitude of indicators to match respondents from different populations in attempting to equalize initial status of groups that cannot be assigned randomly to conditions. Developed by Rosenbaum and Rubin (1983), and widely used throughout the social and life sciences, this type of matching uses measured variables chosen on the basis of theoretical relevance to create a propensity score, the predicted probability of the respondent's membership in a treatment or comparison group. Typically, these scores are determined via logistic regression. Treatment and control respondents who share the same propensity score are compared. With propensity matching, between-group differences are more confidently attributed to the treatment than to initial differences, which presumably are equalized in the matching process. This approach has been used to advantage in numerous quasi-experimental field studies that did not admit to randomization (e.g., see Fendrich, Lippert, Johnson, & Brondino, 2010; Wu, West, & Hughes, 2008; Zanutto, Lu, & Hornik, 2005), but it is not without its detractors. Shadish, Cook, and Campbell (2002), for example, have warned that propensity matching cannot definitively solve the underspecification problem, and that like all regression-based approaches, it is based on the assumption that the variables entered into the matching routine are measured without error. The seriousness of these threats varies from study to study, but to the extent that they hold, the approach can produce misleading results (Michalopoulus, Bloom, & Hill, 2004).

Structural equation modeling

Another statistical/methodological development that builds on and extends multiple regression, structural equation modeling, offers the advantage of allowing multiple predictors, which can be used to forecast one or more criterion variables. The approach allows explicit assessment of measurement errors that can be attributed to measured variables via latent factors (Jőreskog, 1973). The path model approach specifies predictive associations between variables, but lacks a latent measurement component. Path analysis was invented by Sewall Wright, one of the 20th century's foremost contributors to statistical methodology, whose intellectual hero was the great Galton. Wright was said to have published a paper in 1917 that "foreshadowed the analysis of variance and covariance, and . . . illustrated its use in an example which was a forerunner of the biometrical concepts of heritability and genetic correlation. Not to be outdone, in the next year he published a paper using his new technique of path analysis" (Anon, 1990, pp. 277–278).

Wright (1918, 1921, 1923) developed path analysis to investigate the relative contributions of environment and heredity to physical characeristics. In likely the earliest path diagram, Wright (1920) illustrated how parental genes, environment, and developmental differences predicted variations in color patterns of guinea pig offspring from the same litter. Path analysis has been extended and used in an extraordinary variety of social psychological, sociological, medical, and economic investigations. However, the approach is not a statistical philosopher's stone that can change correlational dross into causal gold. At a minimum, path analysis requires specification of a theoretical model in advance of the analysis (Crano & Mendoza, 1987); some have reversed this process, and in so doing have violated Wright's intentions (Denis & Legerski,

2006) and have thereby forfeited the causal logic of the enterprise.

In the 1970s, Wright's technique was refined with the incorporation of a factor analytic component. This hybrid approach has become more commonly known as structural equation modeling, which, along with other latent modeling approaches (e.g., factor analysis, canonical correlation), is an advance over earlier multiple correlational models whose measured indicators are impure representations of the underlying concept, and therefore contain unreliability attributable to measurement error (Ullman & Bentler, 2003). The latent approaches are designed to remove such error, and thereby improve the standard regression models that assume perfect reliability of measures.

Latent multilevel modeling approaches, close cousins of the structural equation modeling techniques, are becoming more widely used in social psychology (Heck & Thomas, 2008). They promise to provide information on both group and individual variations across time (if longitudinal data are available), and sidestep many of the limitations of the more standard random effects multilevel models. Multilevel modeling is, strictly speaking, a statistical innovation, but it will have an incremental effect on research design and methodology in social psychology owing to its capacity to isolate variance attributable to individual vs. group level differences (e.g., Cooper, 2010; McCullough, Luna, Berry, Taback, & Bono, 2010).

Scale construction and factor analysis

All of these techniques became necessary with the development of scales designed to assess internal states, an advance in social psychological methodology fostered by Louis Thurstone (1928, 1931a, 1931b) and his student Rensis Likert (1932), and later revamped and revitalized by Osgood, Suci, and Tannenbaum (1957). The need for scale validation techniques became all the more pressing with the development of these measures, which were designed to tap previously unmeasurable cognitive features—beliefs, attitudes, and values. The scaling methods proposed by these researchers demanded consideration of the reliability of the instruments that were developed. Kuder and Richardson (1937) and Cronbach (1951) made noteworthy and widely used classic test theory-based contributions to the psychometric quality of measurement instruments in social psychology. Alternative and more recent approaches based on the item, rather than the whole scale, also have proved useful (Lord, 1980).

Factor analysis and scale validity

Complementary methods used to determine scale quality made use of basic correlational methods augmented by techniques of exploratory and confirmatory factor analysis. Factor analysis seeks to identify items that are strongly associated within a set (or factor), and that are weakly related (or unrelated) to other variables that fall in different factors. The technique allows a parsimonious grouping of sets of variables into common underlying factors, and is used extensively in development of measures whose items share a common focus. Charles Spearman, another of Galton's admirers, was largely responsible for the initial development of factor analysis (Spearman, 1904), one of the underlying methodologies that is a common feature of all latent variable models. Although he developed this method to test and promote his two-factor theory of intelligence, it is widely applicable across a broad research spectrum.

In pursuing improved measurement of psychological states, researchers have become increasingly concerned with the construct validity of measures, the extent to which a psychological concept plausibly exists (Cronbach & Meehl, 1955). This question, too, is typically investigated via correlational approaches (Crano, 2003). Psychometrically, construct validity is indicated if factors (scales or other indicators) with which the measure theoretically *should* relate are statistically associated with the scale (convergent validity), and factors with which the measure theoretically *should not* relate are not statistically associated with the scale (discriminant validity). In its most rigorous conceptualization, construct validity may be evaluated using the classic multitrait–multimethod matrix (MTMMM) of Campbell and Fiske (1959), which offers researchers the procedural apparatus necessary to disentangle method variance (i.e., score variations attributable to the particular method used to measure a trait) from trait variance (variation attributable to true-score differences among the traits under study). In its originally proposed formulation, the MTMMM was evaluated with a series of commonsense rules and bivariate correlations, but structural equation modeling is more widely used today, as it takes advantage of latent factor approaches for assessing the contribution of both forms of variance (Crano, 2000). Even so, after more than a half-century's work and thousands of studies, consensus regarding the ideal method of evaluating the MTMMM has yet to surface. In a somewhat plaintive retrospective on the approach, Fiske and Campbell (1992) voiced doubts about the possibility of ever disentangling trait from measurement variance.

The future of correlational research can be seen in development of innovative methods designed to reduce rival explanations and mimic causal designs. Longitudinal designs reflect noble attempts to establish the temporal precedence of cause–effect variables. One such possibility is seen in the more widespread use of the underutilized crosslagged panel design (Crano, Kenny, & Campbell, 1972; Pelz & Andrews, 1964; Rozelle & Campbell, 1969), which with appropriate caution uses time order to help estimate the causal preponderance of relationships that are crossed and lagged over time (Kenny, 1975). Several studies across disparate topics have applied this technique to address critical practical and theoretical concerns in social psychology (e.g., Huesmann Eron, & Dubow, 2002; Huesmann, Moise-Titus, Podolski, & Eron, 2003; Huesmann et al., 2006; Kahle & Berman, 1979; Stacy, Bentler, & Flay, 1994).

The evolution of social neuroscience

The debunked practice of phrenology would not seem an appropriate topic in a consideration of the history of research methods in social psychology, but it bespeaks a longstanding interest in the field and foreshadows an important modern development. Headed by Joseph Gall in the early 1800s, the phrenology movement was an early and later discredited approach that sought to identify cranial regions associated with personality and psychological traits (Gallup, Frederick, & Pipitone, 2008; Whitaker, 2000). These cranial regions, or bumps on the head, were erroneously believed to emerge as a result of their corresponding cortical regions. However futile, this attempt at brain mapping assumed the functionality of the brain and suggested that localized regions of the brain were specialized for different cognitive tasks. With recent advances in technology, functional magnetic resonance imaging (fMRI), with its foundations in positron emission tomography (PET) and MRI scans, allows researchers to examine noninvasively the correlation between neural activity and psychological processes via changes in cerebral blood flow (Savoy, 2001). This psychophysiological method is used to identify and isolate the cortical regions underlying specific mental operations.

The promise of social neuroscience research has been widely and eagerly anticipated and touted, and even in its early stages such research has produced extraordinary results by applying, among others, fMRI technology to investigate the ways in which the brain is involved in the mediation of social interactions (Cacioppo, 2002; Cacioppo & Berntson, 1992; Harmon-Jones & Devine, 2003). However, some cautions have been voiced against an uncritical acceptance of the methods used in fMRI research (Vul, Harris, Winkielman, & Pashler, 2009a, 2009b), which relate principally to issues of data selection ("cherry-picking") and the nonindependence of observations. Selecting only values that exceed an arbitrary activation threshold, especially in studies in which a massive number of cortical observations (voxels) are collected, obviously could bias results in hoped-for directions. This issue has yet to be resolved.

The experiment in contemporary social psychology

Regarded as the gold standard not only in social psychology but in empirical research in general, experimental methodology established its roots from the very beginnings of science, as discussed throughout this review. The capacity of experimental methods to shed light on fundamental causal relations has proved an irresistible draw to science, and over the centuries, experimental techniques have been refined and developed on the basis of fundamental experimental/causal logic. The attraction of experimentation was evident from the earliest days of social psychology. In its least stringent form, the experimental method is manifest in any investigation in which some control of extraneous alternative variables (confounds) is deliberately and systematically emplaced by the researcher. This broad conceptualization, which deviates from the modern definition mandating the use of random assignment of subjects to conditions, was the dominant "experimental" methodology in the formative years of psychology.

Introspectionism

A movement that was to have a powerful effect on the development of the field grew out of the efforts of Wilhelm Wundt, who established the first laboratory in experimental psychology at the University of Leipzig in 1879 (Domanski, 2004; Mandler, 2007). In addition to his extraordinary written output—approximately 50,000 printed pages (Boring, 1950)—Wundt also was responsible for mentoring 186 PhD students, who went on to have enormous impact on the growing discipline; these students included Oswald Külpe, James McKeen Cattel, G. Stanley Hall, Hugo Münsterberg, Charles H. Judd, Hugo Chavez, Lightner Witmer, Charles Spearman, and Constantin Rădulescu-Motru. As well, he was indirectly responsible for the first psychology laboratory in the United States at Cornell, having mentored its founder, Edward B. Titchener. Wundt's goal was nothing short of establishing a new experimentally based independent scientific discipline, and he succeeded. As Boring (1963, p. 63) put it, "Wundt meant to found experimental psychology, and it was founded." Clearly this new discipline was built on the shoulders of giants, but without the added stature of Wundt, the field might have evolved very differently.

Wundt's background and training played a significant role in psychology's development. He received his degree in medicine but then studied physics and physiology with von Helmholtz, who also has been nominated as scientific psychology's founder. All of this training was to prove critical as, influenced by the work of Fechner (yet another nominee), Wundt (1874/1910) turned his attention to physiological psychology and published one of psychology's classics, *Principles of Physiological Psychology*. In this work, Wundt insisted on the primacy of experience, and of the experiment, whose techniques he had absorbed from centuries-old traditions in the natural sciences. Owing to Fechner's important psychophysical discoveries, Wundt directed his considerable energies to the discovery of the physiological link between mind and body. His studies were based on a system that has come to be called "structuralism," whose primary investigative methodology was self-observation, or *introspection* (Titchener, 1899a, 1899b). This method was designed to delve into and identify the components of consciousness, and in so doing unearth the elemental mental processes that were the basis of higher thought. This approach shaped the early days of experimental psychological research methods and was to become a prime target of the operationalistic and behavioristic counterattacks (Lyons, 1983; Radford, 1974; Vermersch, 1999).

Titchener's insistence that psychology be a laboratory science, divorced from the confounding influences of the real world, made sense in the early phases of formal psychology's

development. However, it is arguable that this overemphasis on laboratory-based research retarded the impact that social psychological insights and results could have on civil society. It also is arguable that this emphasis has contributed to the various "crises" in social psychology that have arisen over the years (Gergen, 1973; Vygotsky, 1927/1997). It has taken many years to begin to offset this bias, inculcated in the formative years of psychology, but evidence suggests that Titchener's orientation is being brought into a more balanced state in contemporary social psychology partly as a function of a more refined methodological armamentarium and partly in response to societal demands.

Owing to the extreme reaction the approach stimulated, introspectionism has gained perhaps a worse reputation than it deserves. Titchener (1901–1905), for example, instituted strong controls in his use of the technique. Of course, one might argue that any attempts at controlling introspection might constrain the very (free) mental processes it was designed to elucidate, and thus, ultimately defeat the intended purpose of the technique. Though introspectionism has been largely discredited, use of subjective reports of mental processes has not disappeared from the realm of mental discovery. Remnants of the introspectionist orientation can be observed in many of today's social psychological methods, especially in approaches commonly used during the early phases of research. The *think-aloud* protocol, which requires respondents to report on their thought processes and the mental steps engaged in while forming a response, are common and useful in the design of instruments (Sudman, Bradburn, & Schwarz, 1996). Cacioppo, von Hippel, and Ernst (1997) found verbal thought-listing useful in mapping cognitive structures, and Petty, Cacioppo, Strathman, and Priester (2005) have used the approach extensively in their study of the mediators of persuasion. Structured and nonstructured focus groups that encourage respondents to voice their thoughts and opinions openly on a topic of concern also provide grist to the empirical mill (Siegel, Alvaro, Patel, & Crano, 2009).

Contemporary experimentation and randomized assignment

The essence of contemporary experimental design has evolved over the years into a narrower and more restrictive format that requires random assignment of units (subjects) based on chance procedures (Fisher, 1925/1950; Campbell & Stanley, 1963; Crano & Brewer, 1973). This insistence on randomized assignment developed to help ensure that participants across groups were "equivalent" at the onset of the study, consequently justifying causal statements linking manipulation (independent variable) to outcome (dependent variable). Compared to research in which participants self-selected themselves into experimental groups, with their varying histories and individual characteristics confounding the interpretation of differences, the use of randomization in assignment of participants to experimental conditions represents a vital advance.

Though it is not entirely clear who conducted the first deliberately randomized investigation, it likely took place in the 19th century (Greenberg & Shroder, 2004). Statistician Ronald Fisher has been credited as responsible for popularizing random assignment in his book *Statistical Methods for Research Workers* (1925) though, as discussed, Peirce and Jastrow (1885) had used randomization of stimuli to mitigate order effects in their early study of decision making under uncertainty. The importance of random assignment of participants to conditions in social psychological research is difficult to overestimate. It facilitates disentangling between-group differences attributable to the experimental treatment from those that might have occurred as a result of factors that arrived with the respondents, and over which the researcher had no control.

Randomization to conditions is necessary in the "true" experiment (now generally known as randomized controlled trials—RCT), but early studies were hampered by the inability to compare more than two outcomes (treatment vs. control group) at a time (Campbell & Stanley, 1963). These two-group statistical comparisons were enabled by Student's *t* test, invented by William Gosset, who developed the procedure while working for the Guinness Brewery. Owing to commercial secrecy concerns, he adopted Student as his *nom de plume* (Mankiewicz, 2000). The test represented a great advance, but comparing more than two groups at a time was beyond its capacity.

Factorialization and the analysis of variance

Sir Ronald Fisher, probably the 20th century's most renowned statistician, was a young man of 29 years when in 1919 he joined the Rothamstead Experimental Station in Hertfordshire, England. His first job was to make sense of more than 60 years of field trial data. If nothing else, his work demonstrates the importance and utility of secondary data analysis, a methodological approach that should be used with greater frequency and to greater effect in social psychology. To further his secondary analyses, Fisher developed the analysis of variance (ANOVA), which enormously expanded researchers' capacity to study the complex interplay of manipulated *and* measured variables. No longer were scientists limited to comparison of only two groups in their experiments, a limitation imposed by the restrictions of Student's *t*. With Fisher's (1925/1950) analysis of variance, the rigorous statistical decomposition of factorialized experiments became possible. These designs allowed the researcher, in the same study, to examine the direct and interacting effects of multiple independent variables on an outcome measure. Fisher's analytic model accelerated progress in social psychology exponentially. His innovation allowed for discovery of interactions among independently manipulated variables, which scientists otherwise would not have been able to establish unambiguously. Although it is true that many classics of experimental social psychology are based on a comparison of only two mean values, it is even truer that a comprehensive understanding of the outcomes of such studies, their boundaries

and dynamics, almost inevitably requires a more complex investigative orientation, of which testing for interactions among variables plays a central role. The expanded vision permitted by multiple simultaneous manipulations is analogous to that provided in correlational studies with the popularization of statistical techniques to investigate mediation and moderation (Kenny, 1979).

Randomization and its shortcomings

Although randomization is essential, it is not a cure-all for poor design (Sidani, 2006; Strube, 1991). An unintended failing of randomization (or, better, of randomizers) can be realized owing to a misunderstanding of the limits of the procedure. Randomization is effected in the attempt to equalize groups. That it succeeds in this equalization is largely an article of faith for social scientists, for even with very large samples it is not reasonable to assume that randomization matches participants between conditions on every conceivable psychological, physical, and social characteristic that might affect an experimental outcome (Strube, 1991). However, if one were to judge the likelihood of randomization failing to produce equivalent groups, it is obvious that the potential to fail would be negatively related to within-cell sample size (Krause & Howard, 2003). In randomization, size matters. It is for this reason that Crano and Brewer (2002) insisted on the "law of large numbers" in randomization. To believe in the therapeutic effects of randomization without a concomitant acknowledgement that the process cannot reasonably be expected to work as hoped without large numbers of units (or subjects) assigned to conditions is akin to a child's belief in magic. Such beliefs may be comforting, but they do not belong in the scientific study of the interplay of cognition, society, and behavior. It is unfortunate that many of our best journals still see fit to publish research in which the number of observations within conditions seems laughably inadequate to assume that the power of randomization to equalize groups has been harnessed. Arguing that the presence of statistically significant effects with small numbers is persuasive proof of the strength of a treatment's effect betrays a serious misunderstanding of the function of randomization, which is the pretreatment equalization of comparison groups so that individual differences cannot reasonably be judged to interact with treatment variations. If this equalization cannot be accepted because of an overly constricted sample, the experiment is irremediably compromised. Statistical power does not speak to this issue, as it assumes equalization of groups before comparison.

Considerations of internal validity

In addition to problems brought about by an overdependence on the saving grace of randomization, it is essential that experimenters recognize that factors not related to the treatment must be ruled out as explanations of their findings in such a way that they cannot threaten interpretation. In a classic exposition of experimental and quasi-experimental designs, Campbell and Stanley (1963) outlined a series of factors that mitigated confident interpretation of the effects of an independent variable on outcomes. They termed these factors *threats to internal validity*. The usual, and most familiar, of the true experimental designs (or RCTs) is the pretest–posttest control group design (randomization to conditions is understood), which effectively controls for a host of factors other than the experimental manipulation, which thus is attributed as responsible for differences between groups. Having ruled out the likelihood that extraneous alternatives affected the experiment's outcome, we are left, in a Popperian sense, with the experimental treatment as the best (last-variable-standing) explanation of differences between treatment and control groups. These extraneous alternatives, or threats to proper inference, include historical events that occur between pretest and posttest, maturational differences that might obtain between groups, testing artifacts, subject mortality/drop-out, statistical regression, and so on. Campbell and Stanley observed that when properly conducted, the pretest/posttest control group design can offset these problems. In a later discussion, Shadish et al. (2002) appended a series of threats to proper inference that the wise experimentalist also must address. These threats revolve principally around misuse and misinterpretation of statistical routines. They include low power, violations of the statistical assumptions, unreliability of treatment implementation, and so on. As before, a proper understanding of the techniques being used should offset most of these threats, though certainty is never assured. No matter how well designed the study, internal validity is always in question, insofar as chance and other uncontrolled factors can always affect outcomes.

Other designs

An interesting addition to the "standard" experimental design, the posttest-only control group design, dispenses with the pretest altogether. Assuming that random assignment operates as designed, a pretest is not, strictly speaking, necessary. In addition to the obvious economic gains, giving up the pretest can enhance the generalizability (external validity in Campbell & Stanley's (1963) terms) of experimental results, as pretesting at times can unintentionally provide clues to the study's hypothesis and purpose, thereby potentially contaminating the effect of the manipulation and influencing posttest responses.

In contexts in which the extent of reactivity is an important theoretical as well as methodological issue, a sensible and underused factorial design option is available. It combines the pretest/posttest control-group design with the posttest-only control group design. This innovation, the Solomon (1949) four-group design, allows the researcher to determine the effects of pretest, treatment, and their interaction, which provides information regarding the reactivity of pretest with treatment (Braver & Braver, 1988). The design is twice as costly as either of its two component parts, but it provides

valuable information on the effects of the pretest on subjects' sensitivity to the treatment. If pretest sensitization effects are at issue, the Solomon design is worth considering.

Popper

Campbell and Stanley's (1963) orientation to experimental design owed much to the strong voice of Karl Popper, one of the 20th century's most influential philosophers of science. Popper's *critical rationalism* rejected induction's methods of proof, just as it dismissed the possibility of certainty in establishing the validity of any theory. Popper held that a theory's validity could never be confirmed unambiguously, but it could be disconfirmed by a single inconsistent result. This asymmetric view held that theories that were not falsifiable (and Popper considered Marxism and psychoanalysis to be examples of such theories) were not scientific—they belonged in the realm of theology, not science. He believed that all that we consider knowledge—including scientific theories—was hypotheses that were invoked to deal with contemporary issues ("Science must begin with myths, and with the criticism of myths . . ."). Theories inevitably were time- and culture-bound, subject to change as conditions changed. Establishing enduring truths was not in the job description of the working scientist.

The crisis(es)

Most contemporary social psychologists are in accord with Popper's critical rationalist views regarding the impossibility of proving a position—coupled with the ever-present possibility of disconfirmation—and are prepared to live with the constraints imposed by this understanding. Even so, recognizing the limitations imposed by the use of the experimental method in social psychology seems to give rise from time to time to critical observations on the state of the art. Some label these cyclical eruptions *crises*, but as this review suggests, owing to its strong adoption of natural science methods, social psychology has been in a state of crisis from Titchener's time. We have not reached, and probably will not reach, what Thomas Kuhn (1970) dubbed the paradigmatic stage, where the general methodology and fundamental axioms of the science have been accepted by its practitioners. But this does not invalidate the value of the enterprise. We traverse the long road to certainty inch by inch, but as scientists should be well aware that we will never reach this destination.

Of all the discussion of crises in psychology that have been raised, the most perceptive was that of Lev Vygotsky (1927/1997). Although not accessible in the west for many years, this criticism of fundamental assumptions deserves close scrutiny now that it is available. Vygotsky's argument begins with the claim that by adopting the classical methods of the natural sciences, psychology made a fundamental and possibly fatal error. The objects of study in the natural sciences, Vygotsky argued, do not have the capacity for thought or volition—they are immune from the knowledge gained through science. The iron rod used to study heat expansion in our introductory physics laboratory was not aware that it was supposed to expand when we heated it. This state of affairs is not the case in the social sciences, where the objects of investigation usually are thinking, living, human beings, who are not independent of the techniques used in their study, and whose later behavior is affected by the knowledge developed in such studies. As such, the social sciences are largely place-bound and stuck in the present, with knowledge likely to change as conditions change.

The complaints did not stop with Vygotsky. Kim (1999) summarized a virtual litany of researchers' epistemological and gnoseological plaints, including triviality; artificiality; cultural, historical, and methodological boundedness; culture blindness; intellectual sterility; an overly narrow focus on American individualism and world view; and methodological confoundedness. Of all of the modern critiques, that of Gergen (1973) seems to have made the greatest splash. This critique echoed and illustrated the gist of Vygotsky's insightful essay and stimulated considerable critical commentary (e.g., Schlenker, 1974). However, although the mainstream practice of social psychology as a laboratory-based experimental science did not appear to change much in response to this challenge, this work and others that followed from it seemed to encourage a more expansive view of the kinds of information that *could* constitute legitimate social psychological evidence. Research based on social representations (Moscovici, 1988, 2008; Moscovici, Duveen, & Flick, 1998), discourse analysis, and other forms of nonexperimental knowing have all gained momentum from the critical reconsideration of the standard approaches.

This expansion of the methodological store has met with mixed reviews and with mixed success. The gist of the conflicting responses is that any enlargement of our research repertoire is welcome, as new models of knowing and new empirical techniques may open heretofore unexplored avenues for new and creative research; conversely, these approaches are decidedly nonexperimental, and a movement in this direction is seen by some as a regression, a retreat from hard-won gains. Both positions are arguable. To be sure, the expansion of useful research visions is always desirable, so long as the methods that are part and parcel of these expansions are used appropriately and their limitations understood and addressed. The hegemony of experimental techniques in much of social psychology is apparent—and not necessarily beneficial. A one-size-fits-all orientation is not the stuff of which rapid progress is made (Crano & Brewer, 2002). In many instances, the less constricted methods developed over recent years promise a clearer picture of the lie of the land than the highly controlled experiment. On the other hand, for testing (if not developing) causal propositions, the experiment has proved its worth over centuries of refinement. It would be unwise to throw this baby out with the methodological bathwater in our adoption of alternative models. With judicious application, the experiment remains

the gold standard, but it obviously is not the only standard or even the preferred one in all instances.

Field and quasi-experiments

Though experimental methods can be used to infer, if not to establish unambiguously, cause-and-effect relationships, the leap to causation may be entirely futile if the operations used to define the independent and dependent variables are flawed and do not accurately reflect the underlying constructs in question. Even more problematic is the possibility that the controls put in place in laboratory settings paradoxically may present salient social cues that elicit responses that are more a function of the research context than the experimental treatment. Orne and Scheibe (1964) termed these laboratory-based biases *demand characteristics*. In addition, experiments conducted in artificial laboratory environments often lack resemblance to real-world contexts. Studying the temperature–aggression relation in the laboratory, for example, often produces different results from those obtained in field contexts (Anderson, 1989) that make use of social indicators of aggression, such as murder or assault.

To bridge the gap between laboratory and field, investigators have turned increasingly to naturalistic observational techniques, which may include field experimentation, to isolate intrapersonal and interpersonal behaviors and interactions in settings outside the laboratory. These approaches are thought to be informative about behavior in the larger world outside the laboratory. By the 1960s, dissatisfaction with the pace and yield of laboratory experimentation, along with a growing distaste for the ethically questionable games played on laboratory subjects in the apparent pursuit of social psychological knowledge, led to a renewed interest in more applied, field-oriented experimental approaches. These more naturalistic studies avoided the artificiality of the laboratory, and most often did not involve unethical treatment of research participants, though with characteristic ingenuity, this sometimes was accomplished as well. At times, the field orientation was bought at a cost of precise control, but the benefits of the approach were seen by many as worth the price.

Naturalistic experiments designed to capture real-world social behavior, while *apparently* holding a stronger claim to generalizability, lack many of the controls available in carefully controlled laboratory environments. Take, for example, the classic Stanford Prison Experiment, an examination on the effects of deindividualization in which 24 participants were randomly assigned to the role of either prisoner or guard in a simulated prison (Haney & Zimbardo, 1998; Zimbardo, Haney, Banks, & Jaffe, 1974). The extent to which it is plausible to attribute results to the manipulation unambiguously is complicated by the fact that the prisoners were allowed to interact with each other, as were the guards, and participants within both of these cohorts interacted across groups. Even so, the value of insights provided in work of this sort is difficult to question, as is their effect on further thoughtful theoretical development (Zimbardo, 2007). Though randomization to groups might have helped mitigate outcome disparities due to individual differences in the Prison Experiment, dyadic interactions could have allowed group emergent properties to influence results spuriously. Furthermore, the sample sizes of many such randomized naturalistic studies are small, making the sought-for effects of randomization debatable at best (Crano & Brewer, 2002). Perhaps the degree to which any study represents the ideal of the true experimental design should be viewed as falling along a continuum. If this were done, there is little doubt that the appropriate placement of specific investigations often would become a matter of contention.

Coupled with a view of experimentation as falling along a continuum, a concern with social psychology's contribution to society has led to a revitalized emphasis on field and quasi-experimentation, which Cook and Campbell (2002) advanced in their classic volume. This movement was stimulated in no small part by Campbell's (1969b) "Reforms as experiments," which set powerful nascent psychological forces in motion, forces that called for the integration of strong methods with interventions designed to enrich the lives of the citizenry. This work was undoubtedly stimulated by Lyndon Johnson's Great Society programs, whose salutary effects on United States society are in evidence even today, a half-century later. A list of Johnson's, and the majority Democratic Congress' accomplishments would require many pages, but consider some of the Great Society programs and legislation, which included increased protections for civil rights, the War on Poverty, Head Start, the Higher Education Act, the National Teacher Corps, Medicare, Medicaid, the NEH and PBS, and the construction of the Kennedy Center and the Hirshhorn Museum in Washington, DC, among others. And all of the massive expenditures involved in these programs required evaluation of their impacts. This requirement helped move social psychology out of the laboratory and onto the streets, improving both. The forces leading to a more impactful and contributory social psychology had been part of the scientific landscape for years (Lewin et al., 1945; Lewin, Heider, & Heider, 1936), but "Reforms" renewed and reinvigorated the challenge to social psychology to make good on its promise as a force for progressive social change by daring its practitioners to experiment on, and evaluate the worth of, socially relevant interventions.

A defining element of quasi-experimental designs is their lack of random assignment of research subjects to conditions. This limitation had been viewed as a fatal shortcoming for causal inference, but as has been argued (Cook, 2007; Reichardt, Trochim, & Cappelleri, 1995; Shadish & Cook, 2009; Trochim, 1984), even the inability to assign randomly does not inevitably rule out causal inference. It also should be understood that even RCTs (experiments) vary in the degree to which causation may be attributed unambiguously. If experiments and quasi-experiments were aligned along a *certainty with which causation may be inferred* dimension, there would be considerable overlap of various experimental and quasi-experimental techniques, insofar as some methods that clearly do not allow for randomized assignment substantially control for all threats to

internal validity (e.g., the regression-discontinuity design), and hence their utility in forming causal inferences may be as great as that of experiments (Trochim, 2005).

Social psychology is more than ever involved in research with the potential to enrich and refine our contributions as responsible citizens to society at large, and to the science we practice. This renewed social sensitivity is a necessary development for the continued wellbeing of social psychology, and today's practitioners would do well to foster it. As McGuire (2003) presciently predicted, "A Mandarin stance of science for science's sake, however claimed by the high-table elite, would lose support from other segments of society, including funding agencies" (p. 135). We have seen this prophecy come to pass. This movement to the field, using designs that are responsive to the realities of the research context, but which often do not allow for strict randomization of subjects to conditions, sometimes taxes the methodological ingenuity of even our best social psychologists. It results in studies that stretch our comfort zone in terms of the certainty with which we interpret the meaning of our analyses—but the rewards can be great. It also is wise to remember that one study almost never unambiguously establishes anything. If an acausal quasi-experiment is well designed, and if the results of the study cooperate, it may well contribute at least as much to our understanding as a comparably designed RCT.

The use of quasi-experimental methodologies requires considerably greater creativity in offsetting rival alternative hypotheses. Rather than using the brute logic of the design to mitigate alternatives, quasi-experimentalists must draw on creativity and contextual knowledge in developing controls specific to each threat that might offer a reasonable rival explanation of study outcomes. Of course, certainty is never assured, but neither is it in true experiments, as we have argued. This is the fate of the social scientist, and we have lived with this knowledge from the dawn of the discipline. Drawing on centuries of methodological innovations briefly outlined here, it is our belief that the requisite creativity to engage in problems that matter is readily at hand, and its application to important problems will enable social psychology to deliver on its promises. Is certainty assured? Assuredly not. With Heisenberg, we accept that uncertainty is our only certainty.

Acknowledgments

We are grateful to Radmila Prislin and Vanessa Hemovich for their incisive commentary on earlier drafts of this chapter.

Notes

1. We use the term *experimental* here, and in much of the early history of scientific advance up to the modern era initiated by Sir Ronald Fisher, to indicate an approach that involves the systematic controlled variation of variables between groups. It does not imply randomization, which has come to be the hallmark of contemporary experimental design.
2. Al-Biruni's insistence on replication under varying conditions, with different samples, different operations, etc., even might be interpreted as a call for methodological triangulation (Campbell, 1969a; Crano 1981), though this might be projecting more into his writing than is justified.
3. This approach was widely appealing, and remains so in today's social methodology. Indeed, it is a fair bet that operational definitions remain a feature of almost all introductory social psychological research methodology courses. However, many criticisms have been raised to an overly literal operationism—in fact, even Bridgeman came to be uneasy with an excessively rigid interpretation of his ideas. Strictly speaking, every different operation creates a different construct. This denies the utility of multiple operationism, and indeed, may deny the possibility of accumulation of reliable knowledge.

References

Adams, H. F. (1912). Autokinetic sensations. *Psychological Monographs*, *14*, 1–45.

Aiken, L. S., & West, S. G. (1991). *Multiple regression: Testing and interpreting interactions*. Newbury Park, CA: Sage.

Alhazen. (1021/2001). *Alhazen's theory of visual perception: A critical edition, with English translation and commentary, of the first three books of Alhazen's De aspectibus, the medieval Latin version of Ibn al-Haytham's Kitab al-Manazir* (A. M. Smith, Ed.). Philadelphia, PA: American Philosophical Society.

Anderson, C. A. (1989). Temperature and aggression: Ubiquitous effects of heat on occurrence of human violence. *Psychological Bulletin*, *106*, 74–96.

Anon. (1990). Obituary: Sewall Wright 1889–1988. *Biometrics*, *46*(1), 277–279.

Avicenna. (1974). *Avicenna's commentary on the Poetics of Aristotle: a critical study with an annotated translation of the text* (I. M. Dahiyat, trans.). Leiden, The Netherlands: Brill.

Bacon, F. (1620/1855). *The novum organon; or, a true guide to the interpretation of nature* (G. W. Kitchin, trans.). Oxford, UK: Oxford University Press.

Bacon, R. (1267/1859). *Opus tertium, Opus minus, Compendium philosophiae* (J. S. Brewer, trans.). London, UK: Longman, Green, Longman, & Roberts.

Bacon, R. (1267/1972). *Opus majus* (R. B. Burke, trans.). New York, NY: Russell & Russell.

Baron, R. M., & Kenny, D. A. (1986). The moderator–mediator variable distinction in social psychological research: Conceptual, strategic and statistical considerations. *Journal of Personality and Social Psychology*, *51*, 1173–1182.

Boring, E. (1950). A history of experimental psychology (2nd ed.). New York, NY: Appleton-Century-Crofts.

Boring, E. (1963). *History, psychology, and science: Selected papers*. New York, NY: Wiley.

Braver, M. W., & Braver, S. L. (1988). Statistical treatment of the Solomon four-group design: A meta-analytic approach. *Psychological Bulletin*, *104*(1), 150–154.

Breasted, J. H. (1930). *The Edwin Smith surgical papyrus, published in facsimile and hieroglyphic transliteration with translation and commentary in two volumes*. Chicago, IL: University of Chicago Press.

Bridgeman, P. W. (1927). The logic of modern physics. New York, NY: Macmillan.

Bulmer, M. (2003). *Francis Galton: Pioneer of heredity and biometry*. Baltimore, MD: Johns Hopkins University Press.

Cacioppo, J. T., & Berntson, G. G. (1992). Social psychological contributions to the decade of the brain: Doctrine of multilevel analysis. *American Psychologist*, *47*(8), 1019–1028.

Cacioppo, J. T., von Hippel, W., & Ernst, J. M. (1997). Mapping cognitive structures and processes through verbal content: The thought-listing technique. *Journal of Consulting and Clinical Psychology*, *65*(6), 928–940.

Campbell, D. T. (1969a). A phenomenology of the other one: Corrigible, hypothetical, and critical. In T. Mischel (Ed.), *Human action: Conceptual and empirical issues*. New York, NY: Academic Press.

Campbell, D. T. (1969b). Reforms as experiments. *American Psychologist*, *24*(4), 409–429.

Campbell, D. T., & Erlebacher, A. (1970). How regression in quasi-experimental evaluation can mistakenly make compensatory education look harmful. In J. Hellmuth (Ed.), *The disadvantaged child. Vol. 2. Compensatory education: A national debate*. New York, NY: Brunner-Mazel.

Campbell, D. T., & Fiske, D. W. (1959). Convergent and discriminant validation by the multitrait–multimethod matrix. *Psychological Bulletin*, *56*(2), 81–105.

Campbell, D. T., & Stanley, J. (1963). *Experimental and quasi-experimental designs for research*. Chicago, IL: Rand-McNally.

Carnap, R. (1952). *The continuum of inductive methods*. Chicago, IL: University of Chicago Press.

Cohen, J., Cohen, P., West, S. G., & Aiken, L. S. (2003). Applied multiple regression/correlation analysis for the behavioral sciences (3rd ed.). Hillsdale, NJ: Lawrence Erlbaum Associates.

Cook, T. D. (2007). "Waiting for life to arrive": A history of the regression-discontinuity design in psychology, statistics, and economics. *Journal of Econometrics*, *142*, 636–654.

Cook, T. D., & Campbell, D. T. (2002). *Quasi-experimentation: Design & analysis issues for field settings*. Boston, MA: Houghton Mifflin.

Cooper, M. L. (2010). Toward a person–situation model of sexual risk-taking behaviors: Illuminating the conditional effects of traits across sexual situations and relationship contexts. *Journal of Personality and Social Psychology*, *98*(2), 319–341.

Copernicus, N. (1543/1992). *De revolutionibus orbium coelestium (On the Revolutions of Heavenly Spheres)* (E. Rosen, trans.). Baltimore, MD: Johns Hopkins University Press.

Crano, W. D. (1981). Triangulation in social science: A cross-cultural example. In M. B. Brewer & B. E. Collins (Eds.), *Scientific inquiry and the social sciences: A volume in honor of Donald T. Campbell* (pp. 317–344). San Francisco, CA: Jossey-Bass.

Crano, W. D. (2000). The multitrait–multimethod matrix as synopsis and recapitulation of Campbell's views on the proper conduct of social inquiry. In L. Bickman (Ed.), *Research design: Donald Campbell's legacy*. Beverly Hills, CA: Sage.

Crano, W. D. (2003). Theory driven evaluation and construct validity. In S. Donaldson & M. Scriven (Eds.), *Evaluating social programs and problems: Visions for the new millennium* (pp. 145–157). Mahwah, NJ: Lawrence Erlbaum Associates.

Crano, W. D., & Brewer, M. B. (1973). *Principles of research in social psychology*, New York, NY: McGraw-Hill.

Crano, W. D., & Brewer, M. B. (2002). *Principles and methods of social research* (2nd ed.). Mahwah, NJ: Lawrence Erlbaum Associates.

Crano, W. D., & Hannula-Bral, K. A. (1994). Context/categorization model of social influence: Minority and majority influence in the formation of a novel response norm. *Journal of Experimental Social Psychology*, *30*(3), 247–276.

Crano, W. D., Kenny, D. A., & Campbell, D. T. (1972). Does intelligence cause achievement? A cross-lagged panel analysis. *Journal of Educational Psychology*, *63*(3), 258–275.

Crano, W. D., & Mendoza, J. L. (1987). Maternal factors that influence children's positive behavior: Demonstration of a structural equation analysis of selected data from the Berkeley Growth Study. *Child Development*, *58*(1), 38–48.

Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika*, *16*, 297–334.

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

Darwin, C. R. (1859). *On the origin of species by means of natural selection, or the preservation of favoured races in the struggle for life*. London: John Murray.

Denis, D. J., & Legerski, J. (2006). Causal modeling and the origins of path analysis. *Theory & Science*. Retrieved April 1, 2010, from http://theoryandscience.icaap.org

Domanski, C. W. (2004). A biographical note on Max Friedrich (1856–1887), Wundt's first PhD student in experimental psychology. *Journal of the History of the Behavioral Sciences*, *40*(3), 311–317.

Einstein, A. (2002). Galileo. In S. Hawking (Ed.), *On the shoulders of giants: The great works of physics and astronomy*. Philadelphia, PA: Running Press.

Fairchild, A. J., & MacKinnon, D. P. (2009). A general model for testing mediation and moderation effects. *Prevention Science*, *10*, 87–99.

Fechner, G. (1860/1966). *Elements of psychophysics* (D. H. Howes & E. G. Boring, Eds., with an introduction by E. G. Boring; H. E. Adler, trans.). New York, NY: Holt, Rinehart, and Winston.

Fechner, G. (1899). *Nanna oder Über das Seelenleben der Pflanzen*. Hamburg, Germany: L. Voss.

Fechner, G. (1901). *Zend-Avesta, oder, Über die dinge des himmels und des jenseits vom standpunkt der naturbetrachtung*. Hamburg, Germany: L. Voss.

Feigl, H. (1959). Philosophical embarrassments of psychology. *American Psychologist*, *14*(3), 115–128.

Fendrich, M., Lippert, A. M., Johnson, T. P., & Brondino, M. J. (2010). The association between neighborhoods and illicit drug use among adults: Evidence from a Chicago household survey. In P. L. Scheier (Ed.), *Handbook of drug use etiology: Theory, methods, and empirical findings* (pp. 461–474). Washington, DC: American Psychological Association.

Fisher, R. A. (1925/1950). *Statistical methods for research workers* (11th ed., rev.). Edinburgh, UK: Oliver & Boyd.

Fiske, D. W., & Campbell, D. T. (1992). Citations do not solve problems. *Psychological Bulletin*, *112*(3), 393–395.

Galileo. (1632/1967). *Dialogue concerning the two chief new systems, Ptolemaic and Copernican* (S. Drake, trans.; 2nd ed.). Berkeley, CA: University of California Press.

Gallup, G. G., Jr., Frederick, M. J., & Pipitone, R. N. (2008). Morphology and behavior: Phrenology revisited. *Review of General Psychology*, *12*(3), 297–304.

Galton, F. (1869). *Heriditary genius*. London, UK: Macmillan.

Galton, F. (1874). English men of science: Their nature and nurture. London, UK: Macmillan.

Galton, F. (1886). Regression towards mediocrity in hereditary stature. *Journal of the Anthropological Institute of Great Britain and Ireland*, *15*, 246–263.

Galton, F. (1888). Co-relations and their measurement. *Proceedings of the Royal Society*, *45*, 135–145.

Galton, F. (1889). *Natural inheritance*. London, UK: Macmillan.

Galton, F. (1890). Kinship and correlation. *North American Review*, *150*, 419–431.

Galton, F. (1895). *Finger print directories*. London, UK: Macmillan.

Galton, F. (1909). *Essays in eugenics*. London: Eugenics Education Society.

Gergen, K. J. (1973). Social psychology as history. *Journal of Personality and Social Psychology*, *26*(2), 309–320.

Glass, G. V., McGaw, B., & Smith, M. L. (1981). *Meta-analysis in social research*. Newbury Park, CA: Sage.

Glick, T. F., Livesey, S. J., & Wallis, F. (Eds.). (2005). *Medieval science, technology, and medicine: An encyclopedia*. New York, NY: Routledge.

Greenberg, D. H., & Shroder, M. (2004). *The digest of social experiments* (3rd ed.). Washington, DC: Urban Institute Press.

Haney, C., & Zimbardo, P. (1998). The past and future of U.S. prison policy: Twenty-five years after the Stanford Prison Experiment. *American Psychologist*, *53*(7), 709–727.

Harmon-Jones, E., & Devine, P. G. (2003). Introduction to the Special Section on Social Neuroscience: Promise and caveats. *Journal of Personality and Social Psychology*, *85*(4), 589–593.

Heck, R. H., & Thomas, S. L. (2008). *An introduction to multilevel modeling techniques*. Thousand Oaks, CA: Sage.

Hovland, C. I., Janis, I. L., & Kelley, H. H. (1953). *Communication and persuasion*. New Haven, CT: Yale University Press.

Huesmann, L. R., Dubow, E. F., Eron, L. D., Boxer, P., Huston, A. C., & Ripke, M. N. (2006). Middle childhood family-contextual and personal factors as predictors of adult outcomes. In A. C. Huston & M. N. Ripke (Eds.), *Developmental contexts in middle childhood: Bridges to adolescence and adulthood* (pp. 62–86). New York, NY: Cambridge University Press.

Huesmann, L. R., Eron, L. D., & Dubow, E. F. (2002). Childhood predictors of adult criminality: Are all risk factors reflected in childhood aggressiveness? *Criminal Behaviour and Mental Health*, *12*(3), 185–208.

Huesmann, L. R., Moise-Titus, J., Podolski, C.-L., & Eron, L. D. (2003). Longitudinal relations between children's exposure to TV violence and their aggressive and violent behavior in young adulthood: 1977–1992. *Developmental Psychology*, *39*(2), 201–221.

Huff, T. (2003). *The rise of early modern science: Islam, China, and the west*. New York, NY: Cambridge University Press.

Hull, C. L. (1916). Applied aspects of social psychology. *Psychological Bulletin*, *13*(12), 460–462.

Hull, C. L. (1943). *Principles of behavior: An introduction to behavior theory*. New York, NY: Appleton-Century-Crofts.

Hull, C. L. (1951). *Essentials of behavior*. New Haven, CT: Yale University Press.

Hume, D. (1748/1963). Enquiries concerning the human understanding and concerning the principles of morals. Oxford, UK: Clarendon Press.

Hume, D. (1772). *Essays and treatises on several subjects*. London, UK: T. Cadell.

Hunter, J. E., Schmidt, F. L., & Jackson, G. B. (1982). *Meta-analysis: Cumulating research findings across studies*. Beverly Hills, CA: Sage.

Jőreskog, K. G. (1973). A general method for estimating a linear structural equation system. In A. S. Goldberger & O. D. Duncan (Eds.), *Structural equation models in the social sciences* (pp. 85–112). New York, NY: Seminar Press.

Kahle, L. R., & Berman, J. J. (1979). Attitudes cause behaviors: A cross-lagged panel analysis. *Journal of Personality and Social Psychology*, *37*(3), 315–321.

Kant, I. (1781/1887). *Critique of pure reason* (J. M. D. Meiklejohn, trans.). London, UK: Bell.

Kenny, D. A. (1975). Cross-lagged panel correlation: A test for spuriousness. *Psychological Bulletin*, *82*(6), 887–903.

Kenny, D. H. (1979). *Correlation and causality*. New York, NY: Wiley.

Kim, U. (1999). After the "crisis" in social psychology: The development of the transactional model of science. *Asian Journal of Social Psychology*, *2*, 1–19.

Krause, M. S., & Howard, K. I. (2003). What random assignment does and does not do. *Journal of Clinical Psychology*, *59*, 751–766.

Kuder, G. F., & Richardson, M. W. (1937). The theory of the estimation of test reliability. *Psychometrika*, *2*, 151–160.

Kuhn, T. S. (1970). The structure of scientific revolutions (2nd ed.). Chicago, IL: University of Chicago Press.

Lewin, K., French, J. R. P., Jr., Hendry, C., Deets, L. E., Zander, A., Lippitt, R., et al. (1945). The practicality of democracy. In G. Murphy (Ed.), *Human nature and enduring peace: Third yearbook for the Society for the Psychological Study of Social Issues* (pp. 295–347). Boston, MA: Houghton Mifflin.

Lewin, K., Heider, F., & Heider, G. M. (1936). *Principles of topological psychology*. New York, NY: McGraw-Hill.

Likert, R. (1932). A technique for the measurement of attitudes. *Archives of Psychology*, *22*(140), 1–55.

Lindberg, D. C. (1992). The beginnings of Western science: The European scientific tradition in philosophical, religious, and institutional context, 600 B.C. to A.D. 1450. Chicago, IL: University of Chicago Press.

Lord, F. M. (1980). *Applications of item response theory to practical testing problems*. Mahwah, NJ: Lawrence Erlbaum Associates.

Lyons, W. (1983). The transformation of introspection. *British Journal of Social Psychology*, *22*(4), 327–342.

MacKinnon, D. P., Fairchild, A. J., & Fritz, M. S. (2007). Mediation analysis. *Annual Review of Psychology*, *58*, 593–614.

Mandler, G. (2007). *A history of modern experimental psychology: From James and Wundt to cognitive science*. Cambridge, MA: MIT Press.

Mankiewicz, R. (2000). *The story of mathematics*. Princeton, NJ: Princeton University Press.

McCullough, M. E., Luna, L. R., Berry, J. W., Tabak, B. A., & Bono, G. (2010). On the form and function of forgiving: Modeling the time–forgiveness relationship and testing the valuable relationships hypothesis. *Emotion*, *10*(3), 358–376.

McGuire, W. J. (1973). The yin and yang of progress in social psychology: Seven koan. *Journal of Personality and Social Psychology*, *26*(3), 446–456.

McGuire, W. J. (2003). Doing psychology my way. In R. J. Sternberg (Ed.), *Psychologists defying the crowd: Stories of those who battled the establishment and won* (pp. 119–137). Washington, DC: American Psychological Association.

Michalopoulus, C., Bloom, H. S., & Hill, C. J. (2004). Can propensity-score methods match the findings from a random assignment evaluation of mandatory welfare-to-work programs? *Review of Economics and Statistics*, *86*, 156–179.

Moscovici, S. (1988). Notes towards a description of social representations. *European Journal of Social Psychology*, *18*(3), 211–250.

Moscovici, S. (2008). Psychoanalysis: Its image and its public. Oxford, UK: Polity Press.

Moscovici, S., Duveen, G., & Flick, U. (1998). The history and actuality of social representations. In U. Flick (Ed.), *The psychology of the social* (pp. 209–247). New York, NY: Cambridge University Press.

Muller, D., Judd, C. M., & Yzerbyt, V. Y. (2005). When moderation is mediated and mediation is moderated. *Journal of Personality and Social Psychology*, *89*, 852–863.

Newton, I. (1687/1999). *Philosophiae Nauralis Principia Mathematica* (I. B. Cohen & A. Whitman, trans., 3rd ed.). Berkeley, CA: University of California Press.

Nisbett, R. E., & Wilson, T. D. (1977). Telling more than we can know: Verbal reports on mental processes. *Psychological Review*, *84*(3), 231–259.

Orne, M. T., & Scheibe, K. E. (1964). The contribution of nondeprivation factors in the production of sensory deprivation effects: The psychology of the 'panic button'. *Journal of Abnormal and Social Psychology*, *68*(1), 3–12.

O'Rourke, K. (2007). An historical perspective on meta-analysis: Dealing quantitatively with varying study results. *Journal of the Royal Society of Medicine*, *100*, 579–582.

Osgood, C. E., Suci, G. J., & Tannenbaum, P. H. (1957). *The measurement of meaning*. Chicago, IL: University of Illinois Press.

Pearson, K. (1896). Mathematical contributions to the theory of evolution. III. Regression, heredity and panmixia. *Philosophical Transactions of the Royal Society, Series A*, *187*, 253–318.

Pearson, K. (1904). Report on certain enteric fever inoculation statistics. *British Medical Journal*, *3*, 1243–1246.

Pearson, K. (1914). On certain errors with regard to multiple correlation occasionally made by those who have not adequately studied this subject. *Biometrika*, *10*, 181–187.

Pearson, K. (1920). Notes on the history of correlation. *Biometrika*, *13*, 25–45.

Peirce, C. S. (1877). The fixation of belief. *Popular Science Monthly*, *12*, 1–15.

Peirce, C. S. (1878). How to make our ideas clear. *Popular Science Monthly*, *12*, 286–302.

Peirce, C. S., & Jastrow, J. (1885). Of small differences in sensation. *Memoirs of the Academy of Sciences*, *3*, 73–83.

Pelz, D. C., & Andrews, F. M. (1964). Detecting causal priorities in panel study data. *American Sociological Review*, *29*, 836–848.

Petty, R. E., Cacioppo, J. T., Strathman, A. J., & Priester, J. R. (2005). To think or not to think: Exploring two routes to persuasion. In T. C. Brock & M. C. Green (Eds.), *Persuasion: Psychological insights and perspectives* (2nd ed., pp. 81–116). Thousand Oaks, CA: Sage.

Plato. (1892). *Protagoras. With commentary of Herman Sauppe* (J. A. Towle, trans.). Boston, MA: Ginn.

Popper, K. R. (1959). *The logic of scientific discovery*. New York, NY: Basic Books.

Popper, K. R. (1962). *Conjectures and refutations; the growth of scientific knowledge*. New York, NY: Basic Books.

Popper, K. R. (1963/1984). *Conjectures and refutations: The growth of scientific knowledge* (4th ed.). London, UK: Routledge and Kegan Paul.

Preacher, K. J., & Hayes, A. F. (2008). Asymptotic and resampling strategies for assessing and comparing indirect effects in multiple mediator models. *Behavior Research Methods*, *40*, 879–891.

Radford, J. (1974). Reflections on introspection. *American Psychologist*, *29*(4), 245–250.

Reichardt, C. S., Trochim, W. M. K., & Cappelleri, J. C. (1995). Reports of the death of regression-discontinuity analysis are greatly exaggerated. *Evaluation Review*, *19*(1), 39–63.

Rosenbaum, P. R., & Rubin, D. R. (1983). The central role of the propensity score in observational studies for causal effects. *Biometrika*, *70*, 41–55.

Rosenthal, R. (1991). *Meta-analytic procedures for social research* (rev. ed.). Thousand Oaks, CA: Sage.

Rozelle, R. M., & Campbell, D. T. (1969). More plausible rival hypotheses in the cross-lagged panel correlation technique. *Psychological Bulletin*, *71*(1), 74–80.

Russell, B. (1945). *A history of western philosophy, and its connection with political and social circumstances from the earliest times to the present day*. New York, NY: Simon and Schuster.

Sabra, A. I. (1994). *Optics, astronomy, and logic: Studies in Arabic science and philosophy*. Brookfield, VT: Variorum.

Sabra, A. I. (1998). Configuring the universe: Aporetic, problem solving, and kinematic modeling as themes of Arabic astronomy. *Perspectives on Science*, *6*(3), 288–330.

Savoy, R. L. (2001). History and future directions of human brain mapping and functional neuroimaging. *Acta Psychologica*, *107*, 9–42.

Schlenker, B. R. (1974). Social psychology and science. *Journal of Personality and Social Psychology*, *29*(1), 1–15.

Shadish, W. R., & Cook, T. D. (2009). The renaissance of field experimentation in evaluating interventions. *Annual Review of Psychology*, *60*, 607–629.

Shadish, W. R., Cook, T. D., & Campbell, D. T. (2002). Experimental and quasi-experimental designs for generalized causal inference. Boston, MA: Houghton Mifflin.

Sherif, M. (1966). *The psychology of social norms*. Oxford, UK: Harper Torchbooks.

Sidani, S. (2006). Random assignment: A systematic review. In R. R. Bootzin & P. E. McKnight (Eds.), *Strengthening research methodology: Psychological measurement and evaluation*. Washington, DC: American Psychological Association.

Siegel, J. T., Alvaro, E. M., Patel, N., & Crano, W. D. (2009). '. . . you would probably want to do it. Cause that's what made them popular': Exploring perceptions of inhalant utility among young adolescent nonusers and occasional users. *Substance Use & Misuse*, *44*, 597–615.

Solomon, R. L. (1949). An extension of control group design. *Psychological Bulletin*, *46*, 137–150.

Spearman, C. (1904). 'General intelligence' objectively determined and measured. *American Journal of Psychology*, *15*, 201–293.

Stacy, A. W., Bentler, P. W., & Flay, B. R. (1994). Attitudes and health behavior in diverse populations: Drunk driving, alcohol use, binge eating, marijuana use, and cigarette use. *Health Psychology*, *13*, 73–85.

Stigler, S. M. (1989). Francis Galton's account of the invention of correlation. *Statistical Science*, *4*(2), 73–79.

Strube, M. S. (1991). Small sample failure of random assignment. A further examination. *Journal of Consulting and Clinical Psychology*, *59*, 346–350.

Sudman, S., Bradburn, N. M., & Schwarz, N. (1996). *Thinking about answers: The application of cognitive processes to survey methodology*. San Francisco, CA: Jossey-Bass.

Thurstone, L. L. (1928). Attitudes can be measured. *American Journal of Sociology*, *33*, 529–554.

Thurstone, L. L. (1931a). The measurement of change in social attitude. *Journal of Social Psychology*, *2*, 230–235.

Thurstone, L. L. (1931b). The measurement of social attitudes. *Journal of Abnormal and Social Psychology*, *26*(3), 249–269.

Titchener, E. B. (1899a). *An outline of psychology* (rev. ed.) New York, NY: Macmillan.

Titchener, E. B. (1899b). *A primer of psychology* (rev. ed.). New York, NY: Macmillan.

Titchener, E. B. (1901–1905). *Experimental psychology: A manual of laboratory practice*. New York, NY: Macmillan.

Tolman, E. C. (1938). The determiners of behavior at a choice point. *Psychological Review*, *45*, 1–41.

Trochim, W. M. K. (1984). *Research design for program evaluation: The regression-discontinuity approach*. Newbury Park, CA: Sage.

Trochim, W. M. K. (2005). *Research methods: The concise knowledge base*. Cincinnatti, OH: Atomic Dog Publishers.

Ullman, J. B., & Bentler, P. M. (2003). Structural equation modeling. In J. A. Schinka & W. F. Velicer (Eds.), *Handbook of psychology. Volume 2: research methods in psychology* (pp. 607–634). Hoboken, NJ: Wiley.

Vermersch, P. (1999). Introspection as practice. *Journal of Consciousness Studies*, *6*, 17–42.

Vul, E., Harris, C., Winkielman, P., & Pashler, H. (2009a). Puzzlingly high correlations in fMRI studies of emotion, personality, and social cognition. *Perspectives on Psychological Science*, *4*(3), 274–290.

Vul, E., Harris, C., Winkielman, P., & Pashler, H. (2009b). Reply to comments on "Puzzlingly high correlations in fMRI studies of emotion, personality, and social cognition". *Perspectives on Psychological Science*, *4*(3), 319–324.

Vygotsky, L. S. (1927/1997). The historical meaning of the crisis in psychology: A methodological investigation. In R. W. Rieber & J. Wollock (Eds.), *The collected works of L. S. Vygotsky* (Vol. 3, *Problems of the theory and history of psychology*, pp. 233–344). New York, NY: Plenum.

Whitaker, H. A. (2000). Phrenology. In A. E. Kazdin (Ed.), *Encyclopedia of psychology, Vol. 6* (pp. 188–191). Washington, DC and New York, NY: American Psychological Association and Oxford University Press.

Wright, S. (1918). On the nature of size factors. *Genetics*, *3*, 367–374.

Wright, S. (1920). The relative importance of heredity and environment in determining the piebald pattern of guinea-pigs. *Proceedings of the National Academy of Sciences of the United States of America*, *6*, 320–332.

Wright, S. (1921). Correlation and causation. *Journal of Agricultural Research*, *20*, 557–585.

Wright, S. (1923). The theory of path coefficients: A reply to Niles's criticism. *Genetics*, *8*, 239–255.

Wu, W., West, S. G., & Hughes, J. N. (2008). Short-term effects of grade retention on the growth rate of Woodcock-Johnson III broad math and reading scores. *Journal of School Psychology*, *46*, 85–105.

Wundt, W. M. (1910). *Principles of physiological psychology*. Trans. from the 5th German ed. (1902) by E. B. Tichener. London, UK: S. Sonnenschein.

Yule, G. U. (1897). On the theory of correlation. *Journal of the Royal Statistical Society*, *60*(4), 812–854.

Zanutto, E., Lu, B., & Hornik, R. (2005). Using propensity score subclassification for multiple treatment doses to evaluate a national antidrug media campaign. *Journal of Educational and Behavioral Statistics*, *30*, 59–73.

Zimbardo, P. (2007). *The Lucifer effect: Understanding how good people turn evil*. New York, NY: Random House.

Zimbardo, P. G., Haney, C., Banks, C., & Jaffe, D. (1974). The psychology of imprisonment: Privation, power, and pathology. In Z. Rubin (Ed.), *Doing unto others: Explorations in social behavior* (pp. 61–73). Englewood Cliffs, NJ: Prentice Hall.