PALGRAVE STUDIES IN THE THEORY AND HISTORY OF PSYCHOLOGY

Psychology's Misuse of Statistics and Persistent Dismissal of its Critics

James T. Lamiell

palgrave
macmillan

Palgrave Studies in the Theory and History of Psychology

Series Editor
Thomas Teo
Department of Psychology
York University
Toronto, ON, Canada

Palgrave Studies in the Theory and History of Psychology publishes scholarly books that use historical and theoretical methods to critically examine the historical development and contemporary status of psychological concepts, methods, research, theories, and interventions. The books in the series are characterised by an emphasis on the concrete particulars of psychologists' scientific and professional practices, together with a critical examination of the assumptions that attend their use. These examinations are anchored in clear, accessible descriptions of what psychologists do and believe about their activities. All the books in the series share the general goal of advancing the scientific and professional practices of psychology and psychologists, even as they offer probing and detailed questioning and critical reconstructions of these practices.

More information about this series at
http://www.palgrave.com/gp/series/14576

James T. Lamiell

Psychology's Misuse of Statistics and Persistent Dismissal of its Critics

palgrave
macmillan

James T. Lamiell
Georgetown University
Washington, DC, USA

Palgrave Studies in the Theory and History of Psychology
ISBN 978-3-030-12130-3 ISBN 978-3-030-12131-0 (eBook)
https://doi.org/10.1007/978-3-030-12131-0

Library of Congress Control Number: 2018968356

This Palgrave Macmillan imprint is published by the registered company Springer Nature Switzerland AG
The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

This book is dedicated to those Georgetown University colleagues who, over my 36 years at that institution, nourished my efforts both by their encouragement and by their insightful critical appraisals of my work.

Foreword

Most major associations of psychologists have adopted some version of the scientist–practitioner model, first advanced by the American Psychological Association in Boulder, Colorado in 1949. This model proposed to train clinical, applied psychologists in research and scientific practice to ensure that they themselves and the psychological interventions they deliver are well grounded in psychological science. More recently, the APA 2005 Presidential Task Force on Evidence-Based Practice further integrated psychological science with professional psychological practice by insisting that effective psychological intervention must be based on the best available empirical research, defined as 'scientific results related to intervention strategies, assessment, clinical problems, and patient populations in laboratory and field settings as well as relevant results of basic research in psychology and related fields' (2006, p. 274). Acknowledging the importance of 'multiple sources of scientific evidence,' including the clinical expertise of applied psychologists, the Task Force concluded that 'systematic and broad empirical inquiry—in the laboratory and in the clinic—will point the way toward best practice in integrating best evidence' (p. 280). Of particular interest for readers of this book is a statement made by the task force in the penultimate paragraph of their report: 'The application of research evidence to a particular patient always involves probabilistic inferences' (p. 280).

James Lamiell argues herein that psychological research that reports descriptive and inferential statistical analyses of aggregated data drawn from many individuals is ***not*** relevant to understanding individual persons or how any individual person might react to the variables and contexts of psychological research as reported and generalized. This is the case whether or not attempts to categorize, predict, or comprehend any particular individual are absolute or probabilistic. The aggregated data of psychological science, as typically practiced by psychological researchers, cannot be known to apply to any individual person. If Lamiell is correct and his argument valid, any claims by psychologists or their organizations to the contrary must be false and psychology should not be understood as a discipline that has anything scientifically authoritative to say about individuals. In arguing clearly, concisely, and coherently for these conclusions, Lamiell directly challenges the claims of organized psychology to knowledge that can help individual persons to achieve psychological well being. This is a bold, perhaps startling, project but one I think any psychologist or user of psychology should want to take seriously and try to understand.

What perhaps is even more startling, and possibly damning, is the revelation that Lamiell is not the first person to make such arguments. As a respected historian of psychology, as well as a theoretical and quantitative psychologist, Lamiell is well positioned to supply the historical background for his arguments. This makes for a fascinating read as he explains how the first formal psychological research conducted in Germany in the last decades of the nineteenth century made no use of aggregated data but carefully considered all responses of each and every individual who participated in that research. It was not until psychology took a firm foothold in the United States and applied Anglo-American methods of statistical analysis to aggregated data that problems in interpreting the results of such work began to arise. What the early German psychologists were after were results that were common to and true for ***all***, i.e., each one of the particular individuals they studied. What subsequently happened was that later psychologists used statistical methods that revealed what was true ***on average***. Of course, what is common to all is true for all, whereas what is true on average is not common to or true for all, and so cannot be ***known***, absent further inquiry, to be true for ***any*** particular one considered as such. As Lamiell explains, this replacement of ***all*** with ***on average*** opened the floodgates to psychological research that was more statistically sophisticated but yielded results which had to be systematically misinterpreted in order to seem informative about individuals. By filling in many of the historical details that led to this conceptual conflation and the ascendance of aggregated statistical research in psychology, Lamiell joins a number of distinguished scholars in the history of psychology who have pointed to the difficulties and conflation he describes and warned against the increasingly widespread misinterpretation of psychological research thus engendered.

When placed in their relevant historical context, the arguments Lamiell provides cannot help but raise the central question of how mainstream empirical psychology and psychologists managed to push ahead with a program of inquiry that could not possibly yield knowledge of individual persons while simultaneously promoting their research and the professional practices of applied psychology based on this research as speaking directly to individual persons. This, of course, is a question that enters a broad moral and political territory that psychologists mostly have avoided, but which Lamiell maintains cannot be ignored without risking the entire enterprise of psychological science as legitimate and beneficial.

I first approached Jim about writing this book several years ago when I was appointed Editor of a new book series, *The Palgrave Studies in the Theory and History of Psychology*, the series in which this book now appears. I had followed Jim's work from the early 1980s and was curious about why his critique of the ways in which psychologists practiced their discipline seemingly had not affected the manner in which psychological research was being conducted. One possibility was that Jim was just wrong and that those like me whom he had managed to convince were also missing some important pieces to the puzzle of how to conduct psychological inquiry in a way that would make it directly relevant to individual persons. The other possibility was that Jim was right. But in either case, why the lack of attention and response to his critique from mainstream research psychologists and their associations?

I thought there must be an interesting story to be told about this state of affairs and it was with this in mind that I asked Jim if he might consider writing a book about his ideas and his experience in trying to get psychologists to take his ideas seriously. At first, Jim was reluctant but as we talked and emailed, he began to warm to the idea. This book is the result of his 'warming.' In my opinion, it is a book that presents credible and important concerns about the status quo of psychological research and professional practice. My hope is that in presenting his concerns in this format, Jim finally will achieve what has evaded him. This is nothing more than a reasonable expectation that the product of his years of work to improve the core practices of psychological science will be acknowledged, taken up, and engaged with by psychological researchers in ways that might ameliorate psychological science itself. To this end, in his final chapter, Jim not only summarizes his critique but puts forth suggestions, both conceptual, philosophical and methodological, practical for improving psychological science in ways that will make it relevant to individual persons.

Burnaby, Canada

Jack Martin
Burnaby Mountain Chair of
Psychology Emeritus
Department of Psychology
Simon Fraser University

Reference

APA Presidential Task Force on Evidence-Based Practice. (2006). Evidence-based practice in psychology. *American Psychologist, 61,* 271–285.

Preface

Over lunch one day during the 2014 mid-winter meetings of the Society for Theoretical and Philosophical Psychology (Division 24 of the American Psychological Association), held in Atlanta, GA, the Editor of the Palgrave Macmillan series in the Theory and History of Psychology, Dr. Jack Martin, mentioned to me for the first time his interest in my contributing to the series a book along the lines of this work. I was reluctant at first. This was partly because I already had another idea for a contribution to the series. But apart from that, I feared that the work's critical perspective on long-standing methodological practices within scientific psychology would be engaged by mainstream workers to no greater extent—and perhaps even less—than had been my previous efforts along these same lines. If this proved to be so, then, in the end, I would only have diverted myself for a year or more from the other contribution I had originally envisioned.

In time, though, the ever-patient Martin persuaded me of the need within the discipline of psychology for this book. We came to agree that even if the response to it that I feared should materialize, that would, in and of itself, underscore the validity of the work's central premise, namely, that not only has mainstream thinking in scientific psychology long been dominated by interpretive practices that are invalid, but has also been impervious to thoughtful critiques of those practices. To put matters succinctly, the concern is that, beyond being conceptually *misguided*, the field has also been, at least heretofore, effectively *incorrigible*.

For a would-be scientific discipline, such incorrigibility is terribly unhealthy, and this alone would be reason enough to once again elucidate for the scholarly community the fundamental conceptual problems that by now have become embedded within psychology's very epistemological fabric. In its profoundly unsound state, psychology is proffering knowledge claims that are unwarranted, and that is a problem quite enough. But when fundamentally flawed interpretive practices persist while repeated explications of their erroneous nature are simply ignored, the problem deepens. At some point, what might once reasonably have been seen as benign mistakenness mandates regard as deliberate *duplicity*. This is a most worrisome state of affairs.

Moreover, and just because psychology is not only a basic science but also an applied science, the discipline's conceptual confusion is more than just an abstract epistemological problem. For in the applied domain, interventions in the lives of individuals justified by claims to scientific knowledge that are faulty *and known to be so* are socio-ethically problematic as well.

So the stakes here are high, and the fact of this matter is what finally persuaded me to undertake this project. The book is historical in that it offers an account of how mainstream scientific psychology's methodological canon *became* its methodological canon. After all, the aggregate statistical methods of inquiry that have come to thoroughly dominate modern psychological research played no role at all in the experimental psychology launched by Wilhelm Wundt (1832–1920), Hermann Ebbinghaus (1850–1909), and the field's other pioneers in the latter part of the nineteenth century. The book also offers a discussion of how, in the historical course of psychology's development, extant critiques of the ascendant statistical practices either have been rejoined ineffectively or have been met with complete indifference.

Alongside its historical facets, this work is philosophical in that it reiterates and further explicates the deep and irremediable conceptual flaws that are embedded in the interpretive tenets of the contemporary mainstream methodological canon. It is of vital importance to see that the problems here do not result from incompetent execution of mainstream psychology's canonical methods. The problems are inherent in those methods themselves as instruments of psychological research, and hence remain even if the methods are executed flawlessly.

The introductory chapter of the book provides an overview of its contents by expanding upon the points mentioned above. The chapter's objective is to provide the reader, from the very start, with a concise yet clear sense for the nature and gravity of the book's central concerns and their implications for scientific psychology, both basic and applied, moving forward.

Chapter 2 is autobiographical in nature, in that it recounts my own first engagement with the book's central problematic. That engagement unfolded within the context of the disposition-situation debate that dominated the literature of personality psychology throughout the 1970s and into the 1980s. As a young, untenured, and assuredly naive assistant professor, I insinuated myself into that debate with an article published in the *American Psychologist* in 1981. In that article, I explained that the central theoretical question being debated, namely, the degree of consistency over time and across situations in *individuals'* manifestations of their underlying personality characteristics, could not be addressed properly by examining the relationship between trait scores and criterion behaviors in terms of correlations between variables marking *individual differences.* By their very nature, such variables are definable only for *aggregates* of individuals, for at the level of the individual there is no empirically specifiable 'individual *difference*' to examine!

In Chapter 3, I discuss how aggregate statistical methods of investigation entered scientific psychology with the founding and development of that subdiscipline of the field devoted to the systematic study of individual and group differences. Two books published by the German philosopher and psychologist William Stern (1871–1938) were of major importance in this connection. However, other influential contemporaries of Stern, including E. L. Thorndike (1874–1949) and Hugo Münsterberg (1863–1916) played major roles as well, often representing views that differed from those of Stern in several significant respects. In time, the field proved to be more influenced by Thorndike and Münsterberg than by Stern, a development leading ultimately to the predominance of the view that statistical knowledge of populations is, at one and the same time, knowledge of the individuals within those populations.

Chapter 4 is focused on that 'other' of scientific psychology's 'two disciplines' (Cronbach, 1957), existing alongside of differential/correlational psychology, namely, experimental psychology. In particular, the discussion in that chapter documents the widespread failure of critical thinking during the first three decades or so of the twentieth century, as experimental psychologists gradually abandoned an approach whereby research findings were fully defined for individual subjects in favor of a radically different approach whereby research findings are defined by the results of statistical comparisons of group averages. Throughout this historical development, mainstream psychology's experimentalists widely and uncritically assumed full epistemic continuity between the two approaches, so that the newly embraced statistical approach could be regarded as no less suited than the original—and decidedly non-statistical—approach to the pursuit of scientific psychology's original overarching knowledge objective, namely, the discovery of general laws governing the psychological functioning of individual persons. Chapter 4 concludes with a long overdue explication of the implicit conceptual requirements of that assumption.

Chapter 5 offers a critical examination of certain widely accepted interpretive and discursive practices within the mainstream of scientific psychology that sustain the erroneous belief that statistical knowledge of aggregates of individuals also conveys knowledge of the individuals within those aggregates. The objective is to make clear what is problematic about prevailing understandings of (a) the meanings of correlations between variables marking differences between individuals, (b) the prediction and explanation of individual psychological doings, and (c) claims to probabilistic knowledge about individuals. The problem is not simply that an aggregate statistical index cannot be taken to represent *every* individual within the aggregate, a point most mainstream thinkers readily concede, but is, rather, that except under circumstances theoretically conceivable but never realized empirically, such an index cannot *properly* be taken to represent *any* individual within the aggregate. Statistical knowledge of aggregates of individuals is, quite literally, knowledge of *no one*, and, as such, is knowledge of a sort fundamentally and irremediably ill-suited to the objective of advancing our scientific grasp of the psychological doings of living, breathing *some ones.* The need for paradigmatic change becomes evident.

In Chapter 6, the discussion begins with a definition of 'statisticism' as the word I have invented to refer to the virtually unshakable faith held by mainstream psychological researchers in the power of aggregate statistical research methods to generate scientific knowledge about the psychological doings of individuals. Attention is then focused for the remainder the chapter on the socio-ethical facet of this—*ism*. Two examples are used to highlight the issues that arise in this domain; one is drawn from the province of the evidence-based practice movement in psychology, and the other from the province of psychological testing in the service of preemployment screening. The latter portion of the chapter then focuses on broader considerations of a socio-ethical nature, with particular attention being devoted to the consequences of reducing persons to things. I argue that this is an inevitable consequence of regarding persons as instantiations of the categories that define the variables that are the actual focus of investigation in aggregate-level studies.

Chapter 7, which concludes the book, begins with a brief review of the major historical developments, elaborated in previous chapters, that landed mainstream psychology in its current epistemic predicament. I then discuss a variety of matters about which maximum clarity will be essential if the obdurate resistance that has thus far prevailed within psychology's mainstream to the needed change in its investigative methods is ever finally to be overcome. I note that although psycho-demographic inquiry cannot qualify as psychology, I explain both that and why this is not an argument that such inquiry is of no merit. Against the notion that no viable alternatives to the currently dominant research paradigm exist, attention is directed to a promising framework called 'observation oriented modeling.' I emphasize further that the present argument is not against the use of quantitative methods in psychology per se, but is rather against the use of aggregate statistical methods as a way of advancing knowledge of individuals. I then discuss the need to overcome enduring misconceptions in contemporary psychologists' understanding of the concepts of 'nomothetic' and 'idiographic.' The chapter concludes with a plea for reviving a conception of psychology as both a natural science and a human science, noting that such a revival will bring with it a recognition of the need for and scientific legitimacy of both quantitative and qualitative research methods.

By its very nature, this book revisits ideas and arguments that I have articulated in previous publications. This is consistent with the wish of the series Editor that I place the entire argument in its most up-to-date formulation between the covers of a single book. Hence, readers familiar with my earlier works will find here material that, in one form or another, they have encountered before. In particular, those of my previous writings on which I have leaned most extensively in this work include (1) a 1987 book titled *The Psychology of Personality: An Epistemological Inquiry* (New York: Columbia University Press); (2) a 2003 book *Beyond Individual and Group Differences: Human Individuality, Scientific Psychology, and William Stern's Critical Personalism* (Thousand Oaks, CA: Sage Publications); (3) a 2007 journal article titled 'On sustaining critical discourse with mainstream personality investigators: Problems and prospects' (*Theory and Psychology*, volume 17, pp. 169–185); (4) a 2016 book chapter titled 'On the concept of "effects" in psychological experimentation: A case study in the need for conceptual clarity and discursive precision,' in R. Harré, and F. Moghaddam (Eds.), *Questioning Causality: Scientific Explorations of Cause and Consequence Across Social Contexts* (pp. 83–102) (Santa Barbara, CA: Praeger); and (5) a 2017 book chapter, authored in collaboration with Jack Martin, titled 'The incorrigible science,' in H. Macdonald, D. Goodman, and B. Becker (Eds.), *Dialogues at the Edge of American Psychological Discourse* (pp. 211–244) (London: Palgrave Macmillan). In the present work, I have endeavored not only to extend arguments advanced in those various publications, but also to refine some of the prose that I used in earlier writings in hopes of making the arguments more complete, accessible, and compelling.

The intended audience for this book is one comprised mainly of advanced undergraduate students, graduate students, persons actively engaged in the conduct of psychological research, and of other scholars who, while perhaps not psychologists themselves, are concerned with the historical, philosophical, and methodological foundations of psychological research. Of course, I also hope that some of what I have to say here will reach many more individuals, groups, and institutions who make use of the findings of psychological inquiry in their daily activities, such as psychotherapists, educators, administrators, planners, parents, and citizens. Though by no means a textbook, this volume could prove pedagogically useful to instructors offering advanced undergraduate or graduate-level coursework in the history, philosophy, and/or research methods of psychology.

Goodyear, Arizona, USA James T. Lamiell

Reference

Cronbach, L. J. (1957). The two disciplines of scientific psychology. *American Psychologist, 12*, 671–684.

Contents

List of Figures

1

Introduction: Mainstream Psychology's Worrisome Incorrigibility

My central concern in this work is with the deep, abiding, and highly problematic confusion that has become, by now, part of the very fabric of mainstream thinking about the proper use of statistical methods in psychological research. It is a confusion that has persisted for decades, and one that the discipline as a whole has thus far simply refused to acknowledge. The confusion itself can be stated quite simply: it consists of the notion that statistical knowledge about variables defined only for aggregates of individuals entitles scientifically authoritative claims to knowledge about the individuals within those aggregates.

The consequences of this confusion are serious and far-reaching for scientific psychology, and are both epistemic and socio-ethical in nature. In the epistemic domain, the confusion routinely leads psychological investigators to overstate what they may justifiably claim to *know* about individuals on the basis of their research findings. The result is bad science. In the socio-ethical domain, the same confusion routinely leads psychological investigators to overstate what they may justifiably *do*, or endorse doing, under the banner of scientifically licensed interventions, in the lives of individuals. The result is bad professional practice.

J. T. Lamiell, *Psychology's Misuse of Statistics and Persistent Dismissal of its Critics*, Palgrave Studies in the Theory and History of Psychology,
https://doi.org/10.1007/978-3-030-12131-0_1

The pervasiveness of this confusion within contemporary scientific psychology would be troublesome enough were it a phenomenon of but recent vintage. What makes matters profoundly more worrisome is the fact that thoughtful and trenchant critiques of invalid interpretations of aggregate-level statistics have surfaced periodically in the literature of the discipline dating back at least to the 1950s—and comparable critiques can be found dating well back into the nineteenth century if one consults scholarly literature outside of psychology (cf. Porter, 1986). Astonishingly, however, while those critiques have never been successfully rebutted, neither have they been duly heeded within the mainstream of psychology as guidelines to corrective methodological practices. Instead, the basic assumption that population-level research findings can warrant knowledge claims about—and hence justify systematic interventions in the lives of—individuals within those populations have continued to prevail. The long-entrenched methodological canon has thus endured as if the confusions, invalidities, and dubious intervention justifications resulting from and sanctioned by that canon had never been challenged to begin with, much less compellingly defended in the face of the challenges. In this respect, mainstream psychology has proven itself to be, in a word, *incorrigible*. It is difficult to imagine an intellectually more worrisome state of affairs for a putatively scientific field.

In subsequent chapters of this book, major historical developments that have contributed to the emergence of these conceptual difficulties within psychology will be discussed at some length. In this chapter, however, some preliminary observations of an historical nature are in order so as to facilitate the reader's orientation to the material that follows.

Some Orienting Historical Context

The General and the Aggregate

In 1955, psychologist David Bakan (1921–2004) published a brief, two-page commentary in the journal *Perceptual and Motor Skills* in which he averred the following:

> The failure to distinguish between general-type and aggregate-type propositions is at the root of a considerable amount of confusion which currently prevails in psychology. There are important differences in the research methods appropriate to these two types of propositions. The use of methods which are appropriate to the one type in the establishment and confirmation of the other, leads to error. (Bakan, 1955, p. 211)

It seems, however, that Bakan's (1955) altogether valid argument fell on deaf ears, so that more than a decade later he found reason to make the same point again, this time embedded within a broader discussion of fallacious thinking that psychologists were routinely indulging in the course of null hypothesis significance testing (Bakan, 1966). In this latter article, published in the high-profile journal *Psychological Bulletin*, Bakan cast some light on the historical sources of the confusion between aggregate-type and general-type propositions that he had identified in the earlier article. In that connection, he emphasized that at its founding in the latter part of the nineteenth century, experimental psychology was clearly aimed at establishing the validity of

> propositions concerning the nature of man in *general*—propositions of a general nature, with *each individual a particular in which the general is manifest*. This is the kind of psychology associated with the traditional experimental psychology of Fechner, Ebbinghaus, Wundt, and Titchener. (Bakan, 1966, p. 433, emphasis in original)

Here, as before, Bakan (1966) was entirely correct.[1] The general experimental psychology founded by Wundt and prosecuted by him and the other luminaries Bakan mentioned was—and was explicitly referred to at the time as—an *individual* psychology. This was true even as it was recognized that the overarching objective of the discipline was knowledge of the *general laws* presumed to govern various aspects of mental life. As I have pointed out in other discussions of this point (see, for example, Lamiell, 2015, 2016), any apparent contradiction between an investigative approach whereby experimental findings are defined for individuals even as the scientific quest is for knowledge of general laws is clearly resolved once one appreciates the meaning that the early experimentalists attached to the notion of 'general.'

In German, the language of the country where experimental psychology was first formally established, the word for 'general' in the sense relevant here was (and remains) *allgemein.* That word is a contracted form of the expression *allen gemein,* which means 'common to all.' The *allgemeine Gesetze* or 'general laws' of mental life that psychology's founding fathers sought to discover experimentally would thus be laws found to be *common to all* of the investigated individuals and, at least presumptively pending subsequent experimental outcomes, to non-investigated individuals as well.[2] What scientific psychology's first experimentalists were *not* in search of were empirical regularities found merely to be 'true on average' for collections of individuals.

What must also be appreciated is the fact that the individual psychology of the early experimentalists was *not* a discipline concerned in any way at all with *individuality*. Precisely because the knowledge sought was that of laws that would prove generalizable *across* individuals—i.e., from one individual to another and so on—there was no programmatic interest in any empirical indicators of personal idiosyncrasies, i.e., characteristics that, by definition, would not be generalizable across individuals. In this connection, Bakan (1966) aptly cited the anecdote mentioned by E.G. Boring (1886–1968) in his *History of Experimental Psychology* (Boring, 1950), according to which the suggestion to Wundt by his student, James McKeen Cattell (1860–1944), that he be permitted to study individual differences in reaction times was abruptly dismissed by Wundt as *ganz amerikanisch* or '*so* American.'

It was into just this breach that William Stern (1871–1938) stepped in 1900 with the publication of a book proposing a new subdiscipline for scientific psychology that he called 'differential' psychology (Stern, 1900). In that work, Stern boldly proclaimed 'individuality' as the 'problem of the twentieth century' ("*Problem des zwanzigsten Jahrhunderts*"; Stern, 1900, p. 1). What Stern meant by '*Problem*' in this particular context is better rendered in English by the word 'challenge,' for what he wished to convey was the idea that in order for scientific psychology to be viable on through the twentieth century (and, presumably, beyond), some accommodation would have to be made to the possibility—indeed, the certainty—that any given individual's psychological 'doings' would be scientifically graspable only partly in terms of laws common to all, and would therefore have to be viewed partly in terms not applicable to all—and possibly not even to any—other individuals.[3] Stern's view was that a subdiscipline of psychology devoted to the study of nonrandom individual and group differences in any psychological domain would highlight the need for, and thus serve to facilitate the eventual realization of, a scientific psychology worthy of the challenge of individuality.

In the meantime, the differential psychology would make feasible an applied psychology that was beyond the scope of the field's experimental methods. Stern's countryman and friend, Hugo Münsterberg (1863–1916), would eventually address himself to this point as follows:

> The study of individual differences itself is not applied psychology, but it is the presupposition without which applied psychology would have remained a phantom. As long as experimental psychology remained essentially a science of the mental laws, common to all human beings, an adjustment to the practical demands of daily life could hardly come in question. With such general laws we could never have mastered the concrete situations of society, because we should have had to leave out of view the fact that there are gifted and ungifted, intelligent and stupid, sensitive and obtuse, quick and slow, energetic and weak. (Münsterberg, 1913, pp. 9–10)

Importantly, there is clear evidence in Stern's (1900) book to justify his claim, stated many years later in an intellectual autobiography, that 'even then (i.e., even in 1900), I could see that true individuality, the understanding of which was my ultimate objective, cannot be grasped through the channels of differential psychology' (Stern, 1927, p. 142, parentheses added; cf. Lamiell, 2003). Stern made himself even clearer on this point in a sequel to the 1900 book that was published in 1911 under the title (in translation) *Methodological Foundations of Differential Psychology* (Stern, 1911). In that work, Stern made explicit his commitment to the view that the study of *variables* with respect to which individuals have been differentiated yields knowledge of the *variables*, and *not* knowledge of the individuals who have been differentiated in terms of those variables. However, other highly influential differential psychologists of the time, including both Münsterberg and E. L. Thorndike (1874–1949), broke ranks with Stern on this crucial point (Chapter 3 offers a more detailed discussion of this matter), and as the overwhelming majority of mainstream investigators in psychology aligned themselves (wittingly or otherwise) with Thorndike and Münsterberg rather than with Stern, the discipline veered in a direction that would lead eventually to the prevalence of the confusion that concerned Bakan.

In his 1966 article, Bakan explicitly contrasted the original experimental psychology, seeking knowledge of the general laws governing various aspects of individuals' psychological doings, with the later-born differential psychology by pointing out that 'the basic datum for an individual differences approach is not anything that characterizes *each* of two subjects but rather *the difference between them.* For this latter tradition, it is the *aggregate* which is of interest, and not the general' (Bakan, 1966, p. 433, emphasis in original).

This last statement by Bakan (1966) expresses exactly the logical reality that Stern (1911) had recognized in contrasting knowledge about individual differences variables with knowledge about individuals. In studies of individual differences variables, Stern pointed out, individuals are merely placeholders, of use to the researcher only as a means to the end of instantiating empirically the various discrete categories or levels of the variables whose statistical properties within populations (means, variances, covariances) are the actual foci of the research (cf. Stern, 1911, especially p. 318). The individual *differences* under investigation in such inquiries have no existence at the level of the *individual*. This is why, for example, a researcher studying sex differences within some population of children (e.g., with respect to auditory acuity or reaction time to visual stimuli, or any other psychological phenomenon) should never be led to suppose that for any one of the little kiddies portrayed in Fig. 1.1 there could be found a sex *difference* tucked inside his or her diaper!

As the cartoon caption claims, there *is* a difference! However, clear thinking about this requires mindfulness about just where 'there' is. The differences are 'there' *in the data*, and not 'there' in the physical equipment of any one of the individual research subjects who have contributed to those data. Stern grasped this point fully. Münsterberg

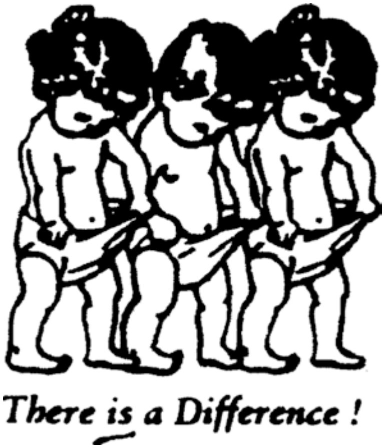

Fig. 1.1 Cartoon with caption 'Where is there?'

and Thorndike did not, nor would the overwhelming majority of their contemporaries and successors in scientific psychology. Therein lies the problem.

Bakan represented an exception to the majority view. Having distinguished knowledge of the general from knowledge of the aggregate, and then properly locating those two quite different knowledge objectives within the original experimental and emergent correlational traditions of psychology, respectively, he was able to reiterate his central point, namely, that mainstream psychological investigators had fallen into the practice of relying on their statistical analyses to secure knowledge of aggregates *as if* the results of those analyses *also* conveyed knowledge that could be generalized across individual subjects and hence give scientific warrant for general-type propositions. In Bakan's (1966) own words: '(T)he data are treated as aggregates while the experimenter is trying to infer general propositions' (p. 433).

It is important to bear in mind here that although differential psychology was, indeed, the research domain within the nascent psychology to which aggregate statistical methods were at first restricted—those methods having had no place in the original $N=1$ experimental psychology— matters had long since changed by the time that Bakan (1955, 1966) was expressing his concerns. This is because over the first several decades of the twentieth century, experimental psychologists gradually abandoned the single-subject approach to experimentation in favor of the *treatment group* approach (Danziger, 1990). In the simplest and by far most widely used version of that approach, the randomized group design, research participants are 'sampled' from 'populations' and randomly assigned to one of two or more experimental treatment conditions defining an independent variable (IV). For each subject, a dependent variable (DV) outcome is represented numerically, and after the outcome for every subject has been recorded, statistical calculations are undertaken to determine if the difference(s) between the DV means of the two or more treatment groups is/are greater than would have been expected on the basis of chance alone (a quantity commonly estimated by computing the variability among subjects assigned to the same treatment condition). Experimental findings are thus defined not in terms of the results obtained with individual subjects, as was true in the original experimental psychology, but rather in terms of the results of analyses of *differences between* outcome *means* defined for treatment *groups*.[4]

Conceptually speaking, the departure of the treatment group form of experimentation from the original $N=1$ approach was a radical one. However, researchers continued to interpret their treatment group experimental findings as if they shed scientific light on some aspect of the psychological doings of their individual subjects, just as *had* been the case in $N=1$ inquiry. It was the illegitimacy of such interpretations to which Bakan (1955, 1966) was calling attention. Alas, the problematic interpretive practices continued despite Bakan's efforts.

A "Troublesome Paradox"

Nonetheless, the conceptual problem that Bakan identified did not go completely unrecognized by others. On the contrary, thirteen years after the publication of Bakan's (1966) article, the author of several highly regarded textbooks in research methods, Fred N. Kerlinger (b. 1910; date of passing unknown) acknowledged puzzlement over what he termed a 'troublesome paradox' (Kerlinger, 1979, p. 275) running throughout scientific psychology's research literature. Evincing no acquaintance with Bakan's writings, Kerlinger wrote:

> The unit of speech in science is always the set, the group. But behavioral scientists, and particularly psychologists, often talk as though the unit of speech were the individual. Psychological theories, for example, are sometimes enunciated as though they were explanations of what goes on inside a single individual. The social psychological scientist, for instance, may talk about the effect of perceived similarity of attitude toward social issues on liking for another person. In explaining the rationale of such a relation, the scientist may talk about the individual and the structure and content of his attitudes toward social issues. Or a cognitive theorist may talk about the structure of memory and its effects on certain behavior. They mean, of course, the attitudes and memories of single individuals. (Kerlinger, 1979, p. 275)

Kerlinger went on to provide two additional examples of the 'troublesome paradox.' He wrote:

> (I)n the conclusion of the report of a stimulating research study on the influence of traits as prototypes on memory, … the following sentence appears: 'Storing material in terms of its relation to a consistent conceptual schema is likely to provide *one* (Kerlinger's italics) with a more stable, less redundant memory structure.' … Here is a passage from another fine study … : 'Self-schemata are cognitive generalizations about the self, derived from past experience, that organize and guide the processing of the self-related information contained in an *individual's* (Kerlinger's italics) social experience.' In these studies the authors could only work with groups of individuals and establish the relations (between the variables) they were studying by using groups of individuals. Both authors, however, slip from the group to the individual unit of speech. They more or less have to because their theories 'explain' what is presumably inside the head of the individual. In the second study, since the hypothesized relations were supported by the empirical *group* evidence (Kerlinger's italics), the author assumes, perforce, that self-schemata existed in the brains of her individual subjects. The paradox, then, is that scientists, especially psychological scientists, must hypothesize and test relations at the group or set level when they in fact often want to talk on the individual level—and may (well) do so. (Kerlinger, 1979, pp. 275–276, italics in original; material in parentheses inserted by author to enhance clarity)

In a footnote added to the passage just quoted, Kerlinger emphasized that he was not singling out for criticism the two studies he had cited, but was rather using those studies to illustrate a widespread phenomenon. He stated: ‘Indeed, it is virtually impossible to escape individual-level talk in psychological research writing’ (Kerlinger, 1979, p. 276).

It is significant that in the above passages, Kerlinger (1979) insisted that the unit of investigation in psychological research ‘is *always*’ and ‘*must* be’ the ‘group’ or the ‘set.’ He did briefly discuss some apparent exceptions to this claim, but ultimately concluded that ‘one requires more generality than the data a single individual can provide’ (p. 278), and that ‘the claim that science is not and cannot be concerned with the individual is in general a valid one’ (p. 278).

For reasons discussed earlier in this chapter, these are claims that would have greatly surprised Wundt and Ebbinghaus and their cohorts who, together, gave birth to psychology as a laboratory science. Likewise surprising to those pioneers would have been Kerlinger’s suggestion, through his wording in certain places, that psychologists’ theoretical interest might in at least some (albeit infrequent) instances be in something other than individual-level phenomena. To the contrary, the founding fathers would have insisted that the unit of investigation in psychology’s experimental laboratories must always be the individual, and that this was so because it is only at the level of the individual where one finds the psychological phenomena of theoretical interest in those laboratory investigations.[5]

All of this said, the point of emphasis here is that Kerlinger (1979) clearly had a sense for the very problem that Bakan had identified. That is, Kerlinger’s ‘troublesome paradox’ and Bakan’s ‘aggregate-general confound’ are one and the same. Unlike Bakan, however, Kerlinger seems to have regarded the problem as rather less serious than it is, for in branding the issue a *paradox*, Kerlinger left open the possibility that a solution to the problem might one day be found *within* the constraints imposed by conventional mainstream research practices. In other words, Kerlinger seems to have believed that the conceptual gap between aggregate-level research findings and general-type theoretical objectives would one day be bridged *validly*, and it was perhaps for this reason that he issued no call for a suspension of traditional mainstream research practices in the interim. On the contrary, there is no discussion of the paradox in the third edition of his research methods textbook, *Foundations of Behavioral Research*, published in 1986, nor does that text make any reference to Bakan's work (which Kerlinger had also left un-cited in the second edition of the *Foundations* text, published in 1973).

In Kerlinger's view, it appears, 'business as usual' in the domain of psychological research practices could and should proceed right along with the persistence of the 'troublesome paradox' about which he wrote in 1979. This is, of course, exactly what has happened. And so, to date, mainstream thinking in psychology has yet to come to terms with the insight so pithily expressed in 1867 by Wundt's senior colleague at Leipzig, the polymath Moritz Wilhelm Drobisch (1802–1896):

> It is only through a great failure of understanding (that) the mathematical fiction of an average man … (can) be elaborated as if all individuals … possess a real part of whatever obtains for this average person. (Drobisch, as quoted in Porter, 1986, p. 171)

In the view of the present author, this Great Failure of Understanding has been and continues to be facilitated in no small measure by the subtle and perhaps largely unwitting conflation of two quite different understandings of the rudimentary statistical concept of *probability*. To illustrate this point and its bearing on our concerns, let us consider a very simple example.

The Blurring of a Crucial Distinction

Let us suppose that a social psychologist has carried out research indicating that among adolescent males growing up in poverty, 80% run afoul of the law by the age of 18. This finding serves as the empirical basis for the knowledge claim that among impoverished male youth the probability, p, of legal transgression by age 18 is .8.

Note that the focal knowledge claim here, $p = .8$, is a claim to acquaintance with an empirical fact pattern that has been established (however provisionally) for a *population*. It is not and cannot properly be understood to be a claim to acquaintance with an empirical fact about any individual within that population, or indeed, about any individual in any population anywhere. This means (among other things) that on the question of the probability that 'this' or 'that' or *any* particular male adolescent—whether within or outside of the target population—will (have) run afoul of the law by age 18, the empirical finding, $p = .8$ uncovered by our hypothetical social psychologist is completely and utterly silent. A probabilistic fact established for a population is, quite literally, an empirical fact about *no one*. These considerations illustrate what is called a *frequentist* understanding of probability, the essence of which is that probabilistic knowledge claims are inherently tied to the consideration of a *series* of empirical instances or events, and can never be articulated validly for any discrete instance or event within that series (cf. Venn, 1888).[6]

On a *subjectivist* understanding of probability, in contrast, probabilistic language is used in speaking of an individual or isolated event, but not to express acquaintance with an empirical *fact* about that individual or isolated event. Such language is used instead to express the strength or degree of a *belief* that one holds about that individual or isolated event. So, for example, someone acquainted with our social psychologist's research finding (perhaps the social psychologist him/herself) might well say of 'this' particular male youth, currently growing up under impoverished circumstances, something such as: 'I strongly believe …' or 'I'm pretty sure …' or 'I think it is very likely that this young person will (have) run afoul of the law by the age of 18.'

It is of crucial importance to recognize—and to appreciate the full implications of—the fact that statements of this latter sort do not advance claims to *knowledge* about the individual under discussion, and this is true even if the speaker gives as the rationale for his/her statement(s) acquaintance with some aggregate-level empirical fact pattern. Such statements remain expressions of *subjective belief*. This can be seen in the inescapable empirical fact about the specific impoverished male adolescent in question that he *either will or will not* (have) run afoul of the law by the age of 18. This is assuredly true (a) *no matter what* some social psychologist's research might have revealed about the population-level probability of impoverished adolescent males running afoul of the law, (b) *no matter what* that social psychologist or anyone else happens to believe about the adolescent in question, and (c) *no matter what* the basis for that belief might be. To underscore the conceptually crucial point here: our hypothetical social psychologist's empirical finding, $p = .8$, expresses an empirical fact pattern that has been observed for a *population*, and not an empirical fact established for 'this' or 'that' or any individual within that population. Hence, that empirical finding cannot scientifically justify a claim to *knowledge* about any *one* at all. Nor could that empirical finding warrant a claim to an objective scientific basis for intervening in the life of some particular adolescent male within the target population.

Probability statements used to express knowledge and probability statements used to express subjective beliefs are both meaningful, but they are so in two entirely different ways. Alas, mainstream psychology has proven itself unable or unwilling to respect this distinction, a recalcitrance manifested by the decades-long and discipline-wide indulgence of statements that are given the grammatical structure of knowledge claims about individuals but that are backed only by aggregate-level research findings.

In his instructive book *Statistics in Psychology: An Historical Perspective*, Cowles (1989) addressed himself directly to this confound. He wrote:

> The fact that probability has to do both with frequencies and with degrees of belief is the … epistemological duality that … we (psychologists) *blur* as we compute our statistics and speak of the confidence we have in our results. (Cowles, 1989, p. 59, emphasis and parentheses added)

Within the context of the hypothetical research example introduced above, locutions illustrative of the sort of blurring to which Cowles alluded would claim that some particular impoverished adolescent male '*has*' a high likelihood of, or '*is*' at risk for, running afoul of the law by age 18, or that he '*has*' a strong tendency toward unlawfulness, etc. Note that all of these statements incorporate wording that makes them, as they stand, not expressions of the speaker's subjective belief about an individual, but, instead, claims to objective knowledge about that individual. They state what that individual may, with objective scientific justification, be said to 'have' or to 'be.' Without doubt, such locutions serve well the goal of making population-level knowledge *seem* interpretable at the level of the individual, and in just this way make it *seem* as if population-level research may properly be regarded as advancing our scientific grasp of individual-level psychological 'doings' even though this is not truly the case. This being so, one might expect Cowles's (1989) discussion of the matter to have issued in a call to psychologists for much greater circumspection in the interpretation of their research findings, so that the blurring to which he alluded could be avoided going forward. Yet this is not what one finds. One finds instead the following:

> The fact that the answer to the question 'Who, in practice, cares (about the blurring)?' is 'Probably very few,' is based on an admittedly informal frequency analysis, but it is one in which we can believe! (Cowles, 1989, p. 59, parentheses added)

So, having pointed directly to a discursive practice within scientific psychology that seems to but actually does not solve the conceptual problem discussed by Bakan (1955, 1966) as the 'aggregate-general confound' and by Kerlinger (1979) as a 'troublesome paradox,' Cowles (1989) then effectively sanctioned that practice on the grounds that the epistemic blurring it entails had proven to be a matter of indifference among mainstream researchers up to that time. That same indifference prevails to this day. Yet, if psychology is a genuine science, and so dedicated ultimately, as are all sciences, to 'the colligation of facts and the clarification of concepts' (Machado & Silva, 2007, p. 680, with a bow to the estimable British scholar William Whewell [1794–1866]), then this paradigmatic indifference to conceptual confusion is intellectually treasonous.

Alas, over the course of the twentieth century and now well into the twenty-first, the ethos of mainstream scientific psychology has proven much more hospitable to the colligation of facts than to the clarification of concepts. Therein lies a major impediment to corrective efforts of the present sort. More than a century ago, the venerable Wilhelm Wundt presciently warned of the untoward consequences for a psychology that would neglect the conceptual aspects of its scientific mission (cf. Lamiell, 2013), and a brief glance back at that work is in order here.

Psychology's Struggle for Existence

On Wundt's (1913) Prescience

In the foreword to an essay that Wundt published in 1913 under the title (in translation) *Psychology's Struggle for Existence* (Wundt, 2013), he wrote:

> In the opinion of some, philosophy and psychology should divorce from each other. … If this matter takes the course that both parties want, philosophy will lose more than it will gain, but psychology will be damaged the most. Hence, the argument over the question of whether or not psychology is or is not a philosophical science is, for psychology, a struggle for its very existence. (Wundt, 2013, p. 197)[7]

A philosopher himself, Wundt well understood that one of the major functions of philosophical inquiry is its quest for *conceptual clarity* (cf. Bennett & Hacker, 2003). In accordance with that mission, the philosophically minded scholar asks doggedly: What do we *mean* when we say X? Wundt (2013) was concerned that to the extent that psychologists isolated themselves from philosophically minded colleagues and preoccupied themselves instead with the technical aspects of experimental design and data analysis, they would eventually lose their appreciation for—and perhaps the intellectual wherewithal to grapple with—basic conceptual questions. Along these lines, he observed that in discussions of the advisability of psychology's prospective divorce from philosophy,

> … there is a question that has hardly been touched upon but should be thought about as it is a decisive one: this is the question of the extent to which it would even be possible for the psychologist to divest himself of philosophy, and to have no need of the assistance of philosophical observations while addressing in depth psychology's own problems. Assuming that such philosophical observations would have value, the question is whether or not psychologists would be able to formulate them on their own. (Wundt, 2013, p. 198)

In the unfolding of his essay, Wundt left no doubt that his answer to this question would be 'no,' and so he worried that the impending divorce, if carried through, would result ultimately in the 'separation of psychology from precisely that domain of science that is indispensable to it' (Wundt, 2013, p. 203). He went on to argue that the questions for an education in psychology that are the most important

> … are so closely connected with the epistemological and metaphysical positions that it is inconceivable that they will at some point disappear from psychology. It is precisely this that shows clearly that psychology belongs to the philosophical disciplines, and this will remain so even after the transformation of psychology into an independent discipline. In a psychology divorced from philosophy, philosophical considerations will be latent, and so it is possible that psychologists who will have abandoned philosophy, and whose education in philosophy will therefore be deficient, will be projecting those considerations (anyway, but) only through an immature metaphysical perspective. As a result of such a separation, therefore, no one will suffer more than psychologists—and, through them, psychology. If philosophers now complain, unjustifiably, that psychology has become merely a technical rather than a purely scientific discipline, that would become even more—and more disturbingly—the case … and then the time truly will have been reached when psychologists will have made themselves into tradesmen *(Handwerker)*, and, at that, not of the most useful variety. (Wundt, 2013, p. 206, parentheses added)

Wundt knew that his views on the need within psychology for philosophically minded thinkers were not widely shared, and that was a major reason for his having written the 1913 essay, to begin with.

Despite his warnings, however, the divorce did eventually come to pass, and, at least in the United States, the ascendance of behaviorism did little to counter that development. Indeed, in his 1928 book *The Ways of Behaviorism*, J. B. Watson (1878–1958) wrote that that school of thought sounded 'a threatening note to the whole of philosophy (Watson, 1928, p. 14),' and he elaborated his point as follows:

> With the behavioristic point of view now becoming dominant, it is hard to find a place for what has been called philosophy. Philosophy is passing—has all but passed, and unless new issues arise which will give a foundation for a new philosophy, the world has seen its last great philosopher. (Watson, 1928, p. 14)

In 2019, it seems apparent that it is behaviorism rather than philosophy that has passed, a development to which B. F. Skinner (1904–1990) seems to have resigned himself in the article he completed on the evening before he passed away (Skinner, 1990). Nevertheless, there remain strong traces within psychology of the positivist-empiricist orientation that was so congenial to behaviorism (cf. Costa & Shimp, 2011), and it is perhaps at least partly for this reason that behaviorism's decline has not led mainstream scientific psychology to a renewed appreciation for the importance to the discipline's overall scientific mission of critical conceptual reflection and analysis. On the contrary, in a discussion relevant to just this point, Machado and Silva (2007) observed the following:

> Within the complex set of activities that comprise the scientific method, three clusters of activities can be recognized: experimentation, mathematization, and conceptual analysis. In psychology, the first two of these clusters are well-known and valued, but the third seems less known and valued. (Machado & Silva, 2007, p. 671)

The incorrigibility thus far of mainstream scientific psychology in the face of periodic critiques of its practices in the interpretation of aggregate-level statistical analyses bears witness to the validity of the view that Machado and Silva (2007) expressed in this passage. I offer here one brief anecdote further illustrative of this point.

In March of 1981, there appeared in the *American Psychologist* my first published article addressing the interpretive problems that Bakan (1955, 1966) and Kerlinger (1979) before me had discussed. Within days of the article's publication, a senior and highly influential colleague in my department followed up his perfunctory congratulations to me with the parting remark 'but (your article) is merely theoretical!'

By 'theoretical,' my colleague meant nonempirical, and in his utterance of the term 'merely' I found nearly palpable his disdain for such conceptual work. Indeed, one year later, I received over that same colleague's signature the letter informing me that my scholarship was 'not of sufficient quality' to warrant promotion and tenure.

Whither Psychology?

Now nearly 40 years on since that encounter, it is at least arguable that the fate that Wundt (2013) forecast for a scientific psychology generally dismissive of conceptual inquiry has largely come to pass. Given that the knowledge that mainstream psychological researchers *actually* (as opposed to allegedly) produce by means of their currently favored statistical exercises is knowledge of *populations* and not knowledge of individuals within those populations, it is fair to say that the field that was once psychology has effectively been transformed into a kind of *psycho-demography*. It is a discipline that is *nominally* psychological in that the variables defined for investigation reflect a theoretical interest on the part of investigators in the psychological doings of individuals. Nevertheless, the discipline is *essentially* demographic because its paradigmatic statistical methods are suited only to the production of knowledge about populations.

Of course, there is nothing categorically 'wrong' with demographic inquiry. However, the extant psycho-demography cannot validly be seen as a discipline that has been able to expand psychology's knowledge objectives beyond those that existed originally while at the same time retaining the original objectives. On the contrary, the extant discipline must be seen as one that, *de facto*, has fatally *compromised* its capacity for pursuing psychology's original knowledge objectives, and is in fact pursuing knowledge of a fundamentally different kind. However, by refusing to acknowledge the true nature of its paradigmatic practices in the interpretation of aggregate-level statistics, the best that can be said of the field of inquiry now widely referred to as 'psychology' is that it is one situated in a kind of disciplinary limbo. Sooner or later, mainstream psychologists will either have to rededicate themselves to the pursuit of knowledge about the psychological doings of individuals, and, in the process, abandon their widespread reliance on aggregate statistical methods, or they will have to explicitly eschew their discipline's original knowledge objectives in favor of the continued pursuit of essentially demographic knowledge objectives. Currently, dominant methodological practices cannot accomplish both objectives, and the persistent pretense that matters are otherwise is intellectually damaging to any scientific *psychology* worthy of the name.

Notes

1. Ebbinghaus's (1885) iconic experimental research on memory stands as a paradigmatic example of Bakan's (1966) point.
2. As Bakan (1966) indicated, the concern of the early experimentalists was, more specifically, with the *normal*, *adult* individual, but that restriction does not gainsay the point being made here, which is that the experimentalists' concern was with general laws governing individual-level phenomena.
3. Two expressions used in this passage merit some elaboration. First, I use the expression 'psychological doings' throughout this work to refer inclusively to phenomena that were initially—and continue to be—of theoretical interest in experimental psychology, i.e., phenomena such as sensations, perceptions, judgments, cognitions, memories, emotions, behaviors, etc. Second, I use the expression 'scientifically graspable' here (and elsewhere in this book) in order to avoid premature foreclosure on the question of whether scientific psychology's knowledge objectives are to *explain* psychological 'doings' or, rather, to *understand* them, or even, perhaps, some combination of both. Following the philosophers Wilhelm Windelband (1848–1915) and Wilhelm Dilthey (1833–1911), I see good reason to distinguish between these two knowledge objectives

(cf. Windelband, 1894; Dilthey, 1894), and these reasons will be discussed at greater length in Chapter 7.

4. Had it not been for experimental psychologists' gradual abandonment of the original $N = 1$ method of experimentation in favor of the treatment group method, the eventual merger of scientific psychology's 'two disciplines' along the lines called for by Cronbach (1957) could never have been accomplished. This point will be discussed in greater detail in Chapter 4.
5. It is important to bear in mind here that Wundt also espoused a non-laboratory *Völkerpsychologie*, i.e., a cultural or anthropological psychology which would *complement* the individual psychology with knowledge of customs and mores of entire cultures or peoples (cf. Wundt, 1912; Jüttemann, 2006). Even there, however, inquiry would rely primarily on ethnographic methods, and not on statistical analyses of the sort employed in contemporary mainstream psychological research. Moreover, in his role as a scientist seeking knowledge of cause–effect relationships, Wundt would have fully agreed with the observation made many years later by the philosopher Rom Harré (b. 1927), that "causal processes occur only in individual beings, since mechanisms of action, *even when we act as members of collectives*, must be realized in particular persons" (Harré, 1981, p. 14, emphasis added). We will return to this point in Chapter 4.
6. Note that this is not to deny the logical possibility of establishing a probabilistic fact about an individual within a population. What that would require, however, is a multiplicity of discrete observations *about that same individual*. This is not what is done in population-level studies, where a discrete data point is defined for each of an indefinitely large number of individuals. It is completely inappropriate to regard these two quite different procedures as alternative means to the same knowledge objective.
7. Throughout this book, all translations from original German texts are the author's own unless otherwise indicated.

References

Bakan, D. (1955). The general and the aggregate: A methodological distinction. *Perceptual and Motor Skills, 5,* 211–212.

Bakan, D. (1966). The test of significance in psychological research. *Psychological Bulletin, 66,* 423–437.

Bennett, M., & Hacker, P. M. S. (2003). *Philosophical foundations of neuroscience*. Oxford, UK: Blackwell.

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

Costa, R. E., & Shimp, C. P. (2011). Methods courses and texts in psychology: "Textbook science" and "tourist brochures". *American Psychologist, 31,* 25–43. https://doi.org/10.1037/a0021575.

Cowles, M. (1989). *Statistics in psychology: An historical perspective*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Cronbach, L. J. (1957). The two disciplines of scientific psychology. *American Psychologist, 12,* 671–684.

Danziger, K. (1990). *Constructing the subject: Historical origins of psychological research*. New York: Cambridge University Press.

Dilthey, W. (1894). Ideen über eine beschreibende und zergliedernde Psychologie [Ideas concerning a descriptive and an analytical psychology]. *Sitzungsberichte der Akademie der Wissenschaften zu Berlin*, 1309–1407. Zweiter Halbband.

Ebbinghaus, H. (1885). *Über das Gedächtnis*. Leipzig: Duncker & Humbolt.

Harré, R. (1981). The positivist-empiricist approach and its alternative. In P. Reason & J. Rowan (Eds.), *Human inquiry: A sourcebook of new paradigm research* (pp. 3–17). New York: Wiley.

Jüttemann, G. (Ed.). (2006). *Wilhelm Wundts anderes Erbe: Ein Missverständnis lost sich auf*. Göttingen: Vandenhoek & Ruprecht.

Kerlinger, F. N. (1979). *Behavioral research: A conceptual approach*. New York: Holt, Rinehart, & Winston.

Lamiell, J. T. (2003). *Beyond individual and group differences: Human individuality, scientific psychology, and William Stern's critical personalism*. Thousand Oaks, CA: Sage.

Lamiell, J. T. (2013). On psychology's struggle for existence: Some reflections on Wundt's 1913 essay a century on. *Journal of Theoretical and Philosophical Psychology, 33,* 205–215. https://doi.org/10.1037/a0033460.

Lamiell, J. T. (2015). Statistical thinking in psychological research: In quest of clarity through historical inquiry and conceptual analysis. In J. Martin, J. Sugarman, & K. L. Slaney (Eds.), *The Wiley handbook of theoretical and philosophical psychology: Methods, approaches, and new directions for social sciences* (pp. 200–215). Hoboken, NJ: Wiley.

Lamiell, J. T. (2016). On the concept of 'effects' in contemporary psychological experimentation: A case study in the need for conceptual clarity and discursive precision. In R. Harré & F. Moghaddam (Eds.), *Questioning causality: Scientific explorations of cause and consequence across social contexts* (pp. 83–102). Santa Barbara, CA: Praeger.

Machado, A., & Silva, F. J. (2007). Toward a richer view of the scientific method: The role of conceptual analysis. *American Psychologist, 62,* 671–681. https://doi.org/10.1037/0003-066X.62.7.671.

Monitor on Psychology, A Publication of the American Psychological Association. November, 2016.

Münsterberg, H. (1913). *Psychology and industrial efficiency*. Boston and New York: Houghton-Mifflin.

Porter, T. R. (1986). *The rise of statistical thinking: 1820–1900*. Princeton, NJ: Princeton University Press.

Skinner, B. F. (1990). Can psychology be a science of mind? *American Psychologist, 45,* 1206–1210. https://doi.org/10.1037/0003-066X.45.11.1206.

Stern, W. (1900). *Über Psychologie der individuellen Differenzen (Ideen zu einer "differentiellen Psychologie")* [On the psychology of individual differences (Toward a "differential psychology")]. Leipzig: Barth.

Stern, W. (1911). *Die Differentielle Psychologie in ihrer methodischen Grundlagen* [Methodological foundations of differential psychology]. Leipzig: Barth.

Stern, W. (1927). Selbstdarstellung [Self-portrayal]. In R. Schmidt (Ed.), *Philosophie der Gegenwart in Selbstdarstellungen* (Vol. 6, pp. 128–184). Leipzig: Barth.

Venn, J. (1888). *The logic of chance*. London and New York: Macmillan.

Watson, J. B. (1928). *The ways of behaviorism*. New York: Harper and Brothers.

Windelband, W. (1894/1998). History and natural science (J. T. Lamiell, Trans.). *Theory and Psychology, 8*, 6–22.

Wundt, W. (1912). *Elemente der Völkerpsychologie*. Leipzig: Alfred Kröner Verlag.

Wundt, W. (2013). Psychology's struggle for existence (J. T. Lamiell, Trans.). *History of Psychology, 16,* 195–209. https://doi.org/10.1037/0032319.

2

Challenging the Canon: The Critique and Its Aftermath in Autobiographical Perspective

Near the conclusion of the previous chapter, it was noted that my initial foray into the issues of central concern in this book happened nearly 40 years ago, in the form of an article published in the *American Psychologist (AP)* (Lamiell, 1981). That article was followed, in steady succession over the ensuing ten years, by several additional publications in which I further elaborated key facets of my critical perspective on certain long-standing tenets of mainstream thinking within the psychology of personality. In the present chapter I adopt a quasi-autobiographical perspective, sharing with readers, first, the inspiration for and substantive core of my critique, and, second, my view of the main features of the reception of that critique among advocates of the established traditions.

Eureka!

Facing Students' Vexing Questions

In August of 1976, clutching a freshly minted Ph.D. in personality psychology from Kansas State University, I arrived in Champaign, Illinois to

J. T. Lamiell, *Psychology's Misuse of Statistics and Persistent Dismissal of its Critics*, Palgrave Studies in the Theory and History of Psychology,
https://doi.org/10.1007/978-3-030-12131-0_2

begin what I hoped would be a life-long career as an academic psychologist with a specialty in the psychology of personality. One of my initial undergraduate teaching assignments at the University of Illinois was as a member of an instructional team offering an introductory course in personality. Depending upon the particular semester, the teaching team consisted of three or four instructors. The students enrolled in the course were divided into three or four sections (as the case might have been in any given semester), each with its own weekly meeting times. Each instructor would teach a given section of the class for one-third (or one-fourth) of the semester, and then rotate to another section, where the same material would be presented again. This procedure was repeated for the entire semester, so that, in the end, each instructor would have presented his/her material to every section of the class.

My specific responsibility in that course was to teach the so-called 'psychometric-trait' perspective on the study of personality.[1] I was well-prepared for my assignment, having worked at some length within the framework of that perspective while in graduate school, and so was familiar both with the tenets of trait theory and with the methodological principles of personality test construction, validation, etc. In this latter connection, I had learned as a graduate student the importance—indeed, the necessity—of acquiring facility with statistical methods, foremost among them being the methods of simple and multiple correlation and regression.[2] Unlike many students of psychology, I found that I much enjoyed working with statistical methods, both while learning them in graduate school, and, subsequently, while teaching them to upper-level undergraduates in a course more advanced than the one mentioned above.

In those years, trait psychology was in the midst of difficult times. Throughout the 1970s (and then on through the 1980s), the most widely debated topic within the field, by far, was the extent of consistency over time and, especially, across situations in individuals' respective behavioral manifestations of their putative personality traits. That being so, I saw myself duty-bound as an instructor to allocate ample time in my course of lectures to discussing the substance of that controversy and its theoretical importance for our scientific understanding of personality, moving forward.

With that in mind, I directed my students' attention to the landmark work on the consistency issue that had been published by Walter Mischel (1937–2018) only a few years earlier (Mischel, 1968). In that work, titled *Personality and Assessment*, Mischel (1968) reported that even the best measures of personality traits that had been developed up to that time by trained psychometricians correlated only $r = +.30$ with criterion measures of behaviors that, theoretically, should have correlated more highly with the trait measures.[3] As all established researchers within the field knew, and as cohort after cohort of their graduate student protégés were dutifully being taught every academic year, this meant that, at best, measures of personality traits were accounting for only 9% of the between-person variance in putatively predictable criterion behaviors—the 9% figure being reckoned as the square of the correlation: when r equals .30, then $r^2 = 09$ or 9%.

While these terms of discussion were—and still are—thoroughly familiar to those in or already well along toward careers in psychological research, they were utterly inscrutable to the students in my introductory personality classes. As a result, those students regularly challenged me with questions that would eventually force a profound shift in my own thinking. Time after time when I lectured on the statistical evidence regarding consistency (or lack thereof) in the manifestation by individuals of their respective personality characteristics, my students would implore me to make clear to them just how a correlation coefficient or its square, 'percent variance accounted for,' spoke to the question at hand. They would ask, for example, if the finding $r^2 = 09$ meant that 9% of the individuals investigated in some study were consistent and 91% of those investigated were not. Of course, my reply had to be that, no, the r^2 statistic could not be interpreted in exactly that way. 'Well then,' they would ask, does that statistic indicate that individuals are consistent 9% of the time and inconsistent the remaining 91% of the time? 'Well, no,' I was obliged to respond, 'the r^2 statistic does not mean exactly that, either.'

Seeing my students on the verge of complete exasperation, and not knowing what I could possibly say that would answer their questions in a way that was both intuitively compelling and technically accurate, I finally resorted to assuring them that after they had completed additional coursework in statistical methods, the entire matter would be altogether clear to them. Although that stratagem reliably placated my students enough for me to continue with my day's presentation and then exit the lecture hall with my professorial dignity intact, I was nevertheless always privately tormented by my inability to provide the students with a clear, straightforward, and technically correct answer to their perfectly reasonable questions. Why, I continued to ask myself, was my graceful exit from the lecture hall always contingent upon persuading my students that they just were not yet sophisticated enough in their thinking to formulate their questions in the scientifically correct way?

Then one day, tortured over the matter yet again in the privacy of my office, it hit me: in effect, my students were asking me how to interpret *with respect to individuals* statistics, specifically r and r^2, that were *not defined* for and hence *could not be* interpreted for individuals. With this, I realized that no scientifically valid explication of the meaning of r or r^2 could be given that would answer the questions my students were asking me. The statistics in question were defined for *variables* marking *individual differences*, and at the level of the individual there is *no such thing* as an empirically specifiable individual *difference*. Finally grasping these simple and obvious but nevertheless conceptually elusive truths, I was able to understand why it was impossible for me to state what the aggregate statistics r and r^2 meant with respect to the question of individual-level (in)consistency, other than to say that *they didn't mean anything!* It was *that* fundamental conceptual reality—and certainly not my students' lack of statistical savvy—that was forcing me to patronize them as I had been, semester after semester, in the face of their penetrating questions!

Realizing all of this was a genuine *Eureka* moment for me, and I immediately set about trying to construct for myself on my office blackboard a graphic illustration of the problem that, once back in the lecture hall, my students would find illuminating. After several iterations, that illustration eventually took the form of Fig. 2.1.

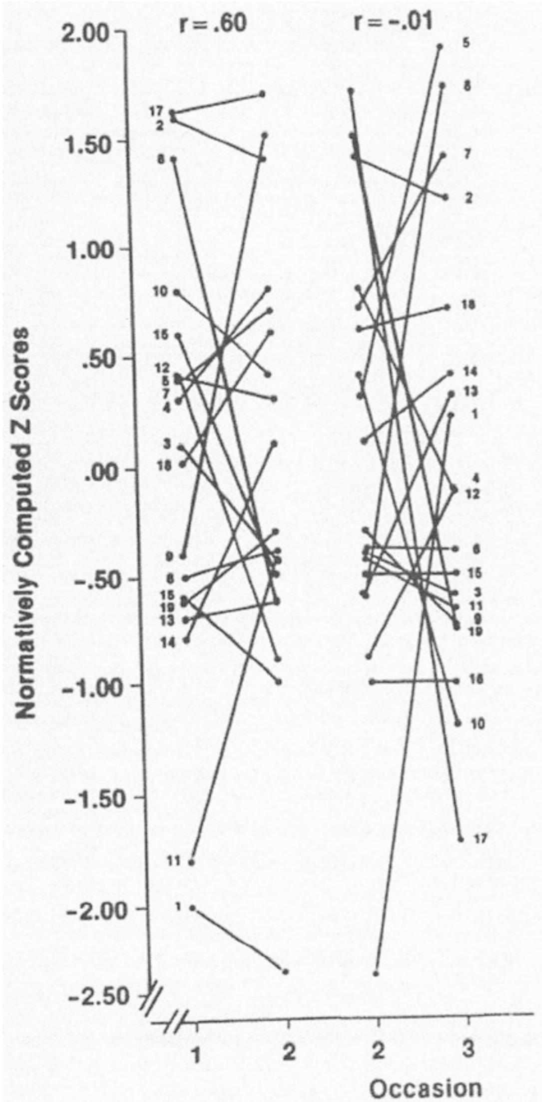

Fig. 2.1 A microscopic picture of data underlying two correlation coefficients

An Empirical Illustration of a Conceptual Problem

In order to vivify the illustration that I envisioned, I decided it would be most effective to work with real data. Accordingly, I cobbled together a brief inventory in the format of a typical personality assessment questionnaire, and administered it on three separate occasions spanning five days (Monday, Wednesday, and Friday) to each of 19 volunteers from a cohort of the students enrolled in my class. In the left panel of Fig. 2.1, each of the lines shown represents one of the 19 student subjects, and connects that subject's standard scores (z-scores) for assessment occasions 1 and 2.[4] Similarly, the lines in the right panel of the figure connect each student's standard scores for assessment occasions 2 and 3.[5] For the data in the left panel of the figure, the Pearson product-moment correlation coefficient proved to be $r = +.60$, whereas for the data in the right panel of the figure, the correlation turned out to be $r = -.01$.

With an image similar to Fig. 2.1 projected onto the lecture hall screen for all of the students to see, I drew their attention to the fact that it is the slope or pitch of an individual line within a panel, and not the single value of the correlation coefficient shown at the top of that panel, that reflects the degree of (in)consistency manifested by the particular individual represented by that line. A perfectly flat or horizontal line indicated perfect consistency, while a steep or sharply pitched line (whether up or down) indicated marked inconsistency. Examining the data in the left panel, for example, where the correlation was very high by the standards prevailing in personality trait studies, it can be seen that subjects 2, 5, 6, 13, and 17 were, indeed, highly consistent from the first to the second assessment occasion. However, it is also apparent that subjects 8, 9, 11, 14, and 15 were quite *in*consistent across those two assessment occasions, the relatively high overall correlation of $r = +.60$ notwithstanding. Conversely for the data in the right panel, where the correlation was essentially zero, and while subjects 1, 5, 8, and 17 were, indeed, highly inconsistent across those two assessment occasions, subjects 2, 6, 15, 16, and 18 were, nevertheless, highly *consistent* across those two assessment occasions.

With the aid of such graphics, I was able to impress upon the students the pedagogical point of the exercise: in *each* of the two panels of the figure, it is shown that some of the 19 research subjects were relatively consistent while others were relatively inconsistent, underscoring the fact that this was true *both* when correlation was relatively high *and* when it was as low (in absolute terms) as it could be. I explained further that the *only* time that the aggregate statistic, r (or its extension, r^2), could be taken as an empirical indicator of individual-level consistency would be if it were to assume the value $r = +1.00$. Under that circumstance, *every* individual subject would have to have been *perfectly* consistent, and a plot of such data in accordance with the conventions adopted for Fig. 2.1 would take the form of the hypothetical data displayed in Fig. 2.2.

Of course, and as my undergraduate students could easily understand, the correlations obtained in actual trait studies never were perfect, meaning that the entire debate within personality psychology over temporal and trans-situational (in)consistency had for its duration up to then been waged with reference to some variant of the empirical circumstances depicted in Fig. 2.1. Seeing this, my students could, in turn, easily see that the arguments being put forward by disputants on *both* sides of that debate were entirely off base: clearly a 'high' (but not perfect) 'consistency coefficient' is *not* grounds for claiming that, in general, the research subjects had manifested the assessed personality characteristics with a 'high' degree of consistency, *nor* is a 'low'—even zero—'consistency coefficient' grounds for the claim that, in general, the research subjects had manifested no consistency in this regard. With the data plots shown in Fig. 2.1 displayed in front of them, my students could readily grasp that (a) the only thing that a personality investigators could validly claim to know on the basis of his/her imperfect aggregate 'consistency coefficients'—*whether* 'high' *or* 'low'—was that his/her research subjects had not been *equally* (in)consistent, and that, for that very reason, (b) that investigator was in no position to claim *any knowledge whatsoever* about the (in)consistency of *any* one of his/her research subjects without disregarding the aggregate r (or r^2) value and examining the data for that particular subject.

Of course, all of the foregoing still left my students to wonder—right along with their instructor—how mainstream trait psychologists could

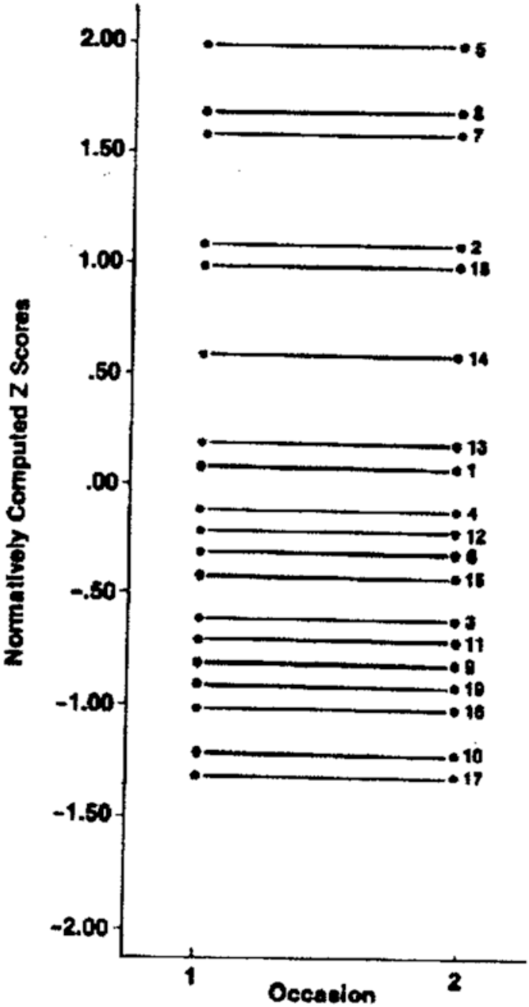

Fig. 2.2 A microscopic picture of correlational data for a hypothetical case in which $r = +1.00$

have adopted so consensually such an egregiously invalid approach to an issue that they held to be of such theoretical consequence. Answering that question would require historical research to which I would eventually turn, but only years later.

From the Lecture Hall to the Professional Literature

Meanwhile, given the extent to which the (in)consistency debate was, in fact, dominating the field of personality psychology during the years being recalled here, it seemed clear to me that sharing with my colleagues in the field the insight to which my 'naïve' *(sic!)* undergraduate students had led me would dramatically alter discourse within the field. Accordingly, I abruptly abandoned my other research efforts and turned all of my scholarly attention to the preparation of an article in which I would publish my newly won insight. The process proved long and difficult: the first three versions of the article were rejected by the *AP* (a blessing, as I would come to appreciate in retrospect), but always with an invitation to revise and resubmit. Seeing the time demands that this project was imposing upon me, a senior colleague concerned for my long-range professional well-being urged me to put aside the project for a while and concentrate on other more modest and readily achievable publication and grant-winning goals, so as to have some prospect for gaining promotion and tenure when my six-year probationary period as an assistant professor elapsed (an eventuality which by that time was not far off). I chose not to follow that advice, and pressed on.

Finally, after some two years had passed, the fourth submitted version of my article was accepted for publication, and in March of 1981, the work appeared in the *AP* bearing the title 'Toward an Idiothetic Psychology of Personality' (Lamiell, 1981). With the neologism in the title, I hoped to convey my intent to coordinate within a single framework two approaches to the study of personality that were widely regarded as disparate from or even antithetical to one another: the so-called 'idiographic' and 'nomothetic' approaches.

The long-dominant approach to personality research widely termed 'nomothetic' was (and still is), in effect, the 'trait' psychology within the context of which the (in)consistency debate had been swirling for, by then, more than 10 years. It was the Harvard psychologist Gordon W. Allport (1897–1967) who was most responsible for bestowing the label 'nomothetic' on mainstream trait psychology, and it was he who, in turn, most visibly and persistently juxtaposed that approach with one he labeled 'idiographic' (cf. Allport, 1937). Allport borrowed the terms 'nomothetic' and 'idiographic' from the German philosopher Wilhelm Windelband (1848–1915), who used them not as labels for two contrasting methods for personality research, but rather to name two different kinds of scientific knowledge. One kind, the 'nomothetic,' takes the form of general laws on the model of the *natural* sciences (*die Naturwissenschaften*), i.e., laws in terms of which to *explain*, as Windelband himself put things, *that, which always is* (in German: *das, was immer ist*) in some particular domain of empirical investigation. The other kind of knowledge, the 'idiographic,' consists of contextually meaningful accounts of non-recurrent happenings or entities on the model of the *human* sciences (*die Geisteswissenschaften*), providing, in any given instance, an *understanding* of what has been, or of *what once was* (in German: *das, was einmal war*) in a given domain of investigation (cf. Windelband, 1894/1998).[6,7] With the neologism 'idiothetic,' I sought to identify an approach to personality studies whereby the determination of those traits relevant to the description of any given individual's personality would be done idiographically, i.e., case by individual case, 'nomothetic' knowledge would be found, if at all, in what might prove common to all in the domain of personality development.

The 1981 *AP* article was divided into two major sections. The first section was devoted to explaining both that and why the correlation coefficients (r and r^2 values) issuing from studies of variables marking individual differences in personality traits were not valid empirical grounds for generalizations about the degree of (in)consistency in individuals' manifestations of those traits. As I had explained to my students in the lecture hall (refer above), those coefficients indicated, when less than perfect, only that the individuals investigated had not been *equally* (in)consistent. This *meant* (a) that no statement about the (in) consistency of individuals *in general* could possibly be valid, and that (b) knowledge of what had transpired with any given individual subject could not validly be claimed without examining the data for that individual subject.

In the second major section of the 1981 *AP* article I argued for the desirability and possibility of reconceiving personality research as a discipline standing completely apart from 'differential' psychology, rather than as one of 'differential' psychology's many subsidiary content domains.[8] In order to achieve this disciplinary separation without necessarily abandoning entirely the theoretical concept of personality traits, I argued for the desirability of a method for representing a given individual's locations along various trait dimensions of personality that would not require the comparison of that individual with others.[9] Such an approach could be devised, I suggested, in accordance with what Raymond B. Cattell (1905–1998) had once termed 'interactive' measurement (Cattell, 1944), and in the second half of the *AP* article I offered an illustration of how such an approach might be put into practice.

Reception of the Critique Within the Mainstream

The immediate reaction to the 1981 *AP* article was encouraging. Privately, a great many requests for reprints of the article were received,[10] as well as numerous letters complimentary of the work and urging further pursuit of the main ideas. In the published literature, the authors of a chapter that appeared in the 1983 issue of the *Annual Review of Psychology* (volume 34 in that series) branded my *AP* article 'the single most important paper' addressing theoretical issues in personality that had been published within the time period under review (Rorer & Widiger, 1983, p. 448). An interview spotlighting my critical perspective on the field was published in a mainstream personality psychology textbook (Ross, 1987), and, two years later, my own book-length elaboration of my critique (Lamiell, 1987) was hailed as a volume 'meriting the careful attention of all workers in the field' (McReynolds, 1989, p. 133).

The *AP* article also prompted an invitation to me from abroad to participate in a symposium planned for the Second European Conference on Personality, to be held in Bielefeld (then West) Germany in May of 1984. Following my presentation at that conference, one in which

I reviewed and elaborated upon the points I had developed in the *AP* article (cf. Lamiell, 1986a), the late Jerry S. Wiggins (1931–2006), a prominent North American personality psychologist and author of a highly regarded textbook on the mainstream approach to personality assessment and research (Wiggins, 1973), told me that my remarks were the most sophisticated critique of mainstream thinking that he had ever encountered.[11,12]

These and similar reactions to the *AP* article, and to related works that I went on to publish over the next decade (e.g., Lamiell, 1982, 1986a, 1987, 1990a; Lamiell & Trierweiler, 1986b; Lamiell, Trierweiler, & Foss, 1983), encouraged me to believe that my critique of mainstream thinking would have a swift and far-reaching effect on research practices within the field. This confidence proved woefully naïve. As reactions to my critique less favorable than those cited above began to surface, I encountered obstacles that I had not anticipated. From my perspective, those obstacles seemed to be rooted in the difficulties discussed below.

Insufficient Reflection Within the Mainstream on the Logic of Basic Statistical Concepts and Methods

One of those difficulties appeared to me as a lack of careful reflection by defenders of mainstream practices on certain rudimentary aspects of their own statistical concepts and methods. I relate here three examples.

In the second of two articles by Paunonen and Jackson (1986a, 1986b) challenging the validity of the arguments I had advanced in the *AP* article and in its immediate sequel (Lamiell, 1982), those authors summarily dismissed my claim that, except under empirical circumstances never realized in practice, aggregate statistics could not provide valid warrant for any claims to knowledge about the psychological doings of individuals. On the contrary, they argued, I had failed to appreciate that aggregate statistics, 'can and must be interpreted *probabilistically* at the level of the individual' (Paunonen & Jackson, 1986b, p. 471, emphasis in original). In elaborating their defense of this counter-claim, Paunonen and Jackson (1986b) blurred to the point of obliteration the distinction between frequentist and subjectivist understandings of probability discussed in the previous chapter, showing no appreciation for the distinction between *claims to knowledge* about one's research subjects, on the one hand, and *statements of subjective beliefs* about those subjects, on the other. This distinction was briefly elaborated in Chapter 1, and will be discussed again at greater length in Chapter 5.[13]

Insufficient reflection on the logic of basic statistical concepts and methods also seemed apparent to me in a publication by Dar and Serlin (1990) commenting critically on a chapter that I contributed to the 1990 volume of the *Annals of Theoretical Psychology* (Lamiell, 1990a). In that chapter, I discussed and illustrated with empirical data the fundamental invalidity of claiming scientific knowledge about the predictability of individual-level occurrences based on the results of aggregate multiple correlation/regression analyses of the sort common in studies of individual differences (Lamiell, 1990a). In their critical commentary, Dar and Serlin (1990) sought to refute my argument on the grounds that the inadequacies of regression analyses that I had claimed to demonstrate were mere artifacts of my peculiar decision to discuss those analyses in terms of variable means, variances, and covariances instead of in the more usual terms of regression line slopes and *y*-axis intercepts. In my rejoinder to that commentary (Lamiell, 1990b), I was compelled to remind Dar and Serlin that regression line slopes and *y*-axis intercepts are *themselves* defined in terms of—and hence *just are* an alternative and wholly equivalent way of talking about—variable means, variances, and covariances. Privately, I saw in that interchange another attempt to derail the argument I was advancing against traditional mainstream statistical practices by means of a counter-argument that was itself manifestly deficient in terms of the conceptual rudiments of those practices.

The last example of this problem to be discussed here is one I draw from a professional interchange that took place at a conference.

In the late 1980s, I attended an invited talk given by a senior and very prominent personality researcher of that era. In the first part of his talk, he took the occasion to chide another equally prominent researcher of the era (not present at the conference) who in published work had drawn inferences about the doings of his individual research subjects on the basis of experimental treatment group means. The speaker was setting up his discussion of his own research findings, which would focus on correlations capturing systematic between-person variance around his treatment group means, variance of just the sort that the target of his criticism had ignored in his own data analyses.

At a social gathering that took place after the presentation, the opportunity arose for me to express to the speaker my appreciation for his point about the inappropriateness of drawing inferences about individuals on the basis of group means. He seemed pleased. But then I asked him if he was at all troubled by the fact that the correlation coefficients by which he was placing such great store as a basis for drawing inferences about his own individual research subjects were *themselves* group means. From the look on my interlocutor's face, it seemed clear that this was a logical fact on the implications of which he had simply not reflected. He responded curtly, 'Well, there are group means and then there are group means,' and that was the end of the conversation.

I believe that this manner of thinking was then and remains now widely prevalent among mainstream personality investigators. Dating back at least to the call issued by Lee J. Cronbach (1916–2001) in 1957 for a merger of scientific psychology's 'two disciplines' (Cronbach, 1957), mainstream psychological researchers have prized the coordinated exercise of experimental and correlational methods, with the latter used as a means of capturing between-person variance around the treatment group means; variance left unaccounted for by the former. Easily overlooked when the results of such research are interpreted is the fact that the correlation coefficients used to capture the otherwise unexplained variance around treatment group means *are themselves group means.* As such, those correlations are subject to the same interpretive constraints that restrict the inferences that may be made on the basis of the treatment group means themselves. My interlocutor in the above interchange had obviously overlooked this basic logical fact, and, as a result, proceeded to commit the very conceptual error for which he had chided his absent colleague.

Insufficiently Careful Readings of the Critique's Assertions

A third obstacle that I found surfacing in reactions to my critique of mainstream thinking was a lack of sustained, careful attention to the actual assertions built into that critique. For example, I often encountered (mostly but not always in unpublished correspondence) the argument that the critical voice I was raising in the 1980s was little or no different in substance from that of Gordon W. Allport (1897–1967)—whose voice, I was advised, had long since been heard at length and then decisively rejected (see, e.g., Eysenck, 1954; Holt, 1962; McClelland, 1951; Sanford, 1963).[14]

There was, to be sure, some overlap between Allport's work and my own early writings. It was, after all, Allport who was most responsible for introducing personality psychologists to the 'nomothetic' and 'idiographic' terminology (see esp., Allport, 1937), and it was out of that terminology that I fashioned the neologism 'idiothetic.' Allport's argument, however, came to be that idiographic knowledge of individuals was needed to *complement* the nomothetic knowledge of individuals generated by traditional studies of individual differences in personality traits (see esp., Allport, 1961, 1966). Quite obviously, that argument entailed Allport's acceptance of traditional trait psychology as a means of generating nomothetic knowledge of individuals.[15]

In contrast to Allport's position on this crucially important point, my own view was, from the outset, that traditional trait psychology never could and never would be able to generate genuinely nomothetic knowledge of individuals. For my part, therefore, there could be no talk of complementing traditional trait psychology with idiographic knowledge, as Allport envisioned.

Following an initial attempt in 1986 to head off confusion on this point (Lamiell, 1986b), I spoke to the matter more forcefully in my 1987 book:

> What Allport failed to see clearly is that the knowledge generated by research conducted within [the traditional trait] paradigm would never be nomothetic … Because such research does exactly what its name suggests,

> its yield is singularly uninformative about any one of the individuals the differences between whom have been studied. This being the case, we are entitled to wonder just how such research manages to carry the name 'nomothetic.' For how can findings that are uninterpretable at the level of the individual possibly be regarded as advancing the quest for general principles of individual psychological functioning? And if the individual differences paradigm cannot advance this quest, then in what sense is it suited at all—let alone 'uniquely'—to personality psychology's nomothetic objectives? (Lamiell, 1987, pp. 15–16, emphasis in original)

Consistent with the distinction discussed in Chapter 1 between knowledge of what is true *in general* and knowledge of what is true *on average*, I pointed out further in my 1987 book:

> (A) general principle of personality would properly be thought of as one that has been found to hold, within the limits of induction that constrain all scientific inquiry, *for each of many individuals.* It follows that research devoted to the search for such principles must be conducted—and its empirical findings must be interpretable—at the level of the individual. (Lamiell, 1987, p. 15, emphasis in original)

It was with these passages in mind that I found worrisome the remark in a 1989 review of my 1987 book, stating that my criticisms of mainstream trait psychology were 'not new,' and were 'essentially the same as—but more extreme than—the view championed 50 years ago by Allport … (McReynolds, 1989, p. 133).' My rejection of the thesis that conventional trait psychology is 'nomothetic' was not simply a 'more extreme' version of Allport's position but, instead, a fundamental *departure* from that position. The recurrence of this criticism of my views, however, would eventually lead me to redouble my efforts to achieve clarity in the matter (Lamiell, 1997, 2000). Those later efforts would be supplemented by the publication of my English translation of Windelband's original 1894 work, a publication warranted by the evidence that can be found throughout the long-running nomothetic vs. idiographic debate in personality psychology that few of the authors who have taken some or another position in the debate over the years have evinced any first-hand familiarity with what Windelband—as opposed to Allport or any other twentieth-century disputant—had actually had to say in connection with those concepts (cf. Windelband, 1894/1998; see also Lamiell, 1998).

In my view, a second example of the need for more careful readings of my critique of mainstream thinking materialized around my discussion of the issue of temporal and trans-situational (in)consistency in individuals' manifestations of their putative personality traits.

For reasons explained at the outset of this chapter, I framed my first published critique of mainstream researchers' use of aggregate statistics within the (in)consistency debate because I was writing at that time as a personality psychologist and that happened to be the substantive issue of greatest concern in the literature of personality psychology at the time. It was thus reasonable (but by no means necessary) to use that issue as a medium for illustrating my points.

Unfortunately, some critics came to see the (in)consistency issue as my primary substantive concern, and, at that, proceeded to characterize my views on that issue in ways I could scarcely recognize. It seemed to me that a more attentive reading of my argument could have prevented this.

I found one example relevant to this point in the previously cited critical commentary by Dar and Serlin (1990). Those authors commented on my discussion of the (in)consistency issue as follows:

> What about Lamiell's (unexplained) requirement of (exactly) equal temporal consistency among persons in regard to the measured attribute? While traditional personality theory does require some evidence of temporal stability, … we fail to understand why the degree of stability (or instability) should be equal for all individuals. … [T]his is clearly an absurd requirement … The point is that perfect measured stability is not a necessary condition for attributes to be considered universal. (Dar and Serlin, 1990, p. 194, parentheses in original)

What so puzzled me about this passage is that nowhere in my discussions of the (in)consistency issue was there any claim that some specific level of consistency, perfect or otherwise, would have to be manifested equally by all individuals in order for a given trait concept to have scientific validity. I was then and remain now in full agreement with Dar and Serlin's (1990) claim that that would have been an absurd theoretical requirement, and I had expressed myself to that effect in print five years earlier (see below).

The reference in my discussion to a hypothetical perfect correlation, one that would reflect equal and perfect consistency by all investigated subjects, was intended simply to illustrate by way of contrast with actual imperfect correlations (a) what *would* have to be the case in order for any such correlation to be interpretable at the level of the individual, and (b) how and why imperfect correlations—be they low *or* high but not perfect—failed to qualify in this regard. As they pertained to the argument I was developing, less-than-perfect 'consistency coefficients' were problematic not because they were less than perfect, but because, being less than perfect, they were uninterpretable for individuals, and knowledge of individual-level doings was what was required from the standpoint of personality theory.

Although, as noted earlier, Dar and Serlin's (1990) specific charge was to comment on my 1990 contribution to the *Annals of Theoretical Psychology* (Lamiell, 1990a), my 1987 book was cited in that contribution, and had Dar and Serlin consulted relevant pages of that book, they would have found the following passage, in which I sought to ensure clarity on the very point that concerned them:

> Note that the argument here is not that individuals must be shown to manifest certain attributes with perfect consistency over time and across situations in order to validate the theoretical assumption of consistency in personality. Indeed, I know of no spokesperson for the consistency thesis who has ever asserted such an extreme view. The argument is that in order to defend the claim that one is studying empirically the degree of consistency with which particular attributes are manifested by individuals, one must at the very least be dealing with empirical findings that are both (a) relevant to the question of (in)consistency and (b) interpretable at the level of the individual. It just so happens that the 'stability coefficients' generated by individual differences research fail to qualify in the latter regard unless they are perfect, and they cannot be perfect unless each of the individuals under investigation is perfectly consistent in the sense captured by normative measurement. (Lamiell, 1987, p. 106)

Though Dar and Serlin's (1990) misrepresentation of my views on this matter might well have been an unfortunate mistake on their part, I was less sure that the same could be said in this next instance.

In 2007, I was asked to review a manuscript that had been submitted to the journal *New Ideas in Psychology.* In the work to be reviewed, the author was making a concerted attempt to coordinate views on the study of personality that I had advanced with certain aspects of traditional mainstream thinking. As is customary, the action editor for the submission in question shared with me the anonymous remarks of the other reviewer. Those remarks included the following:

> If the author wants to earn credibility among readers in the field of personality psychology proper, then he or she would do well to ignore Lamiell's (1981, 1987) ranting about how traits can never apply to the individual. Without putting too fine a point on it, Lamiell is just completely wrong—or else absolutely nobody in the field understands Lamiell. If a person scores high on a trait measure of extraversion, then the probabilistic likelihood is that he or she will indeed show more qualities of extraversion in daily life, compared to a person who scores lower. It is all probabilistic, of course—after all, we are talking psychology here, not chemistry. But Lamiell wants it all to be like chemistry, ironically. He wants us to know that if a person scores high on extraversion that that person must always and forever more, 100% of the time and through thick and thin, be an extravert. If this is not the case even once (say, Saturday, May 6, 2012 for 15 minutes), then we can never, absolutely never, under penalty of eternal censure, use a trait term to describe a person. The position is absurd, and the author of this manuscript does not need to ascribe to it, as far as I can tell. The fact that idiographic structures of traits do not map perfectly on to the Big Five may suggest that the Big Five is not the be all and end all of trait taxonomies—but this finding in no way supports Lamiell, nor does it eviscerate the concept of a personality trait, which is the central goal, to the extent I can discern one, in Lamiell's quixotic agenda. Perhaps I have made my point too strongly. (From an anonymous review of a manuscript submitted to *New Ideas in Psychology*, March 2013)

Conclusion

The level of scholarship reflected by the passage just quoted flatters neither its author nor the 'readership of the personality literature' for whom that author professed to speak. Nevertheless, it might help to explain the dual fact that (a) no intellectually sustainable refutation of the critique I have made of mainstream statistical practices has to date been put forward, and yet (b) those practices continue to the present, essentially unchanged. Indeed, as the Oxford philosopher and social scientist Rom Harré (2006) lamented, my critique of those practices within mainstream personality psychology 'fell on deaf ears' (Harré, 2006, p. 180). He continued as follows:

> It is astounding to see the very same fallacies rife in the field even in the 21st century.
>
> Lamiell (is) a psychologist whose ways of thinking *should* by now have been adopted by everyone interested in the scientific study of personality. Perhaps the reader who turns back to look at the logical slippages in the writings of (mainstream thinkers) will be able to get a sense of how extraordinary it is that the 'Lamiell lessons' have not yet been learned. (Harré, 2006, p. 180, brackets added, emphasis in original)

Harré's words here starkly evince personality psychology's incorrigibility up to 2006, by which time it was fully apparent that the impediments to genuine and lasting change were far more deeply rooted than I had imagined in the early 1980s. By the mid-1990s, I had also come to see more clearly than I did prior to then how the interpretive fallacies about which I had been writing prevailed not just within the subdiscipline of personality studies, but across the other subdisciplines of mainstream psychology as well. Clearly, overcoming those fallacies was going to demand long effort, and it seemed to me that that effort would properly begin with research into the historical developments that had led mainstream psychological researchers to adopt population-level statistical methods in the first place. It is to this aspect of matters that we turn our attention in the next chapter.

Notes

1. Other perspectives taught included, variously, social-cognitive, biological, and cultural.
2. The quality and thoroughness of the training that I received in those methods while in graduate school would prove immensely valuable to me, albeit in ways I had not initially anticipated. Among other things, that training familiarized me with the fundamental equivalence of correlation/regression and analysis of variance (ANOVA).
3. In the professional literature, the lower-case '*r*' is the generally agreed-upon symbol for the Pearson product-moment correlation coefficient, which is, by far, the most widely used index of statistical correlation in psychology. The symbol was first adopted by Francis Galton (1822–1911) as a short-hand referent for the conceptual handmaiden of correlation, i.e., regression.
4. The data for this exercise were displayed in terms of standard scores because the point of the exercise was to demonstrate the impossibility of drawing valid conclusions about the temporal or trans-situational (in)consistency of individuals on the basis of aggregate statistical indices, and, as my students already knew, each such correlation is defined as the average of the cross-products of the standard scores in each of the *N* pairs (where in this case, *N* equaled 19).
5. Busy one day at my desk with these data plots, a colleague stopped by my office and, looking over my shoulder, asked me what I was busy with. 'Those are my data!' I exclaimed. 'Oh,' he exhaled in a tone of condolence, 'they look like the dog's breakfast.' 'No need for pity,' I assured him. 'In this case, that is precisely the point!'
6. In this respect, Windelband's thinking was much in accord with that of his contemporary and countryman, Wilhelm Dilthey (1833–1911; cf. Dilthey, 1894).
7. In actuality, conventional trait psychology does *not* qualify as a 'nomothetic' approach to the study of personality in the sense of 'nomothetic' intended by Windelband, and Allport's mislabeling of trait psychology as he did had major untoward consequences. For one thing, it severely compromised his own objective of persuading mainstream personality psychologists to think and work more idiographically as a *complement* to their preferred variable-centered method of inquiry. For another, it gave advocates for traditional trait psychology license to think that their

work truly was nomothetic, and so conformed to the model of natural science that they so prized anyway. These points have been discussed at length by Lamiell (1998).

8. In this and other important respects my thinking was already running along a path that William Stern (mentioned in Chapter 1) had blazed many years earlier, although this would not become known to me until more than a decade later.
9. At the time of my writing of the *AP* article, I used the term 'measurement' and its cognates, having not yet benefited from the excellent scholarship on this subject that would later be produced by Joel Michell (see, e.g., Michell, 2003, 2011). In time duly chastened by Michell's work, I now regard the expression 'numerical representation' as much more apt in contexts where, consistent with then- and still-prevailing practices within the mainstream, I once referred to 'measurement.'
10. Younger readers of this work should bear in mind that 1981 was well before the era of digitized publications. At that earlier time, actual offprints of articles were produced by publishers, purchased by authors (or their institutional sponsors), and distributed by conventional mail in response to requests from readers.
11. Soon thereafter, Wiggins invited me to contribute an article to an issue of the journal *Clinical Psychology Review*, an invitation that I was most pleased to accept (Lamiell & Trierweiler, 1986a).
12. It was also at the Bielefeld conference where, to my unending gratitude, several European colleagues urged me to look into the works of William Stern. Based on my presentation at the conference, those colleagues believed that I would discover in Stern's writings many points of compatibility with my own thinking. This was a watershed in my intellectual development. After acquiring some facility with the German language, I did, in 1990, begin reading some of Stern's works, and my engagement with his ideas has continued to this day (cf. Lamiell, 2003, 2010).
13. More than 20 years after the referenced publications by Paunonen and Jackson, another vivid example of this deep and abiding conflation was provided by Hofstee (2007) in his critical review of Lamiell (2003).
14. Commenting on a presentation along the lines of the material being discussed here that I was invited to make at a meeting of the Society for Personology in the late 1980s, one of the individuals just cited, who

was present at the meeting, remarked, 'Lamiell, half of what you have said here today is wrong, and the other half is just stupid!' I never was able to find out which half was which, or why.

15. In fact, it was Allport's concession to mainstream thinking on this very point that eventually forced him to 'cry uncle and retire to (his) corner' (Allport, 1966, p. 107; cf. Lamiell, 1997).

References

Allport, G. W. (1937). *Personality: A psychological interpretation*. New York: Holt, Rinehart & Winston.

Allport, G. W. (1961). *Pattern and growth in personality*. New York: Holt, Rinehart, and Winston.

Allport, G. W. (1966). Traits revisited. *American Psychologist, 21,* 1–10.

Cattell, R. B. (1944). Psychological measurement: Normative, ipsative, interactive. *Psychological Review, 51,* 292–303.

Cronbach, L. J. (1957). The two disciplines of scientific psychology. *American Psychologist, 12,* 671–684.

Dar, R., & Serlin, R. C. (1990). For whom the bell curve toils: Universality in individual differences research. In D. N. Robinson & L. P. Mos (Eds.), *Annals of theoretical psychology* (Vol. 6, pp. 193–199). New York: Plenum.

Dilthey, W. (1894). *Ideen über eine beschreibende und zergliedernde Psychologie* [Toward a descriptive and analytical psychology] (Sitzungsberichte der Akademie der Wissenschaften zu Berlin, zweiter Halbband), pp. 1309–1407.

Eysenck, H. J. (1954). The science of personality: Nomothetic! *Psychological Review, 61,* 339–342.

Harré, R. (2006). *Key thinkers in psychology*. Thousand Oaks, CA: Sage.

Hofstee, W. K. B. (2007). *Unbehagen* in individual differences—A review. *Journal of Individual Differences, 28,* 252–253. https://doi.org/10.1027/1614-0001.28.4.252.

Holt, R. (1962). Individuality and generalization in the psychology of personality. *Journal of Personality, 30,* 377–404.

Lamiell, J. T. (1981). Toward an idiothetic psychology of personality. *American Psychologist, 36,* 276–289.

Lamiell, J. T. (1982). The case for an idiothetic psychology of personality: A conceptual and empirical foundation. In B. A. Maher & W. B. Maher (Eds.), *Progress in experimental personality research* (Vol. 11, pp. 1–84). New York: Academic Press.

Lamiell, J. T. (1986a). Epistemological tenets of an idiothetic psychology of personality. In A. Angleitner, A. Furnham, & G. van Heck (Eds.), *Personality psychology in Europe: Current issues and controversies* (pp. 3–22). Lisse, the Netherlands: Swets & Zeitlinger.

Lamiell, J. T. (1986b). What is nomothetic about "nomothetic" personality research? *The Journal of Theoretical and Philosophical Psychology, 6,* 97–107.

Lamiell, J. T. (1987). *The psychology of personality: An epistemological inquiry*. New York: Columbia University Press.

Lamiell, J. T. (1990a). Explanation in the psychology of personality. In D. N. Robinson and L. P. Mos (Eds.), *Annals of theoretical psychology* (Vol. 6, pp. 153–192). New York: Plenum Press.

Lamiell, J. T. (1990b). Let's be careful out there. In D. N. Robinson & L. P. Mos (Eds.), *Annals of theoretical psychology* (Vol. 6, pp. 219–231). New York: Plenum Press.

Lamiell, J. T. (1997). Individuals and the differences between them. In R. Hogan, J. A. Johnson, & S. Briggs (Eds.), *Handbook of personality psychology* (pp. 117–141). New York: Academic Press.

Lamiell, J. T. (1998). 'Nomothetic' and 'idiographic': Contrasting Windelband's understanding with contemporary usage. *Theory and Psychology, 8,* 23–38.

Lamiell, J. T. (2000). A periodic table of personality elements? The "Big Five" and trait "Psychology" in critical perspective. *The Journal of Theoretical and Philosophical Psychology, 20,* 1–24.

Lamiell, J. T. (2003). *Beyond individual and group differences: Human individuality, scientific psychology, and William Stern's critical personalism*. Thousand Oaks, CA: Sage.

Lamiell, J. T. (2010). *William Stern (1871–1938): A brief introduction to his life and works*. Lengerich, Germany: Pabst Science Publishers.

Lamiell, J. T., & Trierweiler, S. J. (1986a). Personality measurement and intuitive personality judgments from an idiothetic point of view. *Clinical Psychology Review, 6,* 471–491.

Lamiell, J. T., & Trierweiler, S. J. (1986b). Interactive measurement, idiothetic inquiry and the challenge to conventional "nomotheticism". *Journal of Personality, 54,* 460–469.

Lamiell, J. T., Trierweiler, S. J., & Foss, M. A. (1983). Detecting (in)consistencies in personality: Reconciling intuitions and empirical evidence. *Journal of Personality Assessment, 47,* 380–389.

McClelland, D. C. (1951). *Personality*. New York: Sloane.

McReynolds, P. (1989). Review of J. T. Lamiell, 'The Psychology of Personality: An Epistemological Inquiry.' *Personality and Individual Differences, 10,* 113.

Michell, J. (2003). The quantitative imperative: Positivism, naïve realism, and the place of qualitative methods in psychology. *Theory and Psychology, 13,* 5–31.

Michell, J. (2011). Qualitative research meets the ghost of Pythagoras. *Theory and Psychology, 21,* 241–259. https://doi.org/10.1177/0959354310391351.

Mischel, W. (1968). *Personality and assessment*. New York: Wiley.

Paunonen, S. V., & Jackson, D. N. (1986a). Nomothetic and idiothetic measurement in personality. *Journal of Personality, 54,* 447–459.

Paunonen, S. V., & Jackson, D. N. (1986b). Idiothetic inquiry and the toil of Sisyphus. *Journal of Personality, 54,* 470–477.

Rorer, L. G., & Widiger, T. A. (1983). Personality structure and assessment. *Annual review of psychology* (Vol. 34, pp. 431–463). Palo Alto, CA: Annual Reviews Inc.

Ross, A. O. (1987). *Personality: The scientific study of complex human behavior*. New York: Holt, Rinehart, & Winston.

Sanford, N. (1963). Personality: Its place in psychology. In S. Koch (Ed.), *Psychology: A study of a science* (Vol. 5, pp. 488–592). New York: McGraw-Hill.

Wiggins, J. S. (1973). *Personality and prediction: Principles of personality assessment*. Reading, MA: Addison-Wesley.

Windelband, W. (1894/1998). History and natural science (J. T. Lamiell, Trans.). *Theory and Psychology, 8,* 6–22.

3

The Entrenchment of Statistical Thinking in Early Twentieth Century Differential Psychology

Scarcely a general psychology textbook is nowadays to be seen that leaves unmentioned Wilhelm Wundt as the founder of psychology as an experimental science. Yet in the model for psychological experimentation that Wundt formally established at Leipzig, Germany in 1879,[1] population-level statistical concepts and methods had no place. Certainly, efforts at quantification played a major role in the early years of experimental psychology, and statistical analyses of experimentally-generated measurements were often carried out. However, all of the measurements submitted to such analyses in any given instance would have been obtained from the same research subject, and the objective of the analyses was to estimate the magnitude of error contained within the measurements.[2] Psychological experimentation on the Leipzig model had nothing to do with estimating the parameters (e.g., means, variances, and covariances) of variables defined only for populations. In short, scientific psychology was, at its inception, a devoid of the sorts of statistical practices that would, in time, come to dominate thinking within the field's mainstream (cf. Danziger, 1990). The present chapter reviews the major early twentieth century developments that would lead to this radical disciplinary transformation.

J. T. Lamiell, *Psychology's Misuse of Statistics and Persistent Dismissal of its Critics*, Palgrave Studies in the Theory and History of Psychology,
https://doi.org/10.1007/978-3-030-12131-0_3

William Stern and the Establishment of Differential Psychology

A watershed event in this connection was the publication in 1900 of a book authored by William Stern (1871–1938) titled (in translation) *On the Psychology of Individual Differences: Toward a "Differential Psychology"* (Stern, 1900). In a retrospective commentary that appeared in a 1914 publication, Stern described the new sub-discipline he had launched in the 1900 book as one devoted 'not to the search for the general laws of psychological life, but rather to the differentiations that result from such factors as age and sex, race and culture, types of temperaments, traits of character and talent, intelligence, etc.' (Stern, 1914, p. 416). It is important to appreciate that Stern did not conceive of differential psychology as an alternative to the general experimental psychology on the Leipzig model, but rather as a sub-discipline that would operate alongside of and in a fashion complementary to that general experimental psychology. He pointed out that the execution of that complementary function entailed the study of large populations of subjects simultaneously, and that 'to this end, survey procedures and methods of statistical analysis made their way into the discipline' (Stern, 1914, p. 416). Stern's language here clearly reflects the historical reality that it was through the portal of differential psychology that population-level statistical methods first gained a foothold in scientific psychology.

Stern's core *theoretical* concern at the turn of the twentieth century was that the established experimental psychology, devoted as it was to the quest for knowledge of the 'general laws of psychological life' in the common-to-all sense of 'general' discussed in Chapter 1, was deliberately blind to any and all manifestations by experimental subjects of some or another aspect of their respective individualities. Thus did Stern exclaim in the very first line of the preface to the 1900 book: '*Individualität, Problem des zwangstigsten Jahrhunderts!*'—'Individuality, problem of the twentieth century!' By 'problem' (*das Problem*) in this passage Stern clearly meant *challenge*. He believed that a scientific psychology could not long endure if it remained oblivious to the fact that there are aspects of any given individual's psychological life that are *not* 'common to all' individuals and, indeed, might not be characteristic of *any* other individual. If differential psychology could sensitize investigators to this fact, he reasoned, that would effectively broaden the bandwidth of their thinking in ways that would, in time, force a paradigmatic accommodation to the empirical realities of individuality.

Quite apart from the serviceability of differential psychology with respect to Stern's ultimate theoretical concern for human individuality, his 1900 book was seen by many among his contemporaries as a useful vehicle for breaking free of the constraints imposed on them by the Leipzig research model, and for pursuing a scientific knowledge that could be put to practical use in various domains of human affairs outside the experimental laboratories. One of the leading spokespersons for those psychologists was the founder of the second laboratory for experimental psychology in Germany (at the University of Freiburg), Hugo Münsterberg (1863–1916), who expressed matters this way:

> As long as experimental psychology remained essentially a science of the mental laws, common to all human beings, an adjustment to the practical demands of daily life could hardly come in question. With such general laws we could never have mastered the concrete situations of society, because we should have had to leave out of view the fact that there are gifted and ungifted, intelligent and stupid, sensitive and obtuse, quick and slow, energetic and weak individuals. … The study of individual differences itself is not applied psychology, but it is the presupposition without which applied psychology would have remained a phantom. (Münsterberg, 1913, pp. 9–10)

Developments in Stern's Vision of Differential Psychology's Mission

Stern's View in 1900

In his 1900 book, Stern declared that differential psychology's mission was defined by three tasks: (1) to isolate empirically the basic dimensions in terms of which between-person differences were to be understood[3]; (2) to identify the causes of those differences in nature and/or nurture; and (3) to explore the ways in which the differences manifest themselves in various non-laboratory domains of human behavior such as school, work, family life, etc. In effect, this circumscription of differential psychology's mission *equated* it with the study of between-person differences. This point bears emphasis because, unlike most mainstream differential psychologists in the twentieth century (see below), Stern himself would not long maintain this view. The reason for this was, as he would note retrospectively in his 1927 intellectual autobiography, that 'even then (i.e., even in 1900), I could see that true individuality, the understanding of which was my ultimate objective, *cannot be grasped through the channels of differential psychology*' (Stern, 1927, p. 142, parentheses and emphasis added).

Stern's belief—and it was, in fact, expressed in the 1900 book—was that no individual's doings could ever be fully grasped in terms of such lawful statistical relationships and typological categories as the assessment and study of between-person differences might establish (see Stern, 1900, pp. 15–16). He maintained that in the course of such work the intrinsic and irreducible unity—the in-divisibility—of a person's psychological life would, of necessity, be analytically fragmented and hence effectively destroyed. What Stern clearly understood, even if he did not explicitly say so in his 1900 book, is that in statistical studies of between-person differences, every individual must be regarded as an instantiation of the categories of the variables that the investigator has singled out for investigation. On that view, each individual is, in effect, substitutable—both for and by—any other individual who instantiates the same categories. Viewed in this light, individualities are not really captured through statistical studies of variables marking between-person differences but are, on the contrary, actually obscured. Hence, while such studies might well *highlight* the need for a viable framework within which to understand human individualities—this simply by bringing to light the fact that some phenomena of theoretical importance in scientific psychology are *not* 'common to all'—those studies could not themselves *meet* the need for such a framework. Instead, Stern saw from the very beginning that in order for psychology to directly address the 'problem of individuality' itself, something other than the assessment and study of between-person differences would at some point be needed.

Stern's View in 1911

Within a decade of the publication of his 1900 book, Stern was primed to re-articulate his conception of differential psychology's scientific mission. He was disinclined to follow his publisher's request to simply 'revise' the 1900 volume, insisting instead that a substantially new—and decidedly more elaborate—treatment was required. This led to the aforementioned *Methodological Foundations of Differential Psychology*, which was published in 1911 (Stern, 1911). On the title page that work, Stern explicitly indicated that it had been undertaken *in place of* a revision of the 1900 book.[4]

The changes that Stern introduced in the 1911 book did not negate the views he had expressed in the 1900 volume, but they did significantly alter the overall picture of differential psychology that he wanted to advance. As noted above, Stern had presented differential psychology in 1900 as a sub-discipline whose mission was fully circumscribed by studies of between-person differences: their basic dimensions, their causes, and their manifestations. In the 1911 book, this was no longer true. Instead, he defined differential psychology as a sub-discipline incorporating *four research schemes*, two of which did *not* entail the study of between-person differences.

The four research schemes that Stern described in the 1911 book were called variation studies, co-variation studies, psychography, and comparison studies, with each of the four being defined by some particular arrangement of the differential psychologist's 'raw materials': *individuals* and *attributes*. Figure 3.1 displays my English rendition of the schematic that Stern used in his 1911 book to elucidate his vision.

The first (uppermost in Fig. 3.1) of the four research schemes depicts *variation* studies, which entail *the study of a single attribute* in terms of the distribution of measurements of that attribute across many individuals within a population. For example, an investigator might be interested in the basic statistical properties (population mean and variance) of the scores on some

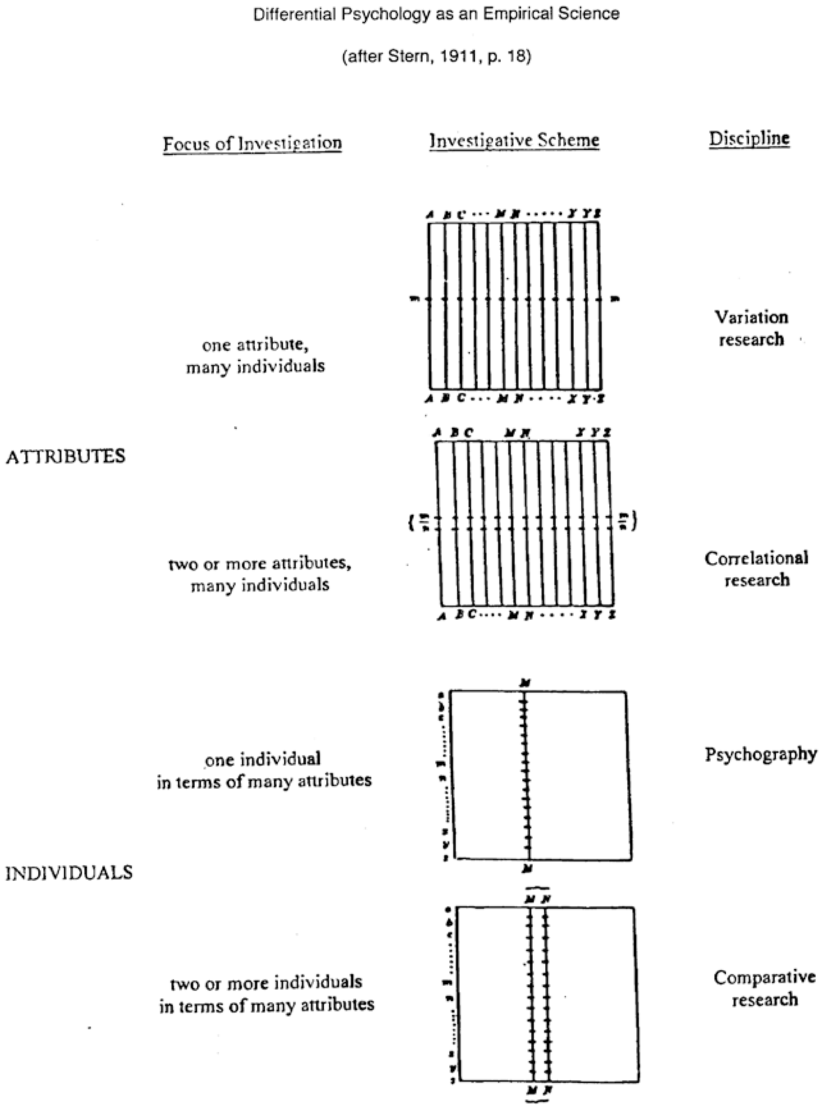

Fig. 3.1 Differential psychology's four research schemes according to Stern (1911)

instrument that s/he has created to assess some particular psychological attribute such as intelligence or some personality characteristic.

The research scheme shown immediately below variation studies in the figure is a portrayal of co-variation (correlation) studies. This scheme is a straightforward extension of variation studies, and entails *the study of two or more attributes* in terms of the distributions of their assessments across many individuals within a population. In addition to the means and variances of each of the attributes, one can in this scheme investigate the correlations between pairs of attributes.

Obviously, both variation and co-variation studies are, in their very essence, studies of between-person differences. Hence, it is in these two research schemes where one finds incorporated Stern's 1900 circumscription of differential psychology's mission. Just as obvious, however, is the fact that those two research schemes no longer capture the whole of differential psychology's mission as Stern had come to view matters in 1911. On the contrary, one finds displayed in the lower half of the figure two additional research schemes, neither of which entails the investigation of between-person differences, and neither of which was explicitly discussed in the 1900 book.

The first of those two additional schemes is one that Stern termed 'psychography' (*die Psychographie*). In a psychographic investigation, the focus is on a *single individual* as characterized in terms of a set of attributes, with the resulting characterization called a 'psychogramm.' The fourth and final research scheme in the figure depicts 'comparison' studies. Such studies are a straightforward extension of psychography, and entail the parallel study of *two or more individuals* through a juxtaposition of their respective psychogramms.

Part II of Stern's four-part 1911 book was devoted to various methodological facets of variation and covariation studies. Particular emphasis was placed on the logic of typological categorization and dimensional measurements, and on the concepts of statistical variation and covariation. In Part III, Stern turned his attention to psychography, doing so with the following observation:

> However varied the problems of differential psychology discussed to this point (in the text) may have been, they all shared something in common:

> *the object of investigation was the attribute in its distribution across individuals*. The individuals (themselves) served only as the means of the research inasmuch as they were carriers of the attributes under investigation. At this point, the direction of research must be altered ninety degrees. The object of investigation is no longer the horizontal distribution of attributes across many individuals, but is instead the vertical structure of an individual with reference to his/her many attributes. This is the problem of psychography, the empirical specification of the psychological features of a person's individuality. (Stern, 1911, p. 318, emphasis and parentheses added)[5]

In the light of this quotation, the reader can appreciate more fully why Stern inserted the terms 'attributes' and 'individuals' into the far-left column of the schematic.[6] They reflect his grasp of the fact that statistical investigations of attribute variables marking between-person differences yield knowledge about those attribute variables. In order to gain knowledge about individuals—including those individuals who have been differentiated from one another in studies of between-person differences—one must actually study those individuals individually, and not variables marking differences between them. Hence, if a sub-discipline called 'differential psychology' were to be made viable as a framework for investigating individualities, that sub-discipline would have to incorporate something akin to the research scheme Stern called 'psychography,' i.e., a scheme that is not itself directed toward the study of between-person differences.[7] This is a distinctive movement beyond the stance that Stern had adopted in the 1900 book.

The full title of the first chapter of Part III of Stern's 1911 book, the part explicitly devoted specifically to 'the investigation of individualities' (*die Erforschung der Individualitäten*; Stern, 1911, p. 317) is 'The Problem of Individuality: Biography' (*Das Individualitätsproblem. Die Biographie*; Stern, 1911, p. 318). This title reflects the importance that Stern attached to biography as a tool in the quest for understanding individualities, and he emphasized that its role should not be understood simply as that of delivering to the psychologist 'raw material' for incorporation into some or another quantitative scheme for representing individualities, but rather as a special method of investigation in its own right.[8] Further emphasizing the value of biography in its own right, Stern noted that psychographic investigations of the sort schematized in Fig. 3.1 could *not* be regarded as a substitute for biography:

> On the contrary, the biographer of the future will be able to make use of a psychographic scheme, or perhaps an already completed psychogramm, as preliminary material for the project. But it is only through an artistic, empathic synthesis of this material that a genuine biography emerges. (Stern, 1911, p. 329)

Here and elsewhere in Stern's writings one finds clear and plentiful evidence of his respect for the importance of qualitative methods in the quest for knowledge about individualities.

Differential Psychology Diverges from Stern

E. L. Thorndike's Perspective on the Study of Individualities

By remarkable coincidence, it was in 1911, the year in which Stern published *Methodological Foundations of Differential Psychology*, that the prominent and influential U.S. psychologist Edward L. Thorndike (1874–1949) published a small monograph bearing the title *Individuality* (Thorndike, 1911). The work appeared as one in a series of 'Riverside Educational Monographs' under the editorship of Henry Suzzallo (1875–1933). That highly-respected scholar, who would soon become President of the University of Washington, wrote effusively in his Editor's Introduction:

> This contribution of Professor Thorndike's … establishes a point of view and indicates a safe method of approach to this intricate study of human nature. With ingenious clarity and brilliant suggestiveness, coupled with scientific caution and accuracy, the author has given us the fundamental modes by which uniformities and variations are to be perceived in human nature … (Suzzallo, Editor's Introduction, in Thorndike, 1911, p. x)

The perspective on individuality that Thorndike advanced in his book differed from that developed by Stern in several important respects. First, where Stern drew a clear distinction between the study of individuals and the study of variables marking individual differences, Thorndike effectively equated the two. This is apparent in the following passage, which appears on page 2 of his monograph:

> We may study a human being in respect to his common humanity, or in respect to his individuality. *In other words*, we may study the features of intellect and character which are common to all men, to man as a species; or we may study the differences in intellect and character which distinguish individual men. (Thorndike, 1911, p. 2, emphasis added)

Another important conceptual difference between Stern and Thorndike was that while Stern emphasized the importance of supplementing quantitative information with qualitative observations in studies of individuals, Thorndike dismissed the qualitative-quantitative distinction as superfluous. He argued that 'a qualitative difference in intellect or character is … really a quantitative difference wherein one term is zero, or a compound of two or more quantitative differences (Thorndike, 1911, p. 5).' This in turn led Thorndike to contend that

> … the difference between any two individuals, if describable at all, is described by comparing the amounts which A possesses of various traits with the amounts which B possesses of the same traits. In intellect and in character, differences of kind between one individual and another turn out to be definable, if defined at all, as compound differences of degree. (Thorndike, 1911, p. 5)[9]

Thirdly, and bearing separate mention here even though it is a logically necessary component of Thorndike's above-mentioned equation of research on individual differences with the study of individuals, Thorndike saw knowledge of the degree of statistical correspondence within a population between the relative magnitudes of two variables marking individual differences as if it also constituted knowledge of, or provided scientific warrant for inferences about, the degree of correspondence between the respective relative magnitudes of those two variables *within individual persons*.

This notion is clearly reflected in Thorndike's claim that the correlation between two variables measuring between-person differences indicates 'the extent to which the amount of one trait possessed *by an individual* is bound up with the amount *he* possesses of some other trait' (Thorndike, 1911, p. 22, emphasis added). Stern's clear distinction between knowledge about individuals, on the one hand, and knowledge about the statistical co-variation of variables that, by their very nature, are defined only for populations, was thus altogether contrary to Thorndike's view on this fundamental epistemological point.[10]

Hugo Münsterberg on the Practical Use of Statistical Knowledge in Psychotechnics

It was noted earlier in this chapter that Stern's call for systematic studies of between-person differences was greeted with much enthusiasm quite apart from the proposed new field's potential as the framework for a basic psychology of human individuality. It was the nascent sub-discipline's capacity for broadly accommodating—indeed, for highlighting—the ubiquitous reality of human differences that made it ideally suited as a framework for applied psychology in a great many domains of life, including school, work, medicine and health care, business and industry, and the military. In the early days of differential psychology, this sort of applied work was widely referred to as 'psychotechnics'—the application of psychology to practical problems in the world outside the laboratory. Among the most important of applied psychology's tools were psychological tests, and one of the most prominent and exuberant proponents of such tests during differential psychology's early years was the aforementioned Hugo Münsterberg.[11]

In 1913, Münsterberg published *Psychology and Industrial Efficiency* (Münsterberg, 1913), which was his own English translation of portions of a work that he had published the previous year in his native German as *Psychologie und Wirtschaftsleben: ein Beitrag zur angewandten Experimental-Psychologie* (Psychology and Economic Life: A Contribution to Applied Experimental Psychology; Münsterberg, 1912). Of most direct relevance to our concerns in this book is evidence in Münsterberg's 1913 publication, a work that became well-known and highly influential (cf. Landy, 1992), that he, in agreement with Thorndike but *contra* Stern, viewed statistical knowledge about variables with respect to which individuals have been differentiated (on the basis of tests of one sort or another) as, at one and the same time, knowledge about the individuals who have been differentiated in terms of those variables.

For example, there is a section in *Psychology and Industrial Efficiency* where Münsterberg discussed the sorts of inferences about individuals that are, or can be, warranted by knowledge of inter-variable correlations discovered through scientifically sound psychotechnical research. He stated that 'with experimental and statistical methods [laboratory psychologists] have gathered ample material which demonstrates the exact degree of probability with which we have a right to expect that certain qualities will occur together' (pp. 134–135). Münsterberg then elaborated on the practical value of such correlational knowledge and proceeded to provide an illustration:

> Inasmuch as one of … two (correlated) traits may be easily detected, while the other may be hidden and can be found out only by long careful tests, it would be valuable, indeed, for the employment manager to become acquainted with such correlations as the psychologist may discover: as soon as he becomes aware of the superficially noticeable symptom, he can foresee that the other disposition is most probably present. To give an illustration: in the interest of such measurements of correlations we have studied in the Harvard laboratory the various characteristics of attention and their mutual dependence. We found that typical connections exist between apparently independent features of attention. Persons who have a rather expansive span of attention for acoustical impressions have also a wide span for the visual objects. Persons whose attention is vivid and quick have on the whole the expansive type of attention, while those who attend slowly have a narrow field of attention, and so on. Hence, *the manifestation of one feature of attention allows us to presuppose without further tests that certain other features may be expected in the particular individual*. (Münsterberg, 1913, pp. 135–136, parentheses and emphasis added)

Here, Münsterberg is effectively embracing the view that knowledge of the correlation between two variables marking differences between individuals, coupled with knowledge of the relative standing of some particular individual on one of those variables, conveys knowledge of 'the exact degree of probability' that that same individual's relative standing on the other variable is within some specified interval of possible test scores. Thus did Münsterberg, like Thorndike, advance a view of the statistical knowledge gleaned from studies of individual differences that is discordant with the distinction Stern drew in his 1911 text between such knowledge and knowledge of individuals themselves.

Consequent Developments

Diverging from Stern's views in the fashion just described, Thorndike and Münsterberg, working independently, blazed a path that would be followed by the overwhelming majority of their contemporaries and immediate successors in early twentieth century differential psychology. This development had two major consequences. One was to render superfluous Stern's (1911) distinction between the study of individual differences and the study of individuals. If aggregate-level statistical knowledge of the sort generated through correlational studies of individual differences can validly be interpreted as Thorndike and Münsterberg claimed, then, in effect, knowledge of variables with respect to which individuals had been differentiated actually *can* be regarded as knowledge, however limited it might be in any isolated instance, of the individuals who have been differentiated in terms of those variables. If that is true, then understanding individualities does not demand some independent research scheme—least of all one needing, as did psychography according to Stern—supplementation by qualitative methods of investigation.

This is not to say that mainstream thinking could not accommodate the construction of individual attribute profiles along the lines of the 'psychogramms' or personal attribute profiles envisioned by Stern as the primary empirical outcomes of psychographic studies. However, such work would be of a strictly derivative nature, dependent upon the prior findings of individual differences studies that would determine, via correlational analyses along the very lines Thorndike (1911) described, the common attribute dimensions that would structure the content of all individual profiles.[12] These are the considerations lying at the core of what would come to be viewed as the properly scientific 'nomothetic' approach in the psychology of personality (cf. Allport, 1937, 1946; Beck, 1953; Eysenck, 1954; Skaggs, 1945), an approach that continues to dominate mainstream thinking within the field to this day (Lamiell, 1987, 1997, 2003).

In the applied realm, the field of psychotechnics quickly became virtually synonymous with the use of psychological tests as the means of generating assessments that could, in turn, be statistically examined to discover the nature and role of between-person differences in many different domains of human performance. Stern himself anticipated—and in Chapter 6 of his 1911 book explicitly warned against—an overreliance on tests and quantitative methods of analysis, to the corresponding neglect of qualitative methods. In time, he would bemoan the continuing failure of most differential psychologists to heed those warnings (Stern, 1921).[13]

A second major consequence of equating the study of individual differences with the study of individuals was to make credible to most an understanding of the knowledge objectives of differential psychology as identical with the knowledge objectives of scientific psychology more generally. In a clear break from Stern's contention that differential psychology's questions were complementary to but still fundamentally different from those central to the general experimental psychology, Anne Anastasi (1908–2001) stated in the preface of her path-breaking 1937 textbook titled *Differential Psychology: Individual and Group Differences in Behavior*, that

> (D)ifferential psychology is … not … a separate field of psychology, but (is) one approach to the understanding of behavior. Its fundamental questions are no different from those of general psychology. It is apparent that if we can explain why individuals react differently from one another, we shall understand why each individual reacts as he does. (Anastasi, 1937, p. vi)[14]

This passage points toward a formal coordination of differential psychology and the general experimental psychology in the pursuit of their putatively common scientific objectives, a possibility toward which Lewis Terman (1877–1956) had likewise pointed some years earlier (Terman, 1924). However, for so long as the general experimental psychology remained dominated by the Leipzig model for experimentation mentioned at the outset of this chapter, such a merger would remain impossible, for there was (and is) no logical way to conform single-subject ($N=1$) experimentation to population-level ($N=$ many) studies of individual differences.

Already by 1920, however, the sea change within psychology whereby single-subject experimentation would be largely abandoned in favor of treatment group experimentation was well underway (Danziger, 1990). Unlike the Leipzig model, the new form of psychological experimentation would incorporate statistical methods of analysis and inference logically equivalent to those employed within differential psychology.[15] So, once treatment group experimentation became sufficiently prevalent, the last formal obstacle to a merger of scientific psychology's experimental and correlational disciplines had been effectively eliminated. This was the development that finally made possible the merger of scientific psychology's two disciplines, a merger for which Lee J. Cronbach (1916–2001) would call for so effectively in his well-known 1957 *American Psychologist* article. It is to that work and its profound consequences that we will direct our attention in the next chapter.

Before doing so, however, it is appropriate and instructive to conclude the present chapter with some observations on the concerns raised by William Stern over a period of more than twenty years as the work of differential psychologists, both in personality psychology and in psychotechnics, became increasingly dominated by psychological tests and statistical methods of data analysis.

Countering an Origin Myth: William Stern as Critic of Differential Psychology

The Stuff of the Origin Myth

With the concept 'origin myth,' the American psychologist Franz Samelson (1923–2015) sought to capture the idea that historians of psychology sometimes present prominent early figures in a particular light so as to legitimize a particular contemporary perspective within their discipline (Samelson, 1974). Such appears to have happened in the case of William Stern (Lamiell, 2006). The aforementioned textbook by Anastasi (1937) can be seen as one early contributor to the creation of differential psychology's origin myth.

Citing Stern's 1900 book as an historical cornerstone of the field, Anastasi (1937) paid special attention to his circumscription of the tasks of differential psychology as he had defined them in that book: identifying the substance, causes, and manifestations of individual differences. Remarkably, she referred to Stern's 1900 book as a 'first edition' (Anastasi, 1937, p. 16), and then, further on, to Stern's 1911 book as a 'revised and enlarged' (p. 17) edition of the 1900 book, despite Stern's own clear insistence that the 1911 book was *not* to be so regarded. Having done that, Anastasi (1937) gave no indication that Stern's 1911 book had significantly re-defined differential psychology's scope so as to include a research scheme, psychography, that does not itself entail the study of individual and group differences, and that could not, therefore, be accommodated by the manner in which he had circumscribed differential psychology's tasks in his 1900 book.

Ten years after the appearance of Anastasi's (1937) book, the author of another widely-read differential psychology textbook, Leona Tyler (1906–1993), did as Anastasi had done in this regard. In Tyler's book, titled *The Psychology of Human Differences*, first published in 1947, she, too, cited Stern's 1900 circumscription of differential psychology's subject matter, and wrote that the three tasks identified by Stern in his 1900 book have remained primary from their day to ours (Tyler, 1947,

p. 13). Tyler did not even mention Stern's 1911 book, let alone the revised view of differential psychology that Stern advanced in that book. It is noteworthy that, years later, Tyler would scold herself for having so long mischaracterized Stern in her lectures and writings—including both subsequent editions of her 1947 text, published in 1956 and 1965, respectively. In an unpublished paper titled *Neglected Insights in Personology*, Tyler wrote:

> Because my teaching specialty has been individual differences, I have known of Stern for a long time, paid proper respect to him in historical introductions to textbooks, but never deepened my understanding. Reading further I realized that I had been giving him credit for just those things he would not have wished to be remembered for. Everybody knows, for example, that Stern invented the IQ. In his later years he indicated in no uncertain terms that he did not regard this as a useful contribution. He is often called the father of differential psychology. But in his autobiography he talks about his realization, after his first book on differential psychology, that 'real individuality, the understanding of which I had made my goal, cannot be reached through the channels of differential psychology. (Tyler, 1985, p. 4)[16]

Alas, it would appear that the vision of Stern projected by the above-cited textbooks is the one that has prevailed—even in Stern's native land. Consider, for example, Fig. 3.2, which displays a likeness of the poster used to publicize the 2003 meeting in Halle, Germany, of the Society for Differential Psychology, Personality Psychology, and Psychological Diagnostics.

One sees in the poster a faint image of the eyes and forehead of Stern, onto which are superimposed likenesses of the schematics for variation studies and correlational studies shown in Fig. 3.1. Conspicuously missing from the poster are likenesses of the schematics presented by Stern (1911) for psychography and comparison studies. This omission is in full accord with the absence of any discussion of those latter two research schemes in the aforementioned textbooks by Anastasi and by Tyler. In graphics as well as prose, Stern has long been misrepresented as the advocate of a differential psychology committed entirely to the

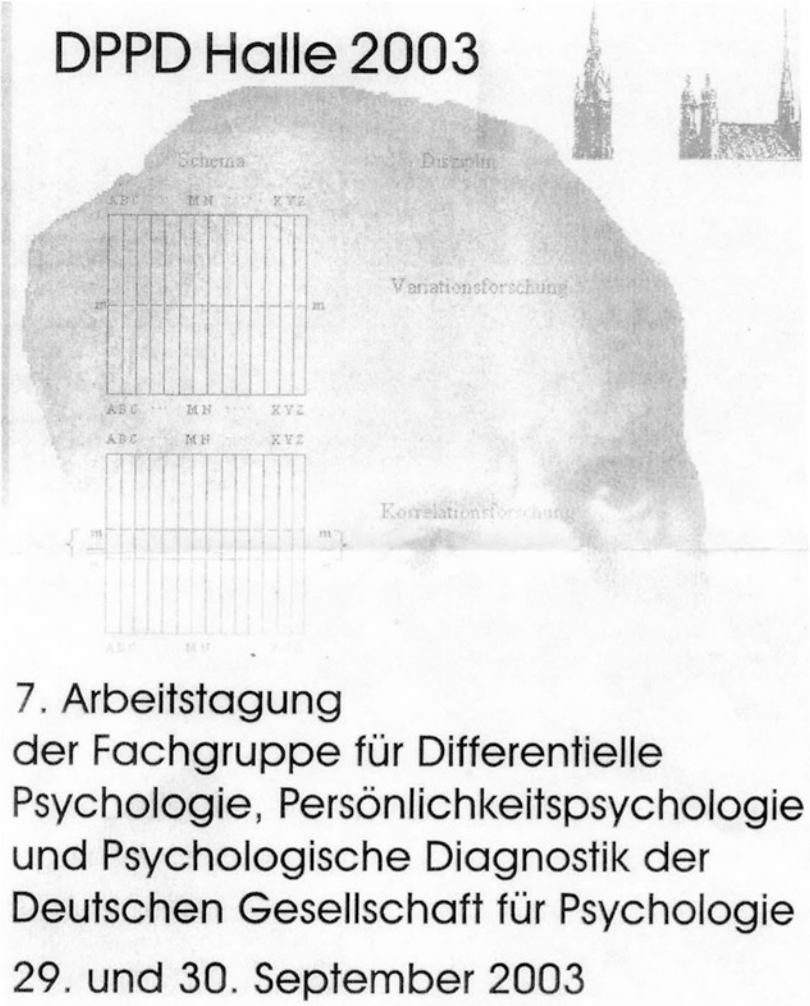

Fig. 3.2 Poster publicizing the 2003 meeting of the German Society for Differential Psychology, Personality Psychology, and Psychological Diagnostics

measurement and statistical analysis of variables marking individual and group differences.

As one final exhibit in this discussion of differential psychology's origin myth, consider the following passage:

> William Stern may be credited with originating the concept of differential psychology,and laying down some of the rules which should govern its methodology. He clearly argued for an empirical and statistical approach and for the separation from orthodox experimental psychology. He anticipated many modern developments, and ranks among the founders of our science. (Eysenck, 1990, p. 249)

Everything stated here by Eysenck is accurate. The problem, as will now be made apparent, lies in what Eysenck left out.

A Fuller View

It has already been noted that, from its inception, the primary instruments for studying variables marking between-person differences have been *tests* and *statistical methods* for analyzing measures generated by those tests. While Stern fully appreciated the potential usefulness of those instruments, he was also a steadfast proponent of qualitative methods, and he repeatedly issued warnings about and criticisms of investigators' widespread neglect of such methods in favor of the quantitative tools.

Already in the 1911 book, he advised his readers, at the conclusion of a chapter devoted entirely to a discussion of psychological tests, as follows:

> The test is only a—and not the—method for examining individuality. By no means does it render non-experimental methods of investigation superfluous.[17] To be sure, tests can supplement such methods. But tests are also supplemented by such methods, are often dependent upon such methods for the confirmation and elaboration of what they reveal, and in many cases must give way to what is revealed by those other methods. Psychological testing per se is to be regarded as a 'psychographic minimum'; it serves as a stopgap measure *(Notbehelf)* when time constraints or other circumstances will not admit of supplemental methods. It also serves as a method of preliminary investigation for the purpose of selecting from a large group some particular individual as a subject of further and more detailed psychographic investigation. (Stern, 1911, pp. 105–106)

Far from seeing test results as the only scientifically admissible indicator of an individual's standing with respect to some attribute variable, Stern stressed in this passage his belief that test results might not even be the best such indicator. There is scarcely any hint of this belief in discussions of Stern's perspective on individuality that are to be found in the extant literature.

In a 1916 discussion of 'IQ' as a measure of intelligence in so-called 'feebleminded' children, Stern warned further of the danger of overreliance on test scores:

> The feebleminded child presents a qualitatively different kind of development. One must resist the temptation to equate the psychological constitution of a 15-year-old feebleminded youth having a mental age of 9 with a 9-year-old of normal intelligence. Just as is true of the normal child, where the investigation of an intelligence type has its own significance over and above the investigation of an intelligence level, so also is it necessary in the study of the child who is not normal to take into account qualitative abnormalities alongside the quantitative subnormalities, and to ascertain the former through special methods of investigation. *The current inclination, prominent in America, to see in the test a single, comprehensive, and universally valid method is to be steadfastly opposed*. (Stern, 1916, pp. 16–17, emphasis added)

Here, too, Stern's grave doubts about the adequacy of purely quantitative information in the investigation of individual cases are evident. Five years later, his concerns in this regard were expressed yet again, in that instance directing his remarks specifically to psychotechnicians:

> Many psychologists, and nearly all of the general public view psychotechnics as consisting of psychological tests and nothing else. … (However), diagnoses based ontests alone are limited not only in fact but in principle, and hence require without exception supplementation through methods of direct observation … For the examinee in question, tests yield a number which makes it possible to assign the examinee a position on a quantitative scale but obscures qualitative particularities. The results of direct observation cannot be compared quantitatively, but enable instead a more nuanced psychogram … For all of these reasons, the method of direct

> observation must always be used as a supplement to the method of tests, and the former must be developed with the same diligence as the latter. (Stern, 1921, pp. 3–4, parentheses added)

Yet again eight years later, and with specific reference to personality testing, Stern emphasized his conviction that

> by dissecting a personality into elementary tests and their isolated application we do not draw closer to the essence of that personality but instead move further from it … (Hence) all attempts to portray a person in terms of an array of test scores are fundamentally false. (Stern, 1929, pp. 63, 65, parentheses added)

Following this statement, Stern renewed his call for the development of more qualitative methods of investigation. In a passage explicitly critical of the abiding dominance of correlational studies over psychographic investigations he wrote:

> If up to now we have been concentrating our efforts on the perspective afforded by correlational research, in which we look horizontally at the relationship between the different existing individual attributes, we must now look vertically at the individual, in a way that leads beyond the surface into the depth, and from the depth then again outward. (Stern, 1929, p. 69)

In September of 1929, Stern attended the Ninth International Congress of Psychology in New Haven, CT. Subsequent to that event, Stern spent several additional weeks visiting universities in the US northeast to gain a first-hand view of doings in the psychology departments of those various institutions. Upon his return to Germany, Stern authored a journal article discussing the impressions he had gained, and in that article wrote the following:

> The face of American psychology is characterized much less by laboratory experimentation than by testing procedures … Since the (First World) War, during which the entire American army was tested by means of a simple, standardized procedure for measuring intelligence,

> testing methods have been extended in ways that are astounding and almost troubling … Seventeen years ago, when I introduced the concept of the 'intelligence quotient' as a measurement principle for such intelligence tests, I had no idea that the 'IQ' would become a kind of worldwide formula and one of the most frequently encountered expressions in American technical jargon … But beyond that, batteries of tests for countless other psychological functions such as spatial perception, manual dexterity, attention, suggestibility, knowledge, arithmetic ability, character traits, etc. have now been developed, standardized, and put into use. … always with emphasis on the objective, quantitative norm, with reference to which the single case is then compared. At times, the primary objective in America seems to be to exercise technique, to obtain numerical measures which can be correlated and statistically analyzed. (What is clear) in all of this is the danger of a mechanization, … and it is to be hoped that the zenith of the testing culture will soon be a thing of the past. (Stern, 1930a, pp. 50–51, parentheses added)

Of course, the testing culture did not soon become a thing of the past (cf. Hanson, 1993). Indeed, correlational studies based on tests designed to measure individual differences in selected personality traits continued unabated, so that another three years later, Stern found occasion to observe disdainfully that in psychology

> (t)he practice is widespread of creating a kind of profile portrait or a list of traits. In America, this sort of exercise is now what passes as 'personality research.' (Stern, 1933, pp. 60–61)

A Concluding Observation

In April of the very year during which the last-quoted lines above were published, the Nazis banned Stern, a Jew, from all activity at the University of Hamburg, where he had been a prominent member of the faculty since helping to found that university in 1919. His last major publication, a general psychology textbook, was published two years later by a Dutch press (albeit in German; Stern, 1935). Had the course of history been other than it was, we can be certain that Stern would have continued his pointed criticisms of differential psychologists' excessive reliance on psychological tests and statistical analyses of the assessments of individual differences generated by those tests.

In any case, the record shows clearly that Stern maintained over the entire course of his academic life an appreciation for the distinction between knowledge of individuals and knowledge of individual differences, and for the indispensability of qualitative observations of individuals alongside of—and even, at times, in preference over—strictly quantitative knowledge (cf. Stern, 1930b). The record also shows that the overwhelming majority of differential psychologists, throughout the twentieth century and now well into the twenty-first, did not share Stern's views in these matters. In the next chapter, we will consider how the majority of experimental psychologists came to share the differential psychologists' views, and how the experimentalists' widespread adoption of the treatment group approach enabled the merger of scientific psychology's 'two disciplines' called for so persuasively by Cronbach (1957).

Notes

1. Henceforth in this work I will follow the lead of Danziger (1990) and refer to that model as the 'Leipzig model.'
2. For a vivid example of this, see Ebbinghaus (1964).
3. I shall use the expression 'between-person differences' as a more economical way of referring simultaneously to both individual and group differences. Actually, it can be shown that all statistically grounded discussions of 'between-person' differences are, finally, discussions of differences between groups of individuals. However, the fact of this matter is not transparent to most, and it will not be pursued here, as it is tangential to our immediate concerns.
4. This is stated in the middle of the book's title page, in the passage that reads: *An Stelle einer zweiten Auflage des Buches: Über Psychologie der individuellen Differenzen (Ideen zu einer differentiellen Psychologie).*
5. In this passage, the terms 'horizontal' and 'vertical' simply refer to the manner in which the four different research schemes are depicted in Fig. 3.1.
6. In the original German text, the words used by Stern here are *Merkmale* and *Individuen* (see Stern, 1911, p. 18).

7. Note that psychography does not qualify as a scheme for investigating individual differences even if in some given instance it incorporates, as it well might, attribute concepts that, at times, *also* are (or have been, or might be) used to investigate between-person differences in variation/co-variation studies. The failure to realize this is the source of the puzzlement expressed by Lundh (2015, p. 25).
8. This point is made in a footnote that appears on p. 321 of Stern's 1911 text: *Hier handelt es sich nicht … um die Biographie, sofern sie dem Psychologen Rohmaterial für seine Zwecke liefert, sondern um die Biographie als besondere Methode der Individualitätsdarstellung.*
9. The reader should note in passing that Thorndike's insistence on a purely quantitative conceptualization of between-person differences, coupled with the aforementioned equation of individuality with the differences, meant that, in his view, every individuality would have to be understood in terms of dimensions of differentiation presumed applicable to all individualities (see especially the graphic illustrations on pp. 22–24 of his 1911 monograph). This is also a point of contrast between Thorndike and Stern, as Stern's thinking allowed for the possibility of meaningfully characterizing individuals in terms of dimensional concepts that did not necessarily apply to all others and, indeed, might conceivably apply to no others (cf. Lamiell, 2003). This was one aspect of Stern's thinking found highly attractive by Gordon Allport (1897–1967; cf. Allport, 1937) but by few other twentieth century personality investigators. However, further pursuit of this point would lead away from our primary concerns in this chapter.
10. Proctor and Xiong (2018, p. 484) "think it likely" that, given Thorndike's strong mathematical proclivities, what he "intended to convey" with the wording 'is bound up with' was knowledge of a probabilistic sort. On that interpretation, a more precise rendition of what Thorndike meant would be that knowledge of the correlation between two variables in a population, coupled with knowledge of where a given individual within that population stands on one of the variables, allows one to know, for that individual, "the probability distribution for possible values on the other variable" (Proctor & Xiong, 2018, p. 484). Note that even if this interpretation of Thorndike is adopted, it still leaves Thorndike in the position of claiming that knowledge about variables within a population entitles claims to knowledge, probabilistic though they may be, about individuals within those populations (cf. Lamiell, 2018). It is that practice to which I am drawing the

reader's attention here. More of a critical nature will be said in Chapter 5 concerning the epistemically problematic nature of such claims.

11. Münsterberg was for a time widely regarded as the initiator of the expression 'psychotechnics,' though his friend and countryman, none other than William Stern, argued in his 1927 *Selbstdarstellung* that it was he who had coined the term in a 1903 publication (Stern, 1903).
12. That approach would eventually become formalized as factor analysis (cf. Cattell, 1952; Eysenck, 1952).
13. In the concluding section of this chapter we will consider Stern's growing disenchantment with developments in differential psychology.
14. This claim by Anastasi was retained in all subsequent editions of her text, which were published, respectively, in 1949, 1958, and 1981.
15. Readers not already familiar with the logical basis for this statement may wish to consult Cohen (1968) and/or Kerlinger and Pedhazur (1974).
16. In the light of these remarks, it is surprising that in a book published in 1978 and titled *Individuality: Human possibilities and personal choice in the psychological development of men and women* (Tyler, 1978), Tyler did not cite William Stern's work at all.
17. By "non-experimental methods of observation" (*nicht-experimentelle Beobachtungsmethode*) Stern meant qualitative observations that an investigator would make of examinees during the course of their being tested. In other places in his writings, Stern used the expression 'direct observations' to convey the same meaning.

References

Allport, G. W. (1937). *Personality: A psychological interpretation*. New York: Holt, Rinehard, & Winston.

Allport, G. W. (1946). Personalistic psychology as science: A reply. *Psychological Review, 53,* 132–135.

Anastasi, A. (1937). *Differential psychology: Individual and group differences in behavior*. New York: Macmillan.

Beck, S. J. (1953). The science of personality: Nomothetic or idiographic? *Psychological Review, 60,* 353–359.

Cattell, R. B. (1952). The three basic factor-analytic research designs—Their interrelations and derivatives. *Psychological Bulletin, 49,* 499–520.

Cohen, J. (1968). Multiple regression as a general data analytic system. *Psychological Bulletin, 70,* 292–303.
Cronbach, L. J. (1957). The two disciplines of scientific psychology. *American Psychologist, 12,* 671–684.
Danziger, K. (1990). *Constructing the subject: Historical origins of psychological research*. New York: Cambridge University Press.
Ebbinghaus, H. (1964). *Memory: A contribution to experimental psychology* (H. A. Ruger & C. E. Bussenius, Trans.). New York: Dover Publications.
Eysenck, H. J. (1952). *The scientific study of personality*. London: Routledge & Kegan-Paul.
Eysenck, H. J. (1954). The science of personality: Nomothetic! *Psychological Review, 61,* 339–342.
Eysenck, H. J. (1990). Differential psychology before and after William Stern. *Psychologische Beiträge, 32,* 249–262.
Hanson, F. A. (1993). *Testing testing: Social consequences of the examined life*. Berkeley, CA: University of California Press.
Kerlinger, F. N., & Pedhazur, E. J. (1974). *Multiple regression in behavioral research*. New York: Holt, Rinehart, & Winston.
Lamiell, J. T. (1987). *The psychology of personality: An epistemological inquiry*. New York: Columbia University Press.
Lamiell, J. T. (1997). Individuals and the differences between them. In R. Hogan, J. Johnson, & S. Briggs (Eds.), *Handbook of personality psychology* (pp. 117–141). New York: Academic Press.
Lamiell, J. T. (2003). *Beyond individual and group differences: Human individuality, scientific psychology, and William Stern's critical personalism*. Thousand Oaks, CA: Sage.
Lamiell, J. T. (2006). William Stern (1871–1938) und der «Ursprungsmythos» der differentiellen Psychologie. *Journal für Psychologie, 14,* 253–273.
Lamiell, J. T. (2018). Rejoinder to Proctor and Xiong. *American Journal of Psychology, 131,* 489–492.
Landy, F. J. (1992). Hugo Münsterberg: Victim or visionary? *Journal of Applied Psychology, 77,* 787–802.
Lundh, L.-G. (2015). The person as a focus for research: The contributions of Windelband, Stern, Allport, Lamiell, and Magnusson. *Journal for Person-Oriented Research, 1,* 15–33. https://doi.org/10.17505/jpor.2015.03.
Münsterberg, H. (1912). *Psychologie und Wirtschaftsleben: ein Beitrag zur angewandten Experimental-Psychologie* [Psychology and economic life: A contribution to applied experimental psychology]. Leipzig: Barth.

Münsterberg, H. (1913). *Psychology and industrial efficiency*. Boston and New York: Houghton-Mifflin.
Proctor, R. W., & Xiong, A. (2018). Adoption of population-level statistical methods did transform psychological science but for the better: Commentary on Lamiell (2018). *American Journal of Psychology, 131,* 483–487.
Samelson, F. (1974). History, origin myth, and ideology: Comte's "discovery" of socialpsychology. *Journal for the Theory of Social Behavior, 4,* 217–231.
Skaggs, E. B. (1945). Personalistic psychology as science. *Psychological Review, 52,* 234–238.
Stern, W. (1900). *Über Psychologie der individuellen Differenzen (Ideen zu einer "differentiellen Psychologie")* [On the psychology of individual differences (Toward a "differential psychology")]. Leipzig: Barth.
Stern, W. (1903). Angewandte Psychologie [Applied psychology]. *Beiträge zur Psychologie der Aussage, 1,* 4–45.
Stern, W. (1911). *Die Differentielle Psychologie in ihrer methodischen Grundlagen* [Methodological foundations of differential psychology]. Leipzig: Barth.
Stern, W. (1914). Psychologie [Psychology]. In D. Sarason (Ed.), *Das Jahr 1913: Ein Gesamtbild der Kulturentwicklung* (pp. 414–421). Leipzig: Teubner.
Stern, W. (1916). Der Intelligenzquotient als Maß der kindlichen Intelligenz, insbesondere der Unternormalen [The intelligence quotient as a measure of intelligence in children, with special reference to the sub-normal child]. *Zeitschrift für angewandte Psychologie, 11,* 1–18.
Stern, W. (1921). Richtlinien für die Methodik der psychologischen Praxis. *Beiheft zur Zeitschrift für angewandte Psychologie, 29,* 1–16.
Stern, W. (1927). Selbstdarstellung [Self-portrayal]. In R. Schmidt (Ed.), *Philosophie der Gegenwart in Selbstdarstellungen* (Vol. 6, pp. 128–184). Leipzig: Barth.
Stern, W. (1929). Persönlichkeitsforschung und Testmethode [Personality research and the methods of testing]. *Jahrbuch der Charakterologie, 6,* 63–72.
Stern, W. (1930a). Eindrücke von der amerikanischen Psychologie. Bericht über eine Kongreßreise [Impressions of American psychology: Report after travel to a conference]. *Zeitschrift für Pädagogische Psychologie, experimentelle Pädagogik und Jugendkundliche Forschung, 31,* 43–51 and 65–72.
Stern, W. (1930b). *Studien zur Personwissenschaft. Erster Teil: Personalistik als Wissenschaft* [Studies in the science of persons. Part one: Personalistics as science]. Leipzig: Barth.

Stern, W. (1933). Der personale Faktor in Psychotechnik und praktischer Psychologie [The personal factor in psychotechnics and practical psychology]. *Zeitschrift für angewandte Psychologie, 44,* 52–63.

Stern, W. (1935). *Allgemeine Psychologie auf personalistischer Grundlage* [General psychology from the personalistic standpoint]. Den Haag, the Netherlands: Nijhoff.

Terman, L. M. (1924). The mental test as a psychological method. *The Psychological Review, 31,* 93–117.

Thorndike, E. L. (1911). *Individuality*. New York: Houghton-Mifflin.

Tyler, L. E. (1947). *The psychology of human differences*. New York: Appleton-Century-Crofts.

Tyler, L. E. (1978). *Individuality: Human possibilities and personal choice in the psychological development of men and women*. San Francisco: Jossey-Bass.

Tyler, L. E. (1985). *Neglected insights in personology.* Unpublished manuscript, Eugene, OR.

4

The Failure of Critical Thinking in the Statistization of Experimental Psychology

The reader may recall from the discussion in Chapter 1 that in a 1913 essay Wilhelm Wundt expressed his concern that if psychology should divorce itself from philosophy, psychologists would, in time, become incapable of dealing effectively with the conceptual issues that, in any science, necessarily arise in concert with the discipline's technical and empirical concerns (Wundt, 2013). In that event, Wundt argued, psychologists will have reduced themselves from true scientists to mere tradesmen (*Handwerker*), and, at that, Wundt opined, tradesmen 'not of the most useful variety (Wundt, 2013, p. 206).' In his view, the very survival of psychology as a distinct and genuinely scientific discipline was at stake.

Revisiting Wundt's critical essay one century later, I argued that his concerns were, alas, proving prescient (Lamiell, 2013); that as a scientific discipline properly devoted, in the words of William Whewell (1794–1866), not only to 'the colligation of facts' but also to 'the clarification of concepts' (quoted in Machado & Silva, 2007, p. 680), psychology was showing clear signs of degenerating into a discipline largely indifferent toward—and sometimes, even, openly disdainful of—the latter of these two objectives. In this same vein, Gantt and

J. T. Lamiell, *Psychology's Misuse of Statistics and Persistent Dismissal of its Critics*, Palgrave Studies in the Theory and History of Psychology,
https://doi.org/10.1007/978-3-030-12131-0_4

Williams (2018) have written recently on the 'hijacking' of a genuinely scienti-*fic* psychology, in the sense averred by Whewell, at the hands of an impostor dominated by an utterly scienti-*stic* ethos of essentially the sort that Wundt (2013) feared (cf. Lamiell, 2018).

The focus in this chapter is on the failure of critical thinking within the mainstream of experimental psychology with regard to matters of major conceptual significance as that side of the discipline transitioned from what Danziger called the 'Wundtian' or 'Leipzig' model for conducting experiments to the radically different 'neo-Galtonian' or 'treatment group' model (Danziger, 1987, 1990).[1] It was in the course of that transition that experimental psychologists allowed their field to become 'statisticized'; i.e., dominated by the statistical methods of investigation that were already being utilized so extensively in correlational (differential) psychology (refer to previous chapter). It was this statisticized version of experimental psychology—and, emphatically, *not* the original Leipzig model—that Lee J. Cronbach (1916–2001) correctly saw as structurally conformable with differential psychology. This realization is what made plausible the call for the merger of scientific psychology's 'two disciplines' that Cronbach (1957) found so urgently needed, and that, to his clear satisfaction, was quickly and widely heeded (cf. Cronbach, 1975).

Conspicuously absent over the entire period of this historical development was a critical discussion of the conceptual implications of the obvious procedural differences between Leipzig model and treatment group experimentation. The rightful place of that much-needed discussion was usurped by a woefully uncritical assumption of full epistemic continuity in the transition from the former model to the latter. That assumption is reflected in the paradigmatic belief that treatment group experimentation would be formally suited to the same overall knowledge objective—that of discovering the general laws presumed to regulate the psychological functioning of individuals—as had been established by the original Leipzig model experimentalists.

The final segment of this chapter is devoted to an elaboration of how the findings of treatment group experimentation must be understood in order to uphold that paradigmatic belief. That discussion is intended to make explicit ideas that long ago should have been clearly set forth and subjected to thorough critical analysis.

Experimental Psychology's Two Disciplines

By 1900, many psychologists were growing impatient with the experimental psychology that Wundt had pioneered in Leipzig scarcely 20 years earlier. The so-called 'brass instruments' research was widely seen to offer little if anything that could be helpful in addressing practical concerns arising outside the university-based research laboratories, in various societal domains such as education, health care, business and industry, and the military. That lack of immediate practical relevance did not trouble Wundt, for while he favored a viable applied psychology in principle, he believed that, at the time, the still youthful science lacked a knowledge base sufficient for informed and scientifically responsible applications. He also believed that premature efforts at practical relevance could misfire, to the detriment of both the intended beneficiaries and psychology as a scientific discipline. Further, he was concerned that a psychology prematurely concerned with practical applications would become a discipline in which research would be driven primarily by questions arising outside of the discipline, instead of by questions of basic theoretical importance arising from within (Wundt, 1909).

In this latter connection, Wundt (1909) elaborated briefly on the then-burgeoning symbiotic relationship between psychology and pedagogy.[2] Wundt wrote that among those psychologists interested in applying the findings of psychological research in pedagogical contexts, the questions deemed worthy of psychological research to begin with were already being formulated with an implicit pedagogical slant. He continued:

> The inevitable consequence of this is a narrowing of perspective, so that the threat arises of psychology becoming more and more an applied pedagogy. Then, the pedagogical psychologist not only defines his tasks in strict accordance with the needs of pedagogy, but, beyond that, also seeks out for the solution of those tasks only resources accessible within the range of (previous) pedagogical observations and experiments, without considering experiences that have been gained outside those circles. In this way psychology is threatened—one must apologize for the word—with becoming the prey *(Beute)* of pedagogy: not only are works relevant to purely psychological themes inadvertently transformed into pedagogical

> tasks, but (in their execution) use is made almost exclusively of resources that have been previously accumulated in the service of (other) pedagogical goals. (Wundt, 1909, p. 17, parentheses added)

Clearly, Wundt believed that a psychology prematurely oriented toward applications could be preyed upon by non-psychological disciplines other than pedagogy in like fashion. In any case, and notwithstanding his reservations, many psychologists continued to seek opportunities to apply their research skills beyond the confines of psychology's experimental laboratories. This meant that many psychologists were forsaking experimental psychology in favor of the newer sub-discipline of differential psychology (Danziger, 1990). There, the prospects for non-laboratory applications of research findings appeared brighter, especially in the domains of intelligence testing and other forms of psychological assessment (cf. Sokol, 1990). However, a downside of that disciplinary migration was that psychological science was becoming, increasingly, a strictly correlational discipline, with its power to establish relationships of a cause–effect nature through controlled experimentation correspondingly compromised. As Danziger (1987, 1990) has so effectively explained, treatment group experimentation rose to favor among psychologists as an apparent means of reclaiming for their discipline the power to establish cause–effect relationships while simultaneously retaining the practical usefulness of the same statistical tools that were being employed so effectively in correlational studies.

The Two Disciplines of Experimental Psychology Briefly Described

In Leipzig model experimentation, an investigator would bring a subject into the laboratory, expose that subject to the experimental manipulations of interest (e.g., systematic variations in the luminance of a visual field, or in the volume of a tone), and then record that subject's responses (e.g., in terms of reaction times). Whatever the particulars, it was in the systematic covariation between the strictly controlled experimental manipulations and an individual subject's responses where investigators would find empirical grounds for cause–effect inferences.

For our present purposes, three points about this model of experimentation merit emphasis.

First (and as has been mentioned previously), there was no role within this model for the methods of statistical analysis that are now regarded by many as essential to psychology's status as a science. When the field's original experimenters carried out statistical analyses at all, the multiplicity of assessments on which the calculations were performed had been generated through observations of the same individual subject, and the knowledge objective of the analyses was to gauge the degree of error in the obtained assessments, and not to estimate population parameters (cf. Danziger, 1987).

Second, the *results* obtained with a given individual subject were the complete *findings* of the experiment. As Danziger (1987) correctly noted in this connection, 'any (subsequent) increase in the number of experimental subjects above one constituted a replication of the experiment (Danziger, 1987, p. 38, parentheses added).'

Third, the feasibility of carrying out such replications with additional subjects, investigated one at a time, is precisely what made single-subject experimentation logically compatible with the scientific quest for knowledge of the general laws presumed to govern various aspects of human mental functioning. Any apparent contradiction in this regard disappears when one appreciates that among scientific psychology's original experimenters, the claim that some pattern of systematic covariation between experimental manipulations and a subject's responses held true 'in general' meant that it held in like fashion for each one of the investigated cases. In the native language of Wundt and a great many of the other original experimental psychologists, the word for 'general' in the sense relevant here, *allgemein*, developed etymologically as a contracted form of the expression *allen gemein*, which means '*common to all.*' Among those investigators, then, an empirical finding would be regarded as holding true 'in general' only if found to hold—within acceptable limits of approximation, hence the previously mentioned need for statistical estimates of assessment errors—for every individual case investigated up to a given point in time.[3]

Treatment group experimentation proceeds altogether differently. In the most basic and still most widely used version of that model, an investigator secures a random sample of indefinitely many subjects (the more, the better) from some population, and then randomly assigns each of those *N* total subjects to one of two or more experimental treatment conditions defining the experiment's so-called 'independent variable' (IV). Each subject's performance under the experimental condition to which s/he had been randomly assigned is then registered in terms of the experiment's so-called 'dependent variable' (DV). The average DV performance across all *n* subjects assigned to each IV condition is then computed, and the statistical significance of the difference(s) between the two or more treatment group averages is then determined by a procedure that, in its essence, evaluates the obtained between-mean difference(s) against some index of how large of a difference between the means could have been expected on the basis of chance alone.

Historically, the exact form taken by this just-mentioned index has varied as the test statistic favored by the experimentalists has changed from the critical ratio to the *t*-test to the F-ratio (cf. Rucci & Tweney, 1980). In any case, the essential function of the index has always been the same: to serve as an empirical estimate of the variation between treatment group means that could be expected quite apart from any effect caused by the experimental treatments. If the magnitude of that index is sufficiently exceeded by the difference(s) between the treatment group averages that an experiment has empirically revealed, it is inferred that those obtained differences did not occur by chance alone, and that, instead, they were causally produced by the differential treatments defining the experiment's IV(s).

Note that in stark contrast to Leipzig model experimentation, no individual subject's results constitute the findings of a treatment group experiment. On the contrary, such an experiment's *findings* remain undetermined until *all* of the individual subjects' results have been registered and entered into the calculations described above. In treatment group experimentation, it is the outcome of the statistical comparison of the treatment group means that constitutes an experiment's findings.

The Historic Failure of Critical Reflection

One searches the archival literature in vain for any critically penetrating discussion of the epistemic implications of the rather prominent procedural differences, just described, between scientific psychology's two experimental disciplines. In the lacuna created by that critical outage, the implicit presumption could—and did—prevail that, epistemically speaking, there was no essential difference between the two experimental psychologies, their obvious methodological differences notwithstanding. This implicit presumption is what enabled Cronbach (1957) to write, as he did, of scientific psychology's 'experimental' discipline as if it had been, since its inception in Leipzig in 1879, a unitary entity. After all, though Cronbach (1957) alluded to Wundt's historical importance in the founding of that entity, he left entirely unmentioned the fact that the version of experimental psychology that Wundt founded was so very different from the experimental psychology that had come to define the discipline by 1957. As it happened, that oversight was not unprecedented.

Some two decades earlier, John Frederick Dashiell (1888–1975), had devoted an appreciable portion of his 1938 APA Presidential address to what he saw as a coming *rapprochement* between psychology's experimentalists, on the one side, and its clinicians on the other (Dashiell, 1939). In his article, too—an article cited by Cronbach (1957)—Dashiell (1939) wrote of experimental psychology in ways that cried out for a critical discourse that never materialized. Within the context of our present concerns, therefore, it is instructive to consider briefly some of what Dashiell (1939) had to say.

Although optimistic about an eventual coordination of psychology's experimental and clinical branches, Dashiell emphasized that he did not regard that rapprochement as having yet been achieved. With that in mind, Dashiell declared himself intent on 'bring[ing] into sharper focus the differences between [the two]' (Dashiell, 1939, p. 13, brackets added).

Dashiell noted that, for the clinician, the primary interest was in 'the peculiar makeup of the individual person' (p. 12). For the experimentalists, he stated by way of contrast, matters were quite different:

> We (experimentalists) study such things as the various factors that modify a *typical* human being's reaction time, or the relationship of time interval to loss of ability to reproduce a memorized passage by adults *in general,* or the dependence of the variable of problem-solving ability upon the age variable in the *average* child. (Dashiell, 1939, p. 13, all emphases in original; parentheses added)

Striking here is Dashiell's effective presumption of full complementarity between the expressions *typical, in general,* and *(on) average*. While such complementarity could, indeed, properly be said to exist within the framework of treatment group experimental psychology, it could *not* properly have been said to exist for the original Leipzig model of experimental psychology employed by Wundt and his contemporaries. To the best of my knowledge, no one ever publicly contested Dashiell on this very significant conceptual point.

To the practitioners of the Leipzig model, as we have already seen, the expression 'true in general' did *not* mean 'true typically' or 'true on average.' It had the altogether different meaning of '*true in common for all*' investigated cases. Yet, and just as would Cronbach two decades later (refer above), Dashiell (1939) alluded to Wundt's historical importance as the founder of experimental psychology without drawing any distinction between the kind of experimental psychology Wundt prosecuted and the treatment group experimental psychology that had come to predominate by 1939 (Danziger, 1987, 1990; Rucci & Tweney, 1980). Failing to draw that distinction, Dashiell blinded himself—and countless readers of his article—to the fundamental conceptual inappropriateness of his presumption of full complementarity in the sense just described.

Striving to contrast further the clinician's concern for the peculiarities of individual cases and the experimentalist's broader knowledge objectives, Dashiell (1939) remarked as follows:

> A natural science, being interested solely in educing uniformities of nature, does not find the individual an object of concern save as it illustrates an old law or suggests a new one… (In) all true experimentation, the specific instance, once it has been noted, is tossed aside like a

> squeezed lemon, its juice having been extracted and compounded with that of numerous others. (Dashiell, 1939, p. 13, parentheses added)

What the contemporary reader must firmly grasp about this claim is that while it is *perfectly* compatible with the views of the Leipzig model experimentalists, its compatibility with the treatment group experimentation Dashiell himself practiced is, in fact, highly questionable. Committed as the Leipzig model experimentalists were to the quest for knowledge of the general laws presumed to govern various aspects of human mental functioning—where, it must be remembered, 'general' meant 'common to all investigated cases'—it is true, just as Dashiell (1939) stated, that they were not interested in or inclined to be distracted by any indications of research subjects' respective peculiarities or *individualities*.[4] For reasons just explained, however, this could not have equated to the complete disinterest of those investigators in *individuals*. The very nature of their scientific mission logically mandated an interest in individuals, because woven into the very conceptual fabric of their discipline was the understanding that only through experimentation with individuals in particular could scientific evidence possibly be adduced in favor of a claim that some lawful regularity held *for individuals in general*, i.e., held in the sense of being 'common to all' of the individual cases investigated (bearing in mind the caveat in endnote 3).

To reemphasize the crucial point here: the language used by Dashiell (1939) in the passage quoted immediately above conforms fully to the epistemic commitments of the Leipzig model experimentalists. He indicated that the individual case is of no interest *except* ('save') 'as it illustrates an old law or suggests a new one.' He noted that the 'juice' of *every* individual case must *first* 'be extracted and compounded with that of numerous others' *before* it is 'tossed aside like a squeezed lemon.'[5] The formal suitability of Leipzig model experimentation to these ideas is transparent just because, in accordance with that model, experimental findings are fully defined by the results obtained in each and every individual case investigated. What is problematic here is the absence of critical attention to the question of just how these same epistemic requirements are supposed to be met by treatment group experimentation. We have already seen that under the terms of that model, experimental findings are *not* defined for *any* individual subject investigated. How, then, are the statistical inferences built into null hypothesis significance testing supposed to accomplish the case-by-case 'extractions' of which Dashiell (1939) wrote?

The dreadful epistemic situation that psychologists were creating for themselves is perhaps at this point in the discussion sufficiently clear: completely escaping critical discussion during the transition from Leipzig model to treatment group experimentation was the assumption that the overriding knowledge objective of the original experimental psychology, which was to discover the general laws presumed to govern various aspects of *individual* functioning, could be pursued at least as well—and perhaps even better—through statistical analyses of the sort built into treatment *group* experimental psychology. Even more pointedly, and consistent with the fact that that notion was not subjected to any serious critical examination, one also finds nowhere any clear explication of just *how* treatment group experimentation, and the logic of statistical inference on which that form of experimentation relies, can be said to reveal cause–effect relationships in individual functioning in the fashion of Leipzig model experimentation.

In fact, the historic transition from Leipzig model to treatment group experimentation proceeded on the widespread and uncritically accepted *assumption* of full epistemic continuity between the two forms of investigation. Due in large measure to the fact that that assumption never was pointedly challenged, neither was the need to thoroughly explicate and justify that assumption ever acknowledged. In effect, its eventual status as dogma happened by default, and it was in the course of that utterly uncritical development that the currently favored statistical methods of investigation became canonical.

In Chapter 1, extensive discussion was devoted to the rare expressions of concern over this matter by Bakan (1955, 1966) and by Kerlinger (1979).

As the reader may recall, Bakan (1955, 1966) insisted that the distinction be drawn and scrupulously maintained between 'general type' and 'aggregate type' propositions. Clearly, his understanding of 'general' was the same as that which had prevailed among Leipzig model experimentalists, and he understood aggregate type propositions as referring to statistical averages on the order of those examined in treatment group experimentation. Alas, Bakan's urgings were never heeded, and they have continued to be ignored right up to the present.

Meanwhile, in the late 1970s, Kerlinger (1979) wrote of what he termed the 'troublesome paradox' created by the conceptual gap between psychologists' *theoretical* concerns for individual-level doings and their *methodological* commitments to group-level experimentation. Kerlinger (1979) offered no solution for this paradox at the time, and it never subsequently became a matter of substantial concern within the discipline—not even to Kerlinger himself.

At long last, the historian of psychology Kurt Danziger did note in his 1987 discussion of statistical methods and the historical development of research practice in American psychology that 'where treatment group experimentation was expected to throw light on psychological processes in individuals, the gap between the goal and the statistical data base became a problem' (p. 45). He then observed that in the face of that problem

> … there arose a kind of modeling in which the statistical structure of the data based on the responses of many individuals is assumed to conform to the structure of the relevant psychological processes operating on the individual level. (Danziger, 1987, p. 45)

Significantly, Danziger (1987) did not cite any publications explicitly advocating this kind of modeling or elaborating on its precise nature, a fact wholly consistent with my contention that 'default' was the mechanism whereby the investigative practices now so widely accepted became paradigmatic. In the remainder of this chapter, I seek to make explicit what the interpretive practice described by Danziger (1987) logically entails.

The Implicit Epistemic Commitments of Treatment Group Psychological Experimentation

In the oft-previously mentioned (and highly influential) article by Cronbach (1957), he noted that

> … (t)he well-known virtue of the experimental method is that it brings situational variables under tight control. It thus permits rigorous tests of hypotheses and confident statements about causation. (Cronbach, 1957, p. 672)

This virtue is one that was certainly recognized and embraced by scientific psychology's original experimenters, a point made explicitly by Danziger (1987) as follows:

> The classical Wundtian (Leipzig-model) experiment was designed to throw light on causal psychological processes operating in individual minds. It did this by systematically varying experimental conditions and observing the results. … These causal processes existed in individual minds, and the experiment was designed to explore them in this individual context. (Danziger, 1987, pp. 37–38, parentheses added)

As argued above, the practitioners of treatment group experimentation implicitly assumed full epistemic continuity between that form of investigation and the procedures of the Leipzig model it was supplanting. Our question at this point is: How must treatment group experimentation be understood in order to be seen as consistent with this assumption? As a practical means of addressing this question, the reader is asked to consider a hypothetical and very simple treatment group experiment, similar to one I have discussed at greater length elsewhere (Lamiell, 2015).

An Illustrative Hypothetical Experiment

Let us suppose that in this experiment each of 30 elementary school pupils was assigned at random to one of two treatment conditions, each of which was defined by a particular method of teaching fourth grade spelling. Fifteen (15) pupils were assigned to each method of instruction. At the conclusion of the instructional period, the performance of each pupil on a common spelling test was recorded, and those results were then analyzed statistically by means of a simple, one-way analysis of variance (ANOVA).[6]

For ease of discussion, let us adopt the following symbolization conventions:

Y symbolizes the dependent variable, spelling test scores;
${}_{G}M_{Y}$ symbolizes the grand mean of Y, the spelling test scores, computed across all 30 pupils included in the experiment;
${}_{1}M_{y}$ represents the mean of the spelling test scores among the 15 pupils submitted to experimental teaching method 1;
${}_{2}M_{y}$ represents the mean of the spelling test scores among the 15 pupils submitted to experimental teaching method 2.

The first step in the ANOVA was, as always, to compute the overall DV mean, ${}_{G}M_{Y}$. Note that the particular value assumed by that mean is something left entirely unexplained by statistical analyses of this sort. The analysis simply ascertains that mean as a starting point and proceeds from there, focusing on the (squared) deviations of the respective treatment group means from that grand mean (and thus, by extension, from each other).

Let us now suppose that the ANOVA has revealed a statistically significant[7] 'effect' attributable to the treatment variable, method of instruction, such that the average test performance among the 15 pupils exposed to teaching method 1 was $({}_{1}M_{y} - {}_{G}M_{Y})$ units on DV scale below the grand mean, while the average test performance among the 15 pupils exposed to teaching method 2 was $({}_{2}M_{y} - {}_{G}M_{Y})$ units on DV scale above the grand mean. The first of these two differences is viewed as the empirical manifestation of the *collective* 'effect' of teaching method 1 on spelling test performance within a sample of children exposed to that treatment, while the second of the two differences is viewed as the empirical manifestation of the *collective* effect of teaching method 2 on spelling test performance within a sample of children exposed to that treatment.

Note that for the purposes of one concerned primarily with the practical implications of such research findings, knowledge of the statistically significant difference between the two treatment group means could suffice. For example, an educational psychologist with applied interests could see in such findings warrant for the recommendation that teaching method 2 be implemented in the local schools, in the scientifically grounded expectation that the spelling achievement of future fourth grade pupils in those schools would be better, on average, under that instructional method than it would be, on average, under instructional method 1. No consideration of a cause–effect relationship would be necessary.

However, given the status of the investigation as a true experiment by the accepted standards of treatment group experimentation, a claim to knowledge of a cause–effect relationship *could* be introduced here, and it is essential to realize that for this purpose more than knowledge of a statistically significant IV–DV relationship is called for. As Harré (1981) noted, the psychological investigator must understand, as a true scientist, that 'causal processes occur only in individual beings, since mechanisms of action, even when we act as members of collectives, must be realized in particular persons' (Harré, 1981, p. 14).

In this illustrative case, the psychological experimenter's putative concern, as a basic scientist, for cause–effect regularities manifesting themselves among his/her research subjects *in general* would necessitate an understanding of the statistical results as indicating that the 'effects' of the treatment variable (IV), method of instruction, were realized not merely collectively, as stated earlier, but *in each one* of the particular persons exposed to one or another of the treatments in question. To put matters concretely, if we are to regard the 'effect' of treatment 1 to have been manifested empirically as the quantity $({}_{1}M_{y} - {}_{G}M_{Y})$, it must be assumed that exposure to teaching method 1 drove down from the grand mean the performance of *each one* of the 15 pupils exposed to that treatment by $({}_{1}M_{y} - {}_{G}M_{Y})$ units on the scale defining the experiment's DV. In like fashion, it must be assumed that exposure to teaching method 2 drove up from the grand mean the performance of *each one* of the 15 pupils exposed to that treatment by $({}_{2}M_{y} - {}_{G}M_{Y})$ units on the scale defining the experiment's DV.[8]

While the 'each one' proviso here might surprise many readers, unaccustomed as most are to thinking critically about the assumptions built into their own experimental work, careful reflection reveals that that proviso is logically essential *if* treatment group experimentation is to be regarded as epistemically continuous with Leipzig model experimentation. As we have seen, the objective of the latter was to gain knowledge about the effects of experimental manipulations on (various facets of) the mental functioning of *individual* subjects *in general*, and Leipzig model experimentation was logically suited to that objective because experimental findings generated by such experimentation were fully defined for individual subjects. It is the fact that the findings of treatment group experiments are not defined for individual subjects that makes it necessary to *assume* the validity of the 'each one' proviso.

The necessity of that assumption is further underscored by briefly considering the consequences that would follow from relaxing it. To be sure, it is always possible for the treatment group experimentalist to concede uncertainty that a given statistically significant treatment variable 'effect' was realized in 'this' or 'that' particular case, since there is no direct empirical evidence on the basis of which to decide the matter one way or the other. However, since this is true in the consideration of *any* particular case, the concession of uncertainty of this sort in one individual case would be, in effect, a concession of uncertainty in each case considered individually. At that point, there would be nothing for the treatment group experimentalist to claim to *know* on the basis of his/her empirical findings other than that the treatment 'effect' discovered in his/her experiment had held, *on average*, for *aggregates* of research subjects, the treatment groups, considered as unitary entities. Conceded here would be that the treatment 'effect' discovered through the experiment had not been shown, and so could not be known, to have held *in general* for the *individual* subjects *within* the respective treatment groups. At that point, all pretenses of epistemic continuity between Leipzig model and treatment group experimentation would have to be surrendered, precisely the endpoint that mainstream thinking has long sought to avoid. The point here is that the 'each one' assumption is the toll exacted for that avoidance.

The Boundless Largesse of Ceteris Paribus

On the face of things, the integrity of the 'each one' assumption just discussed seems empirically contraindicated by the ubiquitous fact of differences in DV status among research subjects who have been exposed to the same experimental treatment. After all, if the causal effect of some given experimental treatment must be assumed identical for every subject submitted to that treatment, then, all other things being equal, all subjects submitted to that treatment should manifest the same DV status. The key here, of course, is the qualifier 'all other things being equal,' a qualifier often articulated in scholarly discourse by the handy Latin expression ceteris paribus.

To appreciate the full epistemic load carried by this seemingly innocent disclaimer, it is once again useful to consider the extreme hypothetical circumstance under which it would be possible to eschew that disclaimer. If for each of the two methods of instruction in our hypothetical experiment, all 15 of the subjects involved had, in fact, scored identically on the spelling test used to define the experiment's DV, then within each treatment group, each pupil's DV score would have been identical to the DV mean for his/her group, and the overall statistical analysis of the data would have revealed perfect covariation between the independent and DV of the experiment. Under those circumstances, and in full accord with contemporary mainstream thinking, the conclusion would be drawn that method of instruction had causally determined, fully and exclusively, the spelling performance of every one of the 30 pupils studied.

In real experiments of this sort, one finds that in at least some of the individual cases exposed to a given treatment—and perhaps even in all of them—the actual DV scores of individual experimental participants do not equal the average score for that treatment. In the thinking of treatment group experimentalists, such findings need not and have not cast doubts on the validity of the 'each one' assumption embedded within the approach. Instead, they are commonly interpreted as evidence that individual subjects' respective performances were influenced by other factors in addition to, in opposition to, or, perhaps, in complex interaction with, *but not instead of*, the effects of the treatments under immediate consideration. This 'not instead of' proviso is crucial, ultimately for the same reason as the 'each one' assumption discussed above: the experimenter who would relax the 'not instead of' proviso in the case of 'this' or 'that' individual subject would have to be prepared to relax it in the case of *any* individual subject considered as such. S/he would ultimately be left with no basis upon which to claim experimental evidence that his/her IV had exerted its putative effect on his/her individual subjects *in general*, and, with that, the presumed epistemic continuity in the transition from Leipzig model experimentation to treatment group experimentation would once again be severed.

Concluding Comment

By thinking along the lines sketched above, treatment group experimentalists seeking knowledge of general lawfulness in individual-level doings, as the Leipzig model experimentalists clearly were, find no reason to doubt the *relevance* of the knowledge generated by their investigations to that overriding knowledge objective, even in the light of empirical evidence that, on its face, could be taken to indicate otherwise. Appeal to ceteris paribus offers psychology's treatment group experimentalists permanent safe harbor, making it possible for them to view aggregate statistical findings as relevant to explanations for individual-level outcomes 'in general,' while at the same time conceding that those explanations are *incomplete*—but not irrelevant—so long as the obtained statistical relationships are not perfect. It is just this incompleteness that is explicitly acknowledged in the refrain, by now virtually *de rigueur* in reports by research psychologists of their empirical findings, that 'further research is needed.' Left unsaid—but unavoidably implicitly assumed—is the notion that the requisite 'further research' should have the same basic design features as its precedent(s). This is how the methodological canon of mainstream experimental psychology perpetuates itself.

Through the lenses of treatment group experimentation, the task of filling out incomplete accounts of individual-level doings in a scientifically valid way is seen as a matter of securing ever stronger statistical relationships between variables. The ultimate objective, of course, is to account for all of the variance in the criterion/dependent variables, whether that variance has been created experimentally or captured non-experimentally (e.g., by tests), or both, within a given study. Even if it is acknowledged that, in practice, the objective of accounting for all of the criterion/dependent variable variance can never be fully achieved, the mainstream conviction is that that objective can and should be pursued to the fullest extent possible. Such can be done through further and perhaps more elaborate treatment group experimentation, or by correlational studies introducing additional variables reflecting preexisting individual and/or group differences (e.g., sex differences, personality characteristics, intelligence levels) or, in full-blown Cronbach-ian (1957) fashion, through the simultaneous consideration of experimental treatments and preexisting individual and/or group differences in 'hybrid' experimental designs (see also Cronbach, 1975).

Whatever the specific design features, the logic of this entire approach as a framework for psychological investigation demands that the statistical relationships between variables defined for *aggregates* of research subjects serve as the empirical basis for claims to *generally* valid accounts of individual-level doings, however incomplete those accounts might be at any given point in time. In the next chapter, we will consider the deep and irremediable conceptual flaws embedded in this entire way of thinking as a paradigm for a scientific psychology.

Notes

1. It is due largely to the invaluable scholarship of Kurt Danziger (b. 1926) that there exists in the archival literature an incisive account of that transition, and in the discussion that follows, I have leaned extensively on that scholarship. Throughout, I follow Danziger's lead in referring to the original model for psychological experiments as the 'Leipzig' model.
2. Ernst Meumann (1862–1915), the acknowledged founder of experimental pedagogy, had completed his doctoral studies under Wundt in Leipzig.
3. We may be sure that psychology's founding scientists, educated scholars that they were, knew well the logical limits of induction that constrain all scientific inquiry. They knew, in other words, that any declaration of the sort 'X is true in general' required, at least implicitly, the caveat 'until further notice,' i.e., unless and until some subsequently investigated case(s) would indicate otherwise.

4. This is precisely why in his 1900 book, Stern began by declaring the 'problem of individuality' to be *the* primary challenge facing twentieth-century scientific psychology (refer to previous chapter).
5. As an aside for now, we note here that Dashiell (1939) continued as follows: 'Note 2 things: (a) we do not isolate the individual person *qua* individual person; but (b) we do isolate the particular function or phenomenon in question' (p. 13). In effect, Dashiell was distinguishing here between knowledge of variables (functions, phenomena) and knowledge of individuals, but doing so while appealing implicitly to the notion that *some* knowledge of the latter is secured along the way to knowledge of the former. This implicit appeal runs counter to the distinction insisted upon by Stern (1911) in differential psychology between knowledge of variables in terms of which individuals are differentiated and knowledge of the individuals differentiated in terms of those variables. In this, then, Dashiell (1939) was aligning himself with the views of the differential psychologists Thorndike and Münsterberg, as discussed in the previous chapter. I will argue in Chapter 5 both that and why Stern was correct in insisting upon the variables-individuals distinction, and that this distinction is no less relevant to treatment group experimental psychology than to differential psychology. Indeed, I will explain why treatment group experimental psychology is, in effect, a species of differential psychology.
6. The results might just as well have been analyzed by means of a simple t-test for independent groups, or by means of correlation/regression procedures. These superficially different methods of analysis are, in fact, equivalent (cf. Cohen, 1968; Kerlinger & Pedhazur, 1974).
7. The reader may suppose that 'statistical significance' was determined according to the standard criterion of $p < .05$.
8. It is perhaps worthy of note in this connection that in the computation of the sum-of-squares for the treatment variable in this experiment, method of instruction, the square of the difference between a given treatment group mean and the grand mean is multiplied by n, once for each of the subjects exposed to that treatment.

References

Bakan, D. (1955). The general and the aggregate: A methodological distinction. *Perceptual and Motor Skills, 5,* 211–212.

Bakan, D. (1966). The test of significance in psychological research. *Psychological Bulletin, 66,* 423–437.

Cohen, J. (1968). Multiple regression as a general data analytic system. *Psychological Bulletin, 70,* 292–303.

Cronbach, L. J. (1957). The two disciplines of scientific psychology. *American Psychologist, 12,* 671–684.

Cronbach, L. J. (1975). Beyond the two disciplines of scientific psychology. *American Psychologist, 30,* 116–127.

Danziger, K. (1987). Statistical method and the historical development of research practice in American psychology. In L. Krueger, G. Gigerenzer, & M. S. Morgan (Eds.), *The probabilistic revolution, Vol. 2: Ideas in the sciences* (pp. 35–47). Cambridge, MA: MIT Press.

Danziger, K. (1990). *Constructing the subject: Historical origins of psychological research*. Cambridge: Cambridge University Press.

Dashiell, J. F. (1939). Some rapprochements in contemporary psychology. *Psychological Bulletin, 36,* 1–24.

Gantt, E. E., & Williams, R. N. (Eds.). (2018). *On hijacking science: Exploring the nature and consequences of overreach in psychology*. New York: Routledge.

Harré, R. (1981). The positivist-empiricist approach and its alternative. In P. Reason & J. Rowan (Eds.), *Human inquiry: A sourcebook of new paradigm research* (pp. 3–17). New York: Wiley.

Kerlinger, F. N. (1979). *Behavioral research: A conceptual approach*. New York: Holt, Rinehart, & Winston.

Kerlinger, F. N., & Pedhazur, E. J. (1974). *Multiple regression in behavioral research.* New York: Holt, Rinehart & Winston.

Lamiell, J. T. (2013). On psychology's struggle for existence: Some reflections on Wundt's 1913 essay a century on. *Journal of Theoretical and Philosophical Psychology, 33,* 205–215. https://doi.org/10.1037/a0033460.

Lamiell, J. T. (2015). Statistical thinking in psychological research: In quest of clarity through historical inquiry and conceptual analysis. In J. Martin, J. Sugarman, & K. L. Slaney (Eds.), *The Wiley handbook of theoretical and philosophical psychology: Methods, approaches, and new directions for social sciences* (pp. 200–215). Hoboken, NJ: Wiley.

Lamiell, J. T. (2018). On scientism in psychology: Some observations of historical relevance. In E. E. Gantt & R. N. Williams (Eds.), *On hijacking science: Exploring the nature and consequences of overreach in psychology* (pp. 27–41). New York: Routledge.

Machado, A., & Silva, F. J. (2007). Toward a richer view of the scientific method: The role of conceptual analysis. *American Psychologist, 62,* 671–681. https://doi.org/10.1037/0003-066X.62.7.671.

Rucci, A. J., & Tweney, R. D. (1980). Analysis of variance and the "second discipline" of scientific psychology: A historical account. *Psychological Bulletin, 87,* 166–184.

Sokol, M. (Ed.). (1990). *Psychological testing and American society*. New Brunswick, NJ: Rutgers University Press.

Stern, W. (1911). *Die Differentielle Psychologie in ihren methodischen Grundlagen* [Methodological foundations of differential psychology]. Leipzig: Barth.

Wundt, W. (1909). Über reine und angewandte Psychologie [On pure and applied psychology]. *Psychologische Studien, 5,* 1–47.

Wundt, W. (2013). Psychology's struggle for existence (J. T. Lamiell, Trans.). *History of Psychology, 16,* 195–209. https://doi.org/10.1037/0032319.

5

Statistical Thinking in Psychology: Some Needed Critical Perspective on What 'Everyone Knows'

Over the past four decades, I have engaged in countless discussions with many different psychologists, both orally and in print, concerning the distinction between aggregate-level knowledge and individual-level knowledge. Almost without exception, my interlocutors in those discussions have pronounced themselves fully cognizant of that distinction in their work, whether as research scientists or as practitioners/consultants. 'Everyone knows,' I have often been reproved, not infrequently with eyes rolling, 'that a statistical index used to represent a group as a whole cannot be taken to represent accurately every individual in the group.' Presumably, it is the combination of widespread understanding of and scrupulous respect for this basic truth that makes further discussion of it 'logically trivial,' as Banicki (2018, p. 269) dismissively branded one of my recent efforts in this direction (cf. Lamiell, 2018a).

Closer and more circumspect consideration, however, reveals a deeper conceptual reality beneath the above-stated maxim that is decidedly more limiting of and problematic for mainstream thinking than is commonly acknowledged. It is not simply that an aggregate index fails to represent

J. T. Lamiell, *Psychology's Misuse of Statistics and Persistent Dismissal of its Critics*, Palgrave Studies in the Theory and History of Psychology,
https://doi.org/10.1007/978-3-030-12131-0_5

every individual in the group; in fact, such an index cannot *properly* be taken to represent *any* individual in the group. This makes every such statistical index—and every collection of them—quite literally, and excepting only circumstances that are theoretically imaginable but never realized empirically, *knowledge of no one.*

The reason that this conceptual reality is problematic for psychology is obvious: a scientific discipline devoted to the production of statistical knowledge of no one is by its very nature a science of no one, and hence a discipline that could not possibly be suited to advancing our scientific grasp of the psychological doings of actual *some ones*.

On the basis of these considerations, I have argued that, henceforth, most of what is currently viewed as 'psychological' research should be recognized and referred to as *psycho-demography* (Lamiell, 2018b). With this term, I have sought to capture the reality that although investigators' professed substantive theoretical interests often do fall within the domain of the psychological, the knowledge that is actually being generated through most of the empirical research is, as a result of its aggregate, statistical nature, fundamentally demographic. It is not knowledge about individuals at all, but is instead knowledge about recognized and/or experimentally created *populations*, and the production of knowledge about populations is the essential business of demography (cf. Proctor & Xiong, 2018; Lamiell, 2018b, 2018c).

By no means is this argument intended to diminish demography. On the contrary, there are, unquestionably, many contexts in which demographic knowledge—including psycho-demographic knowledge—can be extremely informative and practically useful, a point to be discussed further in Chapter 7. However, demography is not psychology—not even when it is psycho-demography—and to the extent that psycho-demography is permitted to masquerade as psychology, it is actually *psychology* that is being diminished, as claims to scientific knowledge about individuals are being advanced that are at best unjustified and not infrequently misleading.

My purpose in the present chapter is to critically examine certain widely shared misunderstandings of aggregate statistical knowledge that *seem* to justify, and hence function to perpetuate, the field's unfounded knowledge claims. Always involved in those misunderstandings is some variant of the notion that aggregate statistical knowledge actually does, in one way or another, advance our scientific grasp of individual-level psychological doings. In the discussion to follow, it will be argued that this notion is false; that, in fact, statistical knowledge about aggregates is *never* correctly interpretable as knowledge about individuals.

Before proceeding further, two points bear explicit emphasis.

First, it is fully recognized that statistical knowledge about individuals is possible. Statistical knowledge is, by its very nature, constituted of a multiplicity of observations, and when an investigator is working with a multiplicity of observations all of which have been made of the *same* individual, then the knowledge generated through a statistical analysis of those observations is, obviously, knowledge of the individual in question. This is not the case, however, when the body of data aggregated for statistical analysis is constituted of numerous isolated observations of *different* individuals, and it is this widespread practice within mainstream psychology that results in what I term psycho-demography.

Secondly, it is also fully recognized that psycho-demographic knowledge can be a valuable source of hypotheses relevant to some or another aspect of individual psychological doings; hypotheses that might, in turn, be tested by methods that actually are suited to the task of gaining knowledge about individuals. Psycho-demography can thus be understood as a kind of pre-psychology, and in that sense can serve a very useful function. This point, too, will be discussed further in Chapter 7.

However, in the intellectual patrimony of the nineteenth-century Leipzig scholar Moritz Wilhelm Drobisch (1802–1896), whom I quoted in Chapter 1, I will argue that it is only the widespread and persistent indulgence within mainstream psychology of a 'Great Failure of Understanding' that statistical knowledge of aggregates of individuals can also properly be interpreted, *as it stands*, as knowledge of some aspect of the psychological functioning of the individuals within those aggregates. This notion is false, and hence scientifically untenable.

An Illustration of the Opacity of Aggregate Statistics with Respect to Individual-Level Phenomena: The Case of Inter-trait Correlations

In Chapter 2, mention was made of the claim by E. L. Thorndike (1874–1949) that the correlation between variables measuring between-person differences along two personality trait dimensions indicates 'the extent to which the amount of one trait possessed *by an individual* is bound up with the amount *he* possesses of some other trait' (Thorndike, 1911, p. 22, emphasis added). A critical analysis of this passage is especially useful for our present purposes because it advances with an explicitness rarely found in the archival psychological literature the view that a statistical index that is literally defined only for an aggregate of individuals, i.e., the correlation between two variables marking individual differences, can properly be interpreted as an empirical indicator of an individual-level reality, namely, the degree of correspondence between the respective levels of the two variables in some given individual.

Against the urgings of differential psychology's founding father, William Stern (refer to Chapter 3), the understanding of inter-trait correlations adopted by Thorndike became the accepted one within the discipline. The person-situation debate that raged within personality psychology throughout the 1970s and 1980s, for example (refer to Chapter 2), could not have arisen otherwise, and although that debate has long-since subsided, the understanding among contemporary mainstream psychologists of permissible interpretations of inter-trait correlations remains aligned with the view that Thorndike espoused over a century ago (see, e.g., Banicki, 2018; Hofstee, 2007; McCrae, 2011; cf. Lamiell, 2018a).

By examining Thorndike's claim closely, its mistakenness can easily be seen. Doing so thus serves us well as an object lesson in the opacity of aggregate statistics with respect to individual level phenomena. The specific phenomenon circumscribed by Thorndike's claim, correspondence in the respective levels of two traits in individual personalities, is, in the broader view of things, incidental.[1]

Looking More Closely

It is helpful to begin with a deliberate consideration of how the value of a (Pearson product-moment) correlation between two individual differences variables is determined. The first step is to calculate the respective means and standard deviations of the two variables. Each mean is defined by summing the 'raw score' observations across the N individuals investigated, and then dividing each sum by N. Each standard deviation is then defined as the square root of the value obtained by dividing by N the sum of the squared deviations of the individual 'raw' scores around their respective means. The standard score (z-score) for each individual on each variable is then defined as his/her 'raw' score minus the group mean, with that quantity then divided by the standard deviation for the variable in question. The sum of the cross-products of the subjects' respective standard scores on the two variables, divided by the number of subjects (N), defines the (Pearson product-moment) correlation between the two variables.

This short review of basic computational procedures is offered to underscore the simple but often overlooked fact that a correlation between two variables marking between-person differences is a group mean—literally, the *average* of the cross-products of standard scores on the two variables—defined for a *set* of N individuals considered as a whole. It is not a statistic that is defined for any one of the individuals within that set.[2]

Reviewing computational steps also serves to make clear why it is that in order for there to be any discussion at all, following Thorndike (1911) of what a correlation does and does not signify about the degree of correspondence between the 'amounts' of the respective variables present within individuals, those 'amounts' must themselves be defined in terms of the individuals' standard scores on the respective variables, because standard scores are the 'stuff' of which correlations are constituted.

With these rudimentary considerations in mind, the reader is now asked to refer back to Fig. 2.1 (p. 27). It will be recalled from our previous discussion of that figure that within each of its two panels, each plotted line represents one of 19 individuals, and connects the standard scores defined for that individual on a personality questionnaire administered on two different occasions. The Occasion 1–Occasion 2 data are displayed in the left-hand panel of the figure, while, for the same 19 individuals, the Occasion 2–Occasion 3 data are shown on the right. In order to coordinate this discussion substantively with the quote of Thorndike (1911) above, the reader may imagine that within each of the panels of Fig. 2.1, the correlation displayed is between assessments of two traits rather than between assessments of a single trait on two different occasions. This slight re-conception does not, of course, have any bearing on the essential nature and meaning of the data displayed in the figure.

Concentrating now on the left-hand panel of Fig. 2.1, we might reasonably characterize the correlation $r = +.60$ displayed there as indicating 'fairly strong' correspondence between the two variables. The question raised for us by the quotation of Thorndike (1911) is: what does this *aggregate* inter-variable correlation indicate about the level of correspondence between the respective standings of some given *individual* on those two variables? That is, to phrase the question in accordance with Thorndike's own wording: what does the correlation $r = +.60$ tell us about the extent to which the amount of one trait possessed by some given one of the 19 individuals investigated is 'bound up with' the amount that individual possesses of the other trait?

The most striking and salient feature of the data displayed in Fig. 2.1 is the unevenness of the slopes of the lines representing the different individuals investigated. Those represented by relatively flat lines are ones about whom we would say that there was high correspondence between the two assessments. Individuals represented by relatively steep lines are ones about whom we would say that the degree of correspondence between the two assessments was low. Now: these considerations just *mean* that the answer to the question of how closely the two trait assessments correspond at the level of the individual depends upon which particular individual is under discussion in any given instance. This in turn just *means* that knowledge of the aggregate level of correspondence between the two variables, in this particular case $r = +.60$, is *no guide at all* to answering the question for *any one* of those 19 individuals. So, we see that, in fact, with respect to the question of correspondence between trait assessments at the level of the individual, the aggregate correlation is effectively opaque, hence irrelevant.

Shifting our attention to the right-hand panel of Fig. 2.1, where the obtained correlation between the two sets of assessments was $r = -.01$, or essentially zero, the epistemic situation remains exactly the same. Once again, the most striking feature of the data displayed there is the unevenness of the slopes of the lines representing the various individuals. In some cases, the slopes are very steep, graphically depicting very low correspondence between the assessed levels of the two traits. In other cases, however, the slopes are virtually flat, depicting close to perfect correspondence. So here again we see that the answer to the question of how closely the two trait assessments correspond at the level of the individual depends upon which particular individual is under discussion in any given instance. No more here than before is knowledge of the aggregate level of correspondence between the two variables, in this case $r = -.01$, any guide at all to answering the correspondence question for any one of the 19 individuals.

The only instance in which knowledge of the aggregate level of correspondence between two trait variables would answer the question of degree of correspondence between the two assessments made of a given individual would be if the aggregate level of correspondence were perfect, i.e., if r were found to equal 1.00. In that case, correspondence could be known to be perfect for each and every one of the 19 individuals investigated because that is the only way that the aggregate level of correspondence could *be* perfect. Under *any* other value of r, which is to say under any value of r ever obtained in actual research, the epistemic situation would be identical to that just described for $r = +.60$ and $r = -.01$: opaqueness hence irrelevance.

A Cautionary Note

To some readers, it might seem at first blush as if the present exercise establishes as too extreme the earlier claim that an aggregate statistic constitutes knowledge of *no one*. After all, having interpreted the correlation $r = +.60$ as empirical evidence of a 'fairly high' degree of inter-trait correspondence, it was then possible to identify several individuals—5, 7, 4, 10, and 15, for example—as individuals for whom the degree of correspondence between the assessed levels of the two traits could also reasonably be judged to have been 'fairly high.' As is true in many contexts, however, first appearances can be misleading, and careful reflection is required.

The correlation coefficients displayed in Fig. 2.1 do not capture or convey any empirical reality whatsoever about any one of the 19 individuals included in the study from which those correlations were obtained. The simplest and most vivid way to illustrate this for oneself is to cover the data plots in the figure so that only the correlations shown at the top of each panel are visible, and then, for either one of the correlations—it does not matter which one—pose to oneself the question:

Knowing the value of that correlation, what can I claim to know about individual X (take your pick) among the 19 subjects included in the study that produced that correlation?

The reader will find it unnecessary to ponder this question long before realizing that the only correct answer is '*Nothing*,' and that this will remain the only correct answer no matter which of the 19 subjects is selected for consideration. As has been shown, this does not mean that no individual could be found, post hoc, for whom results could be said to match the interpretation previously given to the aggregate finding. But with respect to the epistemic issue of concern here, two points of crucial importance must be noted.

First, it must be kept in mind that follow-up, individual-level analyses would be *necessary* in order to answer the question of interest at the level of the individual. This is *because* the aggregate correlation coefficients do not themselves provide the answer. They do not, in other words, function as any sort of 'window' onto individual-level results. As stated above, with respect to individual-level concerns, the aggregate coefficients are opaque.

Second, if follow-up, individual-level analyses are possible, that *means* that some method entirely separates from the computation of the aggregate correlations was available from the start for addressing the question of interest at the level of the individual. In the above exercise, the case-by-case data plots displayed in Fig. 2.1 could have been constructed and examined for research purposes exactly as they were even if the aggregate, inter-trait correlations, demonstrably uninformative about individual-level doings anyway, had never been calculated.

In short, neither did the aggregate correlations answer the individual-level questions, nor was their computation a necessary step toward answering those questions. It must therefore be wondered: Why would an investigator who happened to be interested, for whatever theoretical or practical reason, in the 'correspondence question' at the level of the individual bother with an aggregate-level statistical analysis in the first place?

In order to fully appreciate why aggregate-level statistical knowledge is unhelpful in the quest for knowledge about individual-level phenomena—and, again, the foregoing exercise is merely illustrative of this more comprehensive reality—it is important to clearly grasp that the crucial test question is not whether, by some means or other, one or more individuals can be identified *subsequently* for whom observations can be said to empirically match some inference that has previously been made about those individuals on the basis of aggregate statistical analyses. Rather, the crucial test question is:

Given the results of some aggregate statistical analysis, what can one claim to know about any one of the individuals investigated that was not known prior to that aggregate statistical analysis?

Consistent with the simple exercise just conducted, the conceptual reality here is that (excepting, again, a circumstance that is theoretically imaginable but never realized empirically) the answer to this test question is—always—*nothing*. It is this conceptual reality that has yet to be incorporated into what 'everyone knows' in mainstream psychology about the aggregate-individual distinction. That exclusion is the enduring problem, and it manifests itself in a variety of ways other than the specific one just discussed. One of those ways is reflected in the convictions mainstream investigators share about the serviceability of their statistical analyses with respect to the dual scientific functions of *predicting* and *explaining* individual behavior.

The Statistical Conception of Prediction and Explanation in Contemporary Mainstream Psychology

If there is a feature of psychological science to which the great majority of its practitioners has subscribed since the discipline's commonly acknowledged founding as a science, it is that research should serve the dual objectives of *predicting* and *explaining* various aspects of what I have referred to throughout this work as the psychological 'doings' of individuals—sensations, perceptions, judgments, memories, cognitions, emotions, and behaviors. This commitment has been especially prominent among those who have embraced the view of psychology as a *natural* science (*Naturwissenschaft*), the dominant view within the field since at least as far back as 1879. To be sure, it is a view that was contested from the start by scholars such as Wilhelm Dilthey (1833–1911) and Wilhelm Windelband (1848–1915), who believed that the field should be conceived at least partly, if not entirely, as a *human* science (*Geisteswissenschaft*; Dilthey, 1894; Windelband, 1894/1998). As such, psychology would be aimed not so much at predicting and explaining human doings, but instead at *understanding* them. There have been proponents of this view ever since, right up to and including the present (see, e.g., Schiff, 2017). Still, the natural science view is the one that has long dominated mainstream thinking (cf. Gantt & Williams, 2018), and it is the perspective on prediction and explanation taken by contemporary representatives of that view with which I am concerned in this discussion. A view more hospitable to the notion of psychology is, at least in part, a human science will be presented in Chapter 7.

In mainstream thinking, the objectives of predicting and explaining individual doings are regarded as fully complementary, and both are regarded as well-served by the kind of aggregate-level, statistical knowledge that can be secured through studies of variables marking between-person differences. It matters not whether those differences are of the sort that have arisen outside the laboratory and then 'captured' by tests, or, instead, of the sort that have been created inside the laboratory by the introduction of different treatments. In either case, or, as well, in 'hybrid' studies simultaneously examining variables representing both kinds of differences (refer to Chapter 4), it is understood that the statistical findings of the study can be formulated in terms of expressions having the general structure of Eq. (5.1):

$$Y_i = f_{y,x}(X_i) + e_i \tag{5.1}$$

where

Y_i represents the standing of individual i on the criterion/dependent variable, Y;

X_i represents the standing(s) of that same individual i on the predictor/independent variable(s) X[3];

$f_{y,x}$ represents the mathematical expression, typically a simple or multiple regression equation, describing the statistical relationship that has been discovered to exist, through the empirical research in question, between the criterion/dependent variable, Y, and the predictor/independent variable(s), X, and

e_i represents what is commonly referred to as 'error,' by which is meant the difference between individual i's actual standing on the criterion/dependent variable, Y_i, and the standing on that variable that would be estimated for him/her, symbolized Y_i', given knowledge of the obtained statistical relationship between Y and the predictor/independent variable(s), X. That is:

$$e_i = \left(Y_i - Y_i^{'}\right).$$

It is perhaps because of the form of expressions such as Eq. (5.1) that they can seem to convey knowledge of individuals. Here again, however, appearances can be misleading, and, just as we found previously to be the case in probing the validity of Thorndike's (1911) interpretation of inter-trait correlations, careful reflection is necessary.

At the *aggregate* level, which entails reference to *all* of the research subjects in a given investigation, considered as a unit, the predictive and explanatory power of an expression such as Eq. (5.1) is knowable empirically as the proportion of the total variance in the dependent/criterion variable, Y, that is attributable to the predictor/independent variable(s), X. That proportion is quantitatively specifiable as the square of the correlation, be it simple (r) or multiple (R), between Y and the one or more X-s.[4] In the hypothetical case where r^2 or R^2 would equal 1.00, the value of e_i in Eq. (5.1) would be zero for each and every individual i, rendering every individual's dependent/criterion variable standing fully predictable on the basis of his/her standing on the

X variable(s), and thus fully explainable in those same terms.[5] So, when a simple or multiple correlation between *Y* and the one or more *X*-s is perfect, an expression such as Eq. (5.1) conveys knowledge of *both* aggregate-level *and* individual-level order.

But of course, the statistical relationship between *Y* and *X* never is perfect, and although when it is less than perfect an expression like Eq. (5.1) still conveys knowledge of *aggregate*-level order, it no longer provides knowledge of individual-level order. What makes an imperfect statistical relationship between *Y* and *X* still informative at the aggregate level is the fact that the *average* value of the *e* term in Eq. (5.1) remains mathematically specifiable: it assumes a particular, albeit nonzero, quantity defined for the *set* of research subjects as a whole.[6] By the very same token, however, the reason that an imperfect statistical relationship between *Y* and *X* is no longer informative at the individual level is the fact that the specific value of e_i in individual cases is not specifiable empirically. It is *unknown*.[7] This just *means* that when the statistical relationship between *Y* and *X* is less than perfect, which it always is, the predictive—and hence explanatory—power of an expression such as Eq. (5.1) in individual cases is likewise *unknown*, and it is vital to understand that this is true not just for certain isolated individual cases, but for each and every one of the individual cases considered as such.

In fact, there is nothing in knowledge of a less-than-perfect statistical relationship between the *Y* and *X* components of an expression such as Eq. (5.1) that would preclude the possibility of the standing of any given individual *i* on the criterion/dependent variable *Y* being anywhere within the original range of admissible *Y* values. No greater precision than this in an investigator's knowledge about individual *i* could properly be claimed when the statistical relationship between *Y* and *X* is imperfect, and, of course, that much precision could be claimed before—indeed, whether or not—any statistical analysis of the relationship between *Y* and *X* was ever carried out (cf. Lamiell, 1991; Tryon, 1991a, 1991b).

It perhaps occurs to the reader at this point that these considerations completely undermine the 'each and every' assumption that, as explained in Chapter 4, has always been imposed, wittingly or otherwise, on the interpretation of the 'effects' uncovered by the statistical methods of treatment group experimentation in order to secure epistemic continuity with the earlier Leipzig model. We have just seen that, in practice, the findings of treatment group experimentation can *never* warrant a claim to predictive or explanatory knowledge of *any* particular individual, let alone a claim to such knowledge of 'each and every' one of them. In reality, the presumed epistemic continuity between treatment group experimentation and the earlier Leipzig model never existed. The epistemic gap between the two was complete and unbridgeable from the start.

The Inadequacy of Appeals to Probabilistic Thinking

Putatively, statistical studies of variables defined for populations can warrant *probabilistic* knowledge claims about the individuals within those populations, and in this way serve as a means of advancing our scientific grasp of individual-level doings (a very recent defense of this view can be found in Proctor and Xiong, 2018). This notion can be understood in terms of Eq. (5.1), introduced above.

It was explained earlier that for each individual in an investigator's sample of research participants, the e_i component of Eq. (5.1) is defined as (Y_i–Y_i'), i.e., the difference between individual *i*'s actual standing on the criterion/dependent variable, Y_i, and the best estimate of what that standing would be, Y_i', given knowledge of the obtained statistical relationship between Y and the predictor/independent variable(s), X. Now unless the statistical relationship between Y and X is perfect (in which case e_i will equal zero for every individual in the research sample), the values of e_i will vary across the individuals in the sample, with Y_i' in some instances overestimating and in other instances underestimating its corresponding Y_i.[8] This obtained variability of e_i values can serve as the empirical basis for specifying, at each value of the independent/predictor variable(s), X, an interval of admissible Y values, ranging from l (low) to h (high), that, given a sufficiently large sample of individuals, may be expected to contain some specified proportion, p, of all of the actual Y_i values at that value of X. By convention, p is usually set at .95, and this *l*-to-*h* interval is, in turn, commonly referred to as the '95% confidence interval.'

It is on this basis that an investigator thinking in accordance with prevailing mainstream beliefs and practices will claim the ability to know that, given some given individual's standing on *X*, the probability is .95 that his/her standing on *Y* is contained by the *l* and *h* poles of the confidence interval of *Y* values.

The problem with this line of reasoning is that, logically, the validity of probabilistic knowledge claims hinges on the consideration of a *series* of events. To know that the probability is .5 that flips of a fair coin will turn up heads is to know that over a sufficiently lengthy series of flips, half will turn up heads and half will not. The knowledge claim '$p = .5$' is tied inextricably to the consideration of the entire series of flips. The knowledge claim '$p = .5$' has no validity for any single flip within the series, as it is known from the start that the outcome of each and every such flip, considered by itself, either will or will not be heads. In the language of probabilities, that is, we must say that the probability of an isolated coin flip turning up heads is *always* either one or zero, and *never* .5 (cf. Venn, 1888).

These considerations reflect what is known as the 'frequentist' understanding of probability, and it is to that understanding that psychological researchers must appeal, wittingly or otherwise, in any instance in which they are claiming probabilistic *knowledge* about their research subjects.[9] The egregious error that so often infects such a knowledge claim lies in belief that it is valid not just for an entire sample of subjects considered as a whole, but also for the individual research subjects within that sample.

Having specified a 95% confidence interval in the manner described above, all that an investigator can *validly* claim to know probabilistically about his/her research subjects is that among *n* of those subjects with a given standing on the independent/predictor variable(s) *X* of Eq. (5.1), 95% of them will have values on dependent/criterion variable, *Y*, lying somewhere between the *l* and *h* poles of the confidence interval, while 5% of them will have *Y* values lying outside those poles. On the question of which of those two groups will be found to contain some specific individual person P, the data will be *completely silent*. A claim to *know* that the probability *is* .95 that *this* person's standing on Y will be found to lie somewhere between the l and h poles of the confidence interval is no more sensible than is a claim to *know* that the probability *is* .5 that *the* next flip of a coin will turn up heads. In each instance, the single occurrence being discussed has been extracted from the series of which it was a part, and when that is done, the probabilistic knowledge that has been gained about that series as a whole is rendered irrelevant.

So, yet again here, we find ourselves facing the reality that the empirical knowledge revealed by statistical studies of variables marking differences between individuals is not and cannot validly be made to be knowledge of individuals. This reality cannot be circumvented by phrasing knowledge claims about individuals in probabilistic language.

Some Additional Perspective

Another way to see this is to consider matters from the standpoint of the question: What empirical observation could challenge the validity of a knowledge claim of the sort 'The probability is p that some occurrence, O, will be found to obtain—i.e., transpire or be instantiated—in the case of individual i?' In science, it is widely agreed, claims to factual knowledge must, at least in principle, be susceptible to challenge or disconfirmation by further empirical observation. Claims that do not meet this criterion are called *incorrigible*.

Clearly, if p in the above-stated knowledge claim were stipulated as 1.0, amounting to a claim of certainty that O will obtain in the case of individual i, then the empirical finding that O has *not* obtained for that individual could overturn the claim. Similarly, if p were stipulated as zero, amounting to certainty that O will not obtain in the case of individual i, then the empirical finding that O *has* obtained for that individual could overturn the claim. However, 1.0 and 0.0 are the only two values of p under which the knowledge claim could be overturned by empirical observation. The finding that O *has* obtained for individual i cannot overturn a claim that the probability of that happening was 'low'—but not zero—nor can the finding that O has *not* obtained for individual i overturn a claim that the probability of that happening was 'high'—but not 1.0.

Probabilistic knowledge secured in studies of aggregates of individuals is not knowledge of individuals to begin with, and the inappropriateness of treating that knowledge as if it were knowledge of individuals is reflected in the fact that the resulting knowledge claims are immune from challenge by empirical observations of those individuals.[10] The claims are empirically incorrigible, and such incorrigibility makes for bad science.

Also problematic are claims to knowledge that some occurrence, O, is *X*-times more (or less) 'likely' among 'people' who have certain characteristics or who have experienced certain experimental treatments. For example, referring to the results of an experiment by Isen and Levin (1972), Banicki (2018) stated that those investigators found that 'people' who found a dime in a phone box were 'about 20 times more likely to help a stranger than those who had not been so lucky' (p. 259). The first question always raised by locutions of this sort is: What, exactly, is the intended meaning of the expression 'people'?

On its face, the statement by Banicki (2018) is purely demographic. It refers to two aggregates of 'people' who helped, one comprised of N individuals who had not found a dime in a phone box, and the other comprised of $N * 20$ individuals who had found a dime in a phone box.

Understood psychologically, however, the expression 'people' must be taken to mean not simply aggregates of people, but to mean *any given person* within one or another of those aggregates of 'people.' On this interpretation, Banicki's (2018) statement becomes a claim that the Isen and Levin (1972) experiment showed that *a person* who found a dime was 20 times more likely to help a stranger than *an other* person who did not find a dime. Clearly, however, this is not what the experiment showed.

The problem here is identical to the one discussed above in the context of confidence intervals and probabilities. On the question of which of two groups, the helpers or the non-helpers, would turn out to include 'this' particular person who found/did not find a dime in the phone box, the findings of the Isen and Levin (1972) experiment are silent, and this is true for each and every one of the participants in that experiment. The N-vs.-$N * 20$ outcome of the experiment, which was the empirical basis for Banicki's (2018) '20 times' knowledge claim, is an empirical fact that was defined for the totality of the 'people' who participated in the experiment, and is no valid basis for any claim to knowledge of the 'likelihood' that any one of the participants in the study would help or not help.

Probability on a Subjectivist Understanding

During a conference held in Boston, MA in October of 2015, I had occasion to be interviewed about the ideas discussed in this chapter (cf. Lamiell & Martin, 2017), and in that context was questioned by a conference attendee who was seeking further clarity. He began by proposing that we look at things from what he called 'a common sense point of view,' or what he might just as well have labeled the standpoint of 'what everyone knows.'

Suppose, my questioner suggested, that an observer finds, across numerous encounters with many different individuals over time, that some trait, T, often, even if not always, goes with some other characteristic, C, and that, as a result of these experiences, the observer thinks, in the next encounter with an individual who displays trait T, that it is likely that person will display characteristic C as well. My questioner went on to state explicitly—and altogether correctly—that this is just the sort of thinking practiced by the mainstream psychologists whom I am criticizing, and, as if speaking for those countless mainstream thinkers, he asked me to explain further just what I find wrong with such thinking.

My response to such questioning ran (and continues to run) as follows: In and of itself, there is nothing wrong with the sort of thinking described by my questioner. Actually, however, it is not the sort of thinking I criticize. To see both that and why this is so, one must consider carefully the wording that my questioner chose in stating his concern. He described a scenario in which an observer's experiences with many people prior to encountering person P have inclined that observer *to think it likely* that person P will display some particular trait or characteristic, C. This scenario perfectly illustrates a long-recognized tradition in statistical thinking called *subjectivism*. In that tradition, probabilistic statements are made—and properly understood—not as *claims to knowledge*, the understanding of probability adopted in the *frequentist* tradition which has been the focus of the discussion throughout this chapter up to this point, but rather as *expressions of subjective belief*. The scenario described by my questioner culminated not in the observer's claim to *know* that the probability (likelihood) that person P will display characteristic C *is p*, but rather in a statement about the observer's *belief in the likelihood* that person P will display characteristic C.

There is nothing categorically 'wrong' with using the language of probability to express statements of subjective belief, and if mainstream psychologists were forthright about labeling as expressions of subjective belief their probabilistic statements about individuals based on aggregate-level statistical analyses, the problems that have been identified and discussed earlier in this chapter could be avoided. Such forthrightness would entail a clear understanding and full acceptance of the reality that (to return to the hypothetical scenario introduced above) (a) the probability that person P will display characteristic C is *always* 1.0 or zero—the characteristic either will or will not be displayed; (b) this is true *no matter what* the frequency distribution is in the data that an observer has accumulated through his/her prior experiences; (c) *no matter what* that observer's subjective belief about person P happens to be, and, for that matter; and (d) *regardless* of whether or not that belief is based on those prior experiences. Just these crucial points, however, are the ones that get lost when psychologists 'blur' the distinction between frequentist and subjectivist thinking about probability in the fashion stated by Cowles (1989) and discussed in Chapter 1.

Mainstream psychologists have always been determined to distance themselves as much as possible from any hint of subjectivity in their scientific pronouncements. This is a major reason for their wholesale rejection of the Bayesian framework for understanding probability (cf. Gigerenzer, 1987, 2004; Papineau, 2018; van Zyl, 2018), and it is vividly reflected in the tradition of expressing probabilistic statements about individual research subjects not as statements of subjective belief, on the order of 'My/our research findings incline me/us to think it (un) likely that person P is/has/did/will do X,' but rather, and quite invalidly, as claims to knowledge, on the order of 'My/our research findings establish that 'the probability *is p* that person P is/has/did/will do X.' The findings of studies investigating statistical relationships between variables marking differences between individuals can never justify knowledge claims of this sort.

Following up his initial question to me at the aforementioned Boston conference, my interlocutor asked me to consider another example with a slightly different twist. Suppose, he requested, that a medical doctor had found over time that among patients for whom she had been prescribing penicillin as treatment for certain symptoms, 85% improved. Would that doctor not be well-advised, my interlocutor asked, to continue prescribing penicillin for patients exhibiting those same symptoms? Moreover, he asked, might not any given patient with knowledge of his doctor's outcome findings prefer to be given the penicillin rather than not?

I began my response by agreeing that, indeed, the doctor might be well-advised to continue prescribing penicillin for her patients, believing on the basis of the already accumulated data that, going forward, the improvement rate among her patients would continue to be about 85%. But I emphasized that the established 85% improvement rate was an empirical fact about a *population*, and that, contrary to widely prevailing belief, could not justify any claim by the doctor to know that *this* patient standing in front of her today 'has an 85 percent chance' of improving. *This* patient's chances of improving are one or zero, and this is true *no matter what* the improvement rate within the population had thus far been found to be, and *no matter what* the doctor subjectively believed would happen in this or any other individual case.

I further emphasized to my interlocutor the importance of making the truth of this latter point clear to the patient faced with deciding whether to take the penicillin or not. That is, the patient must be counseled to understand that it has by no means been scientifically established that *his* chances of getting better by taking the penicillin are 85%. They are one or zero, and if, knowing that, the patient still feels, subjectively, that it would be better for him to take than to not take the penicillin, then so be it. But whatever choice the patient makes, the probability of improvement in his particular case remains one or zero.

In 1865, the French physiologist Claude Bernard (1813–1878) made a point in discussing statistical knowledge of direct relevance to the dialog recounted above. Bernard wrote:

> Statistics can allow (the doctor) to tell (her patient) that, of every hundred such cases, eighty are cured … but that will scarcely move him. What he wants to know is whether he is numbered among those who are cured. (Bernard, 1865, as quoted in translation from the French by Porter, 1986, p. 160)

What Bernard clearly understood over 150 years ago is that although population-level statistics can inform *subjective judgments or beliefs* about individual cases, they do not and cannot supply objective *knowledge* about what was, is, or will be true about individual cases. Sooner or later, mainstream thinking in psychology will have to align itself with this reality. When that happens, the domain of 'what everyone knows' about the nature of aggregate statistics in the social sciences will have been significantly expanded and, with that, the thinking of mainstream psychologists on the subject will be significantly improved.

Notes

1. Commenting on a recent article in which I discussed Thorndike's (1911) claim (Lamiell, 2018b), Proctor and Xiong (2018) have argued that I failed to grasp the probabilistic nature of the meaning that Thorndike 'intended to convey' (Proctor & Xiong, 2018, p. 484). The inadequacy of psychologists' appeals to probabilistic thinking will be discussed later in this chapter.
2. Lest this point be seen as overly pedantic, I would entreat the reader to recall the anecdote discussed in Chapter 2, involving an interchange between myself and a senior and internationally respected personality investigator who had transgressed the conceptual boundary circumscribed by this very point during a conference presentation in which he was discussing his own correlational research findings. More recently, the authors of an article published in the *Journal of Experimental Psychology* concluded their research report with the observation: 'Most research on lexical-semantic processing has examined group-level data. The present findings suggest that *additional* insights can be gleaned from individual differences analyses' (Pexman & Yap, 2018, p. 1105, emphasis added). This wording clearly reflects the authors' mistaken

belief that individual differences analyses entail something *other* than the examination of group-level data. The confusion here within the mainstream of psychology is widespread and of long standing.

3. Note the proviso here that the symbol X in Eq. (5.1) can stand either for a single independent/predictor variable or for a combination of several such variables.
4. The reader should bear firmly in mind that expressing the statistical findings of a research study in terms of a simple or multiple correlation does not somehow make a treatment group experiment into a correlational investigation. If a study qualifies as a treatment group experiment by virtue of its design features (random assignment of subjects to different treatment conditions), it remains a true experiment even if its statistical findings are expressed in the terminology and symbols commonly employed in correlation/regression analyses (cf. Kerlinger & Pedhazur, 1974).
5. This point reflects the full complementarity of prediction and explanation within the canon of research methods in contemporary mainstream psychology.
6. Note that under these circumstances, the set of e-values in a given study will also have a defined variance, and the ratio of that variance to the total variance of Y will be the proportion of the total Y variance left *un*accounted for by the X variable(s), i.e., $1 - r^2$ or R^2.
7. Were matters otherwise, the sensible researcher would use the putatively known value of e_i to correct its corresponding Y_i', and thus insure that the corrected Y_i' values would always align perfectly with their corresponding Y_i values. Errors of estimation/prediction would thus completely disappear!
8. The sum of the squared e_i values will be less than it would be under any scheme for generating the Y_i' values other than that specified by the $f_{y,x}$ component of Eq. (5.1). This is the basis for regarding that particular set of Y_i' values as the 'best' estimates of the actual Y_i values.
9. Alongside the frequentist understanding of probability, scholars have also long recognized a *subjectivist* understanding according to which probabilistic statements are made not as claims to knowledge but rather as *expressions of subjective belief*. This understanding of probability will be discussed further below.
10. To underscore a point stated earlier, this does not mean that probabilistic knowledge of individuals cannot be obtained. If the multiplicity of

empirical readings required to formulate probabilistic knowledge is secured through repeated observations of the same individual, then meaningful claims to probabilistic knowledge about that individual can be made, and those claims can, in turn, be subjected to further empirical test through the accumulation of more such observations about that individual. However, this is not what is done in studies of variables defined for aggregates of individuals, and it is specious to suppose that research of this latter sort can be regarded as an adequate substitute for research of the former sort.

References

Banicki, K. (2018). Psychology, conceptual confusion, and disquieting situationism: Response to Lamiell. *Theory and Psychology, 28,* 255–260. https://doi.org/10.1177/0959354318759609.

Cowles, M. (1989). *Statistics in psychology: An historical perspective*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Dilthey, W. (1894). Ideen über eine beschreibende und zergliedernde Psychologie [Ideas concerning a descriptive and an analytical psychology]. *Sitzungsberichte der Akademie der Wissenschaften zu Berlin*, Zweiter Halbband, 1309–1407.

Gantt, E. E., & Williams, R. N. (Eds.). (2018). *On hijacking science: Exploring the nature and consequences of overreach in psychology*. New York: Routledge.

Gigerenzer, G. (1987). Probabilistic thinking and the fight against subjectivity. In G. Gigerenzer, L. Krueger, & M. S. Morgan (Eds.), *The probabilistic revolution: Ideas in the sciences* (Vol. 2, pp. 11–33). Cambridge, MA: MIT Press.

Gigerenzer, G. (2004). Mindless statistics. *The Journal of Socio-Economics, 33,* 587–606.

Hofstee, W. K. B. (2007). Unbehagen in individual differences: A review. *Journal of Individual Differences, 28,* 252–253. https://doi.org/10.1027/1614-0001.28.4.252.

Isen, A. M., & Levin, P. F. (1972). Effect of feeling good on helping: Cookies and kindness. *Journal of Personality and Social Psychology, 21,* 384–388.

Kerlinger, F. N., & Pedhazur, E. J. (1974). *Multiple regression in behavioral research*. New York: Holt, Rinehart, & Winston.

Lamiell, J. T. (1991). Problems with the notion of uncertainty reduction as valid explanation. *Journal of Theoretical and Philosophical Psychology, 11,* 99–105. https://doi.org/10.1037/h0091520.

Lamiell, J. T. (2018a). On the concepts of character and personality: Correctly interpreting the statistical evidence putatively relevant to the disposition-situation debate. *Theory and Psychology, 28,* 249–254. https://doi.org/10.1177/095935431774837.

Lamiell, J. T. (2018b). From psychology to psycho-demography: How the adoption of population-level statistical methods transformed psychological science. *American Journal of Psychology, 131,* 471–475.

Lamiell, J. T. (2018c). Rejoinder to Proctor and Xiong. *American Journal of Psychology, 131,* 489–492.

Lamiell, J. T., & Martin, J. (2017). The incorrigible science: A conversation with James Lamiell. In H. Macdonald, D. Goodman, & B. Becker (Eds.), *Dialogues at the edge of American psychological discourse: Critical and theoretical perspectives* (pp. 211–244). London: Palgrave Macmillan.

McCrae, R. R. (2011). Facts and interpretations of personality trait stability: A reply to Quackenbush. *Theory and Psychology, 11,* 837–844. https://doi.org/10.1177/0959354301116009.

Papineau, D. (2018, June 18). Thomas Bayes and the crisis in science. *The Times Literary Supplement*.

Pexman, P. M., & Yap, M. Y. (2018). Individual differences in semantic processing: Insights from the Calgary semantic decision project. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 44,* 1091–1112. https://doi.org/doi.org/10.1037/xlm0000499.

Porter, T. M. (1986). *The rise of statistical thinking: 1820–1900*. Princeton, NJ: Princeton University Press.

Proctor, R. W., & Xiong, A. (2018). Adoption of population-level statistical methods did transform psychological science but for the better: Commentary on Lamiell (2018). *American Journal of Psychology, 131,* 483–487.

Schiff, B. (2017). *A new narrative for psychology*. Oxford: Oxford University Press.

Thorndike, E. L. (1911). *Individuality*. New York: Houghton-Mifflin.

Tryon, W. W. (1991a). Uncertainty reduction as valid explanation. *Journal of Theoretical and Philosophical Psychology, 11,* 91–98. https://doi.org/10.1037/h0091519.

Tryon, W. W. (1991b). Further support for uncertainty reduction as valid explanation. *Journal of Theoretical and Philosophical Psychology, 11,* 106–110. https://doi.org/10.1037/h0091508.

van Zyl, C. J. J. (2018). Frequentist and Bayesian inference: A conceptual primer. *New Ideas in Psychology, 51,* 44–49. https://doi.org/doi.org/10.1016/j.newideapsych.2018.06.004.

Venn, J. (1888). *The logic of chance*. London and New York: Macmillan.

Windelband, W. (1894/1998). History and natural science (J. T. Lamiell, Trans.). *Theory and Psychology, 8,* 6–22.

6

Statisticism in Psychology as a Socio-ethical Problem

The seemingly unshakable belief among contemporary mainstream psychologists that aggregate statistical methods of inquiry offer the best available means of advancing our scientific grasp of the psychological doings of individuals defines that paradigmatic malady I have elsewhere branded 'statisticism' (cf. Lamiell, 2013). This unfortunate –*ism* is reflected in the ascendant—and continuing—practice within the mainstream, noted by Danziger (1987) and discussed in Chapter 4 of this work, of assuming that the statistical structure in data sets comprised of the responses of many individuals reveals the structure of the relevant psychological processes operating in individuals. Routinely, researchers' aggregate-level findings of slight/moderate/strong statistical covariations between variables marking differences between individuals are discussed in publications, conference presentations, and informal conversations as empirical evidence of correspondingly slight/moderate/strong 'tendencies' that the investigated individuals 'have.'

Statisticism is epistemically problematic in that it blinds its carriers to the unbridgeable conceptual gap between the probabilistic knowledge claims about individuals that are commonly made within scientific psychology's mainstream, on the one hand, and the probabilistic

J. T. Lamiell, *Psychology's Misuse of Statistics and Persistent Dismissal of its Critics*, Palgrave Studies in the Theory and History of Psychology,
https://doi.org/10.1007/978-3-030-12131-0_6

knowledge claims about populations that the extant empirical evidence will actually support, on the other hand. In this connection, Banicki (2018) has explicitly posed the question: 'Is [this] conceptual gap really so unbridgeable?' (Banicki, 2018, p. 258). For reasons I hope to have made abundantly clear in Chapters 4 and 5, the unequivocal answer to this question is *yes*. The gap truly is completely and irremediably unbridgeable, just as Drobisch claimed it to be in 1867 (Porter, 1986).

It should be obvious that this problem is much more than the 'troublesome paradox' once noted by Kerlinger (1979, p. 275) but then ignored both by him and by mainstream researchers more broadly. It is a large and highly problematic epistemic issue warranting unambiguous acknowledgment throughout the field. After all, as prosecutors of a putatively scientific discipline, psychological researchers are presumably striving for knowledge claims about the psychological doings of individuals that are *valid*. Yet knowledge claims about the psychological doings of individuals backed only in the coin of statistical fact patterns established for populations are profoundly *in*valid. This is not because the claims can be known to be false for all of the individuals studied, for, as we saw in Chapter 5, that is not necessarily the case. It is, rather, because the claims cannot be known to be true for *any* of the individuals studied. In other words, the claims are invalid in the specific sense that they lack the scientific warrant that is commonly attributed to them.

As elaborated in Chapter 5, a discipline devoted to the creation and dissemination of knowledge of statistical relationships between variables in terms of which individuals have been differentiated—whether by tests or by experimental treatments or by some combination of the two—is, by its essential nature, a discipline devoted to the creation and dissemination of knowledge about no one. As has been emphasized, this does not mean that such knowledge is worthless. On the contrary, such knowledge can be highly valuable, both as a species of demography and as a kind of 'pre-psychology,' prompting hypotheses about the possible nature of specified psychological doings that might subsequently be tested using methods that are appropriate for securing knowledge about individuals.[1] However, when population-level statistical knowledge is presented and discussed *as if it already is* knowledge about the individuals within the investigated populations, the result leads, exactly as Bakan (1955) claimed, to error, and the systematic purveyance of error is just bad science.

Yet as large and serious as this epistemic problem is, it is not statisticism's only untoward consequence. Due to the fact that psychology is not only a basic science but also an applied discipline, the problem of statisticism has a socio-ethical facet as well, and it is on this facet of the problem that the discussion in the present chapter is focused. The discussion begins with a consideration of two specific examples illustrative of what is problematic in this domain. The first is drawn from the province of experimental psychology, specifically, the evidence-based practice movement in psychology. The other example is drawn from the province of correlational psychology, specifically, the use of psychological tests in the service of preemployment screening.

Following these discussions, the final portion of the chapter is devoted to broader considerations of a socio-ethical nature.

Specific Examples of Problematic Practices

Advocacy of Randomized Controlled Trials as the 'Gold Standard' for Evidence-Based Practice in Psychology

Paralleling the groundswell of support within the medical community for so-called 'evidence-based practices,' i.e., for having decisions about health care interventions guided by research findings documenting the relative effectiveness—or, as the case may be, relative ineffectiveness—of different treatments, there has likewise developed within the community of scientific psychologists a sizable movement in favor of evidence-based practice in psychology (EBPP). Although the report of a 2006 American Psychological Association (APA), Presidential Task Force emphasized the importance of maintaining a broad view on the question of what kinds of research evidence can and ought to be considered in attempts to determine just what the best practices are for remediating various particular difficulties, there is a broad consensus within the mainstream community that randomized controlled trials (RCTs) are the scientific 'gold standard' in this domain (see, e.g., Kazdin, 2008; Lilienfeld, Ritschel, Lynn, Cautin, & Latzman 2013).

Viewed from a methodological standpoint, a randomized controlled trial is just a special case of treatment group experimentation. In the conduct of an RCT, participants are drawn from the population of individuals suffering from the difficulty for which some given treatment is being investigated for its possible remedial effectiveness. The participants are then assigned at random to one of two experimental conditions, typically a treatment condition or a no treatment control condition,[2] with the two conditions defining the investigation's independent variable, and the trial begins. At the conclusion of the trial, the respective averages of the experimental conditions on designated outcome (dependent) variables are compared statistically in the search for any significant difference(s) that might be identified. All of this is formally identical to what has been discussed in previous chapters as treatment group experimentation, with 'treatment' in this case referring specifically to some form of intervention intended to be effective in remediating some particular psychological difficulty. The overarching rationale here is that treatments leading to outcomes that are statistically superior to non-treatment, or to some alternative treatment(s), should be prioritized by clinicians when deciding which treatment possibility to pursue in the case of a given client.

As things have developed, the EBPP movement has met with considerable resistance in the clinical psychology community, to the point where some advocates of the movement have attempted to soften that resistance by identifying and addressing its 'root causes' (Lilienfeld et al., 2013, p. 883). Among those root causes, Lilienfeld et al. (2013) specifically discussed the doubts harbored by many practicing clinicians that probabilities extracted from aggregate-level data analyses can be applied meaningfully to their respective individual clients. As Lilienfeld et al. framed the issue:

> Many students and beginning clinicians presume *erroneously* that group probabilities, which are all that RCTs can hope to deliver, cannot apply to the individual case. They assume that they cannot bridge the nomothetic and idiographic realms of analysis. Hence, they may conclude that there is

> no reason to rely on EBP, because 'every individual is unique.' Of course, there is a kernel (of) truth in this assertion: Each individual is indeed unique. Yet this undeniable fact does not imply that one cannot deduce probabilistic generalizations from controlled group studies that apply to individual clients, because groups are, after all, composed of individuals. (Lilienfeld et al., 2013, p. 891, emphasis and parentheses added)

The first epistemic problem that one encounters in this passage is the manifest belief on the part of its authors that RCTs produce 'nomothetic' knowledge. Indeed, in the paragraph immediately prior to the one just quoted, the authors stated that 'EBP relies primarily on nomothetic findings, which strive to extract universal or quasi-universal laws that apply to all or most individuals within the population' (Lilienfeld et al., 2013, p. 891).

However, and as was discussed at length in Chapter 4, such lawfulness as treatment group experimentation is in principle capable of revealing is *not* of a sort that can validly be claimed to apply 'to all or most individuals' within the population that has been sampled in the conduct of the investigation. No more than brief reflection is required to discern the logical impossibility of adducing empirical evidence of lawfulness applying 'to all or most individuals' without studying individuals, case by case. Such evidence would have to warrant a claim of the sort 'yes, the putative lawfulness applies in this case; yes, it applies in this case, too; yes, it applies again in this case; oops, it appears not to apply in this case,' and so on. Unarguably, treatment group experimentation—including but not limited to RCTs—does *not* generate evidence of this sort.

As is true of all treatment group experimentation, the *findings* of RCTs are defined by the outcome of statistical comparisons of the treatment condition *averages*. Such findings are not even defined for individuals, and this is what makes it impossible to claim validly that those findings 'apply to all or most individuals.' Moreover, it is for just this reason that such findings do not and cannot qualify as *nomothetic* knowledge about the psychological doings of individuals in the sense of 'nomothetic' intended by the scholar who minted the term, Wilhelm Windelband (1848–1915; cf. Windelband, 1894/1998). Indeed, to anyone who has ever read the relevant treatise by Windelband (1894/1998) and then examined the literature that has accumulated over the years bearing on the nomothetic–idiographic distinction, it is clear that very little of that literature has been authored by people who have likewise read and understood Windelband (1894/1998).

Windelband coined the term 'nomothetic' to refer to 'general' lawfulness in that sense of 'general' meaning *common to all*, i.e., true of *each* of the many instances investigated. He did not mean 'general' in the radically different sense of 'true on average' *across* the many instances investigated (refer to discussion in Chapter 1; see also Lamiell, 1998). Yet, as Lilienfeld et al. (2013) themselves noted, it is only the latter sense of 'general' that can possibly apply to the findings of RCTs, and it is for just this reason that those findings cannot properly be regarded as 'nomothetic' in the sense of *nomothetic* meant by Windelband (1894/1998).

All of that said, what is of more immediate concern in the present context is the contention by Lilienfeld et al. (2013) that students and beginning clinicians resisting the EBPP movement are 'erroneous' in their belief that the probabilistic knowledge issuing from RCTs does not apply to their respective individual clients. The specific reference here to students and beginning clinicians seems to imply that this putative error is much less prevalent among more experienced clinicians. If this is true, it only reflects tellingly the force of mainstream thinking as an instrument of paradigmatic indoctrination, because for reasons explained in Chapter 5, the beliefs of the students and beginning clinicians in this regard are entirely *correct*. It is Lilienfeld et al. (2013) (and, as those authors imply, more experienced clinicians as well) who are in error. This point can perhaps best be developed with reference to the examples that Lilienfeld et al. (2013) themselves used to prosecute their argument.

Those authors first appealed to an example they attribute to Meehl (1973), one in which an individual is forced to play the game of Russian roulette and is offered two options. Lilienfeld et al. (2013) continued:

> In one condition, the barrel of the gun contains four bullets, with one canister left blank; in the other, the barrel of the gun contains only one bullet, with four canisters left blank. If the player followed the

> rationale that 'probabilities don't apply to the individual case' to its logical (or in this case, illogical) conclusion, the choice of the condition would not matter, as the player would be equally likely to live or die regardless of her choice. … Yet this reasoning is obviously fallacious, as her odds of dying are four times higher with the first gun than with the second. (Lilienfeld et al., 2013, p. 891, parentheses in original)

Lilienfeld et al. (2013) then continued with an example of their own:

> Similarly, imagine a patient who has recently experienced a severe myocardial infarction. His physician presents him with two treatment options associated with identical side effect profiles: one that has been found in controlled studies to be associated with an 80% survival rate, and another that has been found to be associated with a 50% survival rate. Again, the logic that group probabilities are irrelevant to the individual would imply incorrectly that he has no legitimate grounds for selecting the former treatment over the latter. (Lilienfeld, et al. 2013, p. 891)

In both of these examples, the one borrowed from Meehl (1973) as well as the one of their own design, Lilienfeld et al. (2013) err in exactly the way discussed in Chapter 5, i.e., by extracting one single instance from the *series* of instances to which the probabilistic knowledge is inextricably bound (Venn, 1888). In the first scenario, the claim of Lilienfeld et al. (2013) that 'the odds of [the player] dying are four times higher with the first gun than with the second' is simply wrong. The odds of the player dying *in that one game* of Russian roulette are zero or one, and this is true no matter which gun is selected.

It is true that if the 'player' *(sic!)* were to participate in, say, 100 games of Russian roulette, about 80 (or 20, depending upon the gun selected) would end in no shot being fired. But there will not be 100 games. There will be only one game, and when the player of that one game squeezes the gun's trigger, a bullet either will or will not be discharged. In that single instance, the concept of a '20% chance' or '80% chance' simply does not apply, and that just means that the claim by Lilienfeld et al. (2013) that the player's odds of shooting herself are 'four times higher with the first gun than with the second' is false.

Likewise in the second example: were the patient to undergo, say, 100 courses of a given therapeutic treatment for his difficulty, the (hypothetical) data indicate that, depending upon the therapeutic course selected, about 80 (or 50) of those courses will result in survival, and 20 (or 50) will not. But the patient is not going to undergo 100 courses of treatment. There will be only one course of treatment. That single course of treatment either will or will not result in survival, and that is true no matter which treatment is selected. Again in this context, the concept of an '80% chance' or '50% chance' is inapplicable.

What allows Lilienfeld et al. (2013) to seem to win their argument that probabilistic knowledge gleaned from group studies is relevant to individual cases is that those authors have subtly changed the question from 'What is the factual probability (likelihood) of the desired outcome in this one instance?' to 'In which choice would the actor (Russian roulette player or medical patient) have a greater subjective sense of confidence in obtaining the preferred outcome?'

The first of these two questions asks for *objective knowledge*, and the only correct answer is 'zero or one' regardless of the choice made. To reiterate: the reason for this is that probabilistic knowledge is always and of its very essence factual knowledge about a *series* of instances considered as a single entity. It cannot properly be regarded as knowledge about any single instance within that series (Venn, 1888). At the level of the single instance, the 'probability' (if that is the term to be used) is always zero or one, because what we *know* is that the outcome, *X*, either will or will not occur.

The second of the two questions posed above asks about *subjective belief*. Assuming a desire on the part of both the Russian roulette player and the medical patient to survive, we may suppose that the Russian roulette player's subjective sense of confidence will be greater with the gun having four empty canisters, and that the medical client's subjective sense of confidence will be greater with the treatment associated with an 80% survival rate. The apparent rationality of these *beliefs* does not alter what aggregate statistical knowledge can entitle us to claim *know* in advance about the 'probability' that the result of one's choice will be the desired one.

Undoubtedly, a person's knowledge of aggregate statistics often influences the decisions and choices that s/he makes. This is a psychological phenomenon worth investigating and reflecting upon in its own right, as Daniel Kahneman (b. 1934) and Amos Tversky (1937–1996) so amply demonstrated (see, e.g., Kahneman & Tversky, 1982). But establishing that probabilistic knowledge influences individuals' subjective beliefs, choices, etc. ought not to be mistaken for establishing that knowledge of aggregate-level probabilities 'applies to'—i.e., entitles claims to objective probabilistic knowledge about—individual cases. This is the mistake embedded in the baldly patronizing attempt by Lilienfeld et al. (2013) to discredit the resistance of many 'students and beginning clinicians' to EBPP. Knowingly or otherwise, the students' and early career clinicians' resistance is grounded in the logically *correct* notion that some given therapeutic treatment *either will or will not* work with some particular one of his/her clients, and that this is true no matter what the aggregate statistical findings of RCTs have been found to be. Hence, that resistance is, in fact, fully justified, and the claim by Lilienfeld et al. (2013) to the contrary is entirely invalid.

Socio-ethical Implications

In the previously cited article by Kazdin (2008), which addresses many issues that have arisen in connection with the EBPP movement, he mentions concerns of practicing clinicians about efforts among third-party payers and states that would allow them to prescribe, presumably on the basis of the empirical findings of RCTs, those treatments that will be allowed and reimbursed. To the best of my knowledge, no such prescriptions are currently in place in the United States, but the very prospect of their being installed is reason enough to be concerned about them from a socio-ethical standpoint.[3]

Clearly, psychotherapy is aimed at remediating difficulties being experienced by individual clients. Prescriptions of the sort just mentioned would restrict a client's eligibility for reimbursement by insurance companies for his/her out-of-pocket expenses to those treatments that have been associated with statistically favorable outcomes in RCTs. As we have seen, however, the evidence secured through RCTs is not of a sort that can be known to be applicable to this or that or *any* individual case considered as such (refer to the argument developed at greater length in Chapter 5).

Advocates for restricting treatment choices to those that are associated with statistically favorable outcomes in RCTs (cf. Baker, McFall, & Shoham, 2008; Lilienfeld et al., 2013) might argue in terms of probabilities, but, as we have seen, this would require the adoption of a subjectivist understanding of probability according to which probabilistic statements about an individual client could express only the speaker's—and, by extension, the payer's—confidence that some given outcome will be achieved in the case of that client (refer to Chapter 5), and not factual knowledge about the likelihood that that outcome will be achieved. As has been emphasized repeatedly, that likelihood is always either zero or one in the individual case: the desired outcome either will or will not be achieved, and this is true no matter what the statistical findings of RCTs have revealed.

In the implementation of treatment prescriptions (and hence also of proscriptions) of the sort mentioned by Kazdin (2008), the alternatives made available to the individual client and his/her therapist would necessarily be limited not by objective knowledge known to be relevant to that individual client, but, instead, by the subjective belief(s) of one or more individuals tasked with that implementation. If practicing clinicians—whatever their level of experience—are inclined to resist such treatment prescriptions, this is just as it should be, and any movement on the part of mainstream psychologists to oppose the resisting clinicians in their efforts would be not only epistemically unjustified but, as well, socio-ethically questionable.

The Use of Psychological Tests as Instruments of Preemployment Screening

There are many societal contexts, e.g., school, business and industry, medicine, the military, in which important decisions affecting individuals' lives are made on the basis of the results of psychological tests (cf. Hanson, 1993). To illustrate the socio-ethical concerns to which statisticism can give rise in such contexts, attention is directed here to the use of such tests for guiding preemployment screening decisions, i.e., decisions about the suitability of job applicants for work of the sort they are seeking (e.g., Hogan, Hogan, & Roberts, 1996).

In an article that appeared in *The Wall Street Journal* dated September 29, 2014, it was stated that, as of then, tests of the personality characteristics, skills, cognitive abilities, and other presumed traits of job applicants were being administered to 60–70% of prospective workers in the United States. This percentage was up from the 30 to 40% it had been just five years previously. The article cited an estimate provided by Hogan Assessment Systems, Inc., of Tulsa, OK that such testing has become a $500 million per year industry, and is growing at a rate of 10–15% per year.

Interestingly, the President of Hogan Assessment Systems, Inc., Robert Hogan (b. 1937), co-authored an *American Psychologist* article published in 1996, which presented a forceful defense of testing in connection with pre–employment decisions. In that article, it was argued that the screening of job applicants by means of well-constructed and properly validated tests is both warranted scientifically and consistent with concerns for social justice (Hogan et al., 1996).

Illustrating the manner in which personality tests can be—and, in the view of Hogan et al., should be—put to use in preemployment screening, the authors cited empirical evidence indicating that 'truck driver performance is predicted by high scores for prudence and adjustment and low scores for sociability, because high sociability is associated with impulsivity, and impulsive truck drivers get in trouble on the job' (Hogan et al., 1996, p. 472).

It is thinking along the lines reflected in this passage that prompted the following hypothetical vignette, which thematically reprises one that I introduced some years ago (Lamiell, 2003). Let us imagine that, in full knowledge of the statistical relationship cited by Hogan et al. (1996), a psycho-technical expert retained (and doubtless well-compensated) by a trucking company has recommended to the chief personnel officer of that company that an applicant whom I will call here 'Lesley' not be considered further for employment as a driver. After learning of this recommendation, Lesley requests a meeting to discuss the reason(s) for that decision. The trucking company's personnel officer arranges for

Lesley to speak directly with the psycho-technical expert, and on the assumption that the latter is fully conversant with the epistemic limitations of the statistical considerations according to which the test administered to Lesley has been validated (an assumption that is far from always safe; cf. Hake, 2001; Valsiner, 1986), we may suppose that the conversation unfolds more or less as follows:

Lesley: I'm here to ask why I am not being considered further for a position as a driver with this company.
Psycho-technician: Well, Lesley, on the battery of personality tests that you took as part of the application process, you scored high on sociability.
Lesley: Yes, and?
Psycho-technician: Well, being highly sociable can often be a good thing, but in this instance, I'm afraid it is a liability. Scientific psychological research has shown that high sociability is statistically predictive of impulsivity, and impulsivity in truck drivers leads to trouble.
Lesley: I see. So, given my highly sociable nature, just how much trouble would I cause the company were I to be hired?
Psycho-technician: Well, it is not really possible to answer that question in your specific case, Lesley, but what we do know based on the scientific research is that, on average, individuals with the same score as yours on our measure of sociability end up causing more trouble in this profession than do their less sociable peers.
Lesley: But in my particular case, you really can't say what the outcome would be?
Psycho-technician: That is correct.
Lesley: So, allow me to see if I understand all of this properly. It has been recommended by you that I as an individual be considered no further for employment as a truck driver with this company, on the basis of a statistically-guided prediction about my future performance in the position, a prediction the accuracy of which in my individual case is completely unknown.
Psycho-technician: That is correct.

It is not inconceivable that Lesley's next comment would be something along the lines of 'I think I'll call my lawyer!' In any case, the question that this little vignette raises is: how, if at all, can professional practices that entail treating the Lesleys of the world in this way be justified?

Contrary to what the advocates of such practices (e.g., Hogan et al., 1996) believe,[4] and for reasons discussed at length in Chapter 5, there is no recourse here in the notion that well-executed scientific research can establish the validity of certain tests, or batteries of tests, for predicting an individual's job performance, so that, ultimately, it is the scrupulous adherence to accepted standards for conducting such research that can justify the practice just described. We have seen that, in fact, knowledge of the correlational validity of scores on some specified assessment instrument (or collection of instruments) for *generating predictions about*—as opposed to 'predicting'—individuals' scores on some specified criterion performance measure(s) provides no basis whatsoever for any claim to any knowledge of the test's predictive *accuracy* in the case of any one individual—neither Lesley nor anyone else. In correlational data of the sort on which claims concerning an instrument's predictive accuracy are commonly based, the accuracy of a prediction about an *individual's* criterion standing based on a score generated by that instrument is entirely *unknown*. Were it otherwise, the prediction made for each individual, respectively, would be adjusted in advance to correct for its putatively known degree of inaccuracy, and the result would be perfectly accurate predictions in every individual case! Clearly, this is not the way things are, have ever been, or could ever be.

Nor can the apologist for such practices find safe harbor in claiming probabilistic knowledge that some given individual's criterion score will fall within the upper and lower limits of some 'confidence interval.' As also explained in Chapter 5, the probability that some specified interval on the scale defining a criterion variable will contain some particular applicant's criterion score is always either one or zero: either it will or it will not.

To reiterate: all that a psycho-technician could *validly* claim to *know* is that Lesley's (or any given applicant's) actual criterion score would fall somewhere on the scale of admissible criterion score values, and this is true no matter what the predictive validity of the test has been found to be (so long as it is not perfect, which, of course, it never is). In other words, the level of predictive precision that a psycho-technician adhering to the accepted scientific standards for test validation can *legitimately* claim in any individual case is precisely equal to the level of predictive precision that could legitimately be claimed in that individual case even if the statistical analyses were never carried out!

To be sure, statistical evidence of the usual sort might be cited by a psycho-technician as the basis for his/her belief about what would happen if Lesley (or some other individual) were to be hired. Transparently, however, this would make the basis for recommending or not recommending Lesley a subjective opinion held by the 'expert' psycho-technician, and not some objective, scientifically secured truth about Lesley. Beyond the epistemic considerations here, the deeper socio-ethical concern is for the equitable treatment of the Lesley's of the world as individual persons, and it is arguable that a discipline that sanctions testing-based interventions of the sort just described is effectively blind to this concern.

In any and all investigations of variables marking differences between individuals, those individuals are being regarded, de facto and wittingly or otherwise, as instantiations of the categories that define the variables under investigation (male/female for the variable 'sex'; caucasian/non-caucasian for the variable 'race'; years since birth for the variable 'age'; IQ for the variable 'intelligence'; level of gregariousness for the variable 'sociability'; etc.). Under the logic of the statistical analyses conducted on empirical observations recorded in terms of the variables, each individual instantiation of a category (or specific combination of categories) is fully interchangeable for each other individual instantiation of the same category/category combination.

In applied contexts such as the one exemplified by the example of Lesley, each individual is handled in accordance with the dictates of an actuarial scheme (e.g., a regression equation) that, if followed strictly, will optimize certain payoff functions in the long run, i.e., on average, and hence serve the best long-term interests of the exercisers of the scheme. In the context of preemployment screening of job applicants, proceeding in accordance with this way of thinking means that decisions will be made and actions taken that, inevitably, will disserve some of the affected persons: some individuals will be barred from jobs in which they would have performed well, and, just as inevitably, other individuals will be hired, only to later experience failure. But in the mainstream ethos, it is accepted that concern for such inequities must be subordinated to the concern for the greater overall interests of the institution (be it an insurance company or other corporation, or a school, or a governmental agency) in the service of which the scheme is being exercised. In practices of this sort, individuals are treated as *commodities*, to be handled, deployed, and even manipulated in consideration of those other, super-ordinate interests (cf. Hanson, 1993).

Socio-ethical concerns in these contexts are only exacerbated by the claims issuing from the mainstream that probabilistic knowledge about populations can be understood as conveying probabilistic knowledge about the individuals within those populations. From that standpoint, statistically defined algorithms for decision-making followed in the best long-term interests of an institution can be made to seem to be also in the best interests of the individuals affected, and scientifically justifiable as such, even if the individuals in question would subjectively disagree. The highly sociable Lesley might not believe that, if hired as a truck driver, trouble caused by impulsivity would ensue, but scientific research seems to have established that that would 'probably' be the case, and so a recommendation by a psychometrician that Lesley not be hired would seem to well serve not only the best interests of the trucking company but also the laudable objective of saving Lesley from Lesley, by heading off professional disappointment, high bills for damages resulting from troubles caused, and/or, possibly, even death.[5]

Psycho-technical 'experts' might well acknowledge, if pressed on this point, that in the course of treating individuals in accordance with statistical schemes based on population-level studies, 'mistakes' will occur. But what statisticism prevents the 'experts' from seeing—or, as the case might be, from explicitly acknowledging—is the fact that such 'mistakes' are the *inevitable* result of practices that assume that statistical schemes generated from population studies can justly be applied to (i.e., imposed upon), in algorithmic fashion, each and every individual within the studied populations. Then, when the inevitable 'exceptions to the rule,' sometimes referred to as 'outliers,' manifest themselves, they are seen as indications that more regularities of the same basic—i.e., statistical—sort must be discovered and implemented accordingly. In this way, the paradigm perpetuates itself.

The nature of mainstream thinking about these matters is such that its advocates can readily concede that outcomes dictated by the population-level probabilistic schemes can be *mistaken* in certain individual cases without ever acknowledging that the very implementation of such schemes is *inappropriate* in individual cases. Seen through lenses fogged by statisticism, all of the mistakes built into such practices are 'honest' ones—regrettable, perhaps, as a kind of collateral damage, but finally eradicable only by further and more thorough applications of the same statistical methods. In the meantime, it is regarded not only as practically necessary but also as scientifically justifiable to subordinate the interests of adversely affected individuals to the larger, long-term interests of the institutions.

Broader Socio-cultural Considerations

Thinking about these matters in broader terms invites further critical reflection on the basic assumptions about the nature of human personhood that are built into a conceptual framework for psychological inquiry dominated by population-level statistical considerations. As has already been noted, any such framework demands that persons be regarded as instantiations of person categories, where each instantiation of a given category (which might itself be a combination of categories) is, in principle, substitutable by and for any other instantiation of the same category. Such a framework effectively reduces persons to things. Mainstream thinking in psychology finds scientific justification for such reduction in what is taken to be evidence of the lawfulness of individuals' respective psychological doings, i.e., the lawfulness putatively revealed by the statistical relationships found to exist between 'independent' (predictor) variables marking between-person differences along selected dimensions and the 'dependent' (criterion) variables reflecting between-person differences in the psychological 'doings' of interest (sensations, perceptions, judgments, emotions, cognitions, behaviors, etc.).

Lurking in the shadows of claims to knowledge of this sort of 'lawfulness' are intimations of causal determination. To be sure, researchers usually carefully avoid the word 'cause' and its cognates when discussing the findings of non-experimental studies, ever mindful of the textbook adage, drilled into every introductory psychology student, that 'correlation does not imply causation.' Curiously, however, the word 'effect' proliferates even in discussions of strictly correlational findings. Conference presentations and published research articles are peppered with references to statistical evidence of the 'effects' of differences in race, age, socio-economic factors, intelligence, personality variables, etc., on criterion measures, subtly exploiting the widespread understanding that where there are 'effects' there must be *causes*. Obviously, variables of the sort just mentioned (sex, age, race, etc.) do not represent different experimental treatments to which research subjects are randomly assigned. Nevertheless, the word 'effect' is often used in discussing statistically significant covariations involving such variables, and it is in this way that the notion of causal determination seeps into the discourse surrounding even non-experimental psychological research.[6] It is this discursive practice that sustains and broadens the illusion that research productive of statistical knowledge about average psychological doings within specified person categories can and does serve the quest for scientific knowledge of the causal determinants of the psychological doings of the individuals whom researchers have chosen to regard in terms of those categories.

Among its many other untoward consequences, the paradigmatic requirement that persons be regarded simply as instantiations of person-categories effectively eliminates from mainstream discourse an entirely different conception of persons as unique and inherently valuable beings whose behaviors and other psychological 'doings' are *self*-determined and hence not adequately accounted for in terms of the alleged (though ever inscrutable) causal powers of person-categories. Such an alternative view of persons has been proffered by many critics of mainstream thinking over the years—including but by no means limited to William Stern (1871–1938; cf. Stern, 1917/2010), Gordon Allport (1897–1967; cf. Allport, 1968), and Joseph F. Rychlak (1928–2013; cf. Rychlak, 1988), to cite examples from three different historical eras in our discipline. One of the major concerns shared by those thinkers was for the ethos that is created and maintained by the mechanistic conception of persons' psychological doings that the mainstream view must necessarily foster, though, often, only implicitly, uncritically, and even unwittingly (cf. Wendt & Slife, 2007).

Since very few of Stern's works have ever been published in English translation, I highlight here some observations reflecting his mindfulness of the untoward consequences of a disciplinary ethos dominated by a mechanistic—and hence impersonal—conception of human nature. Against such a view, Stern repeatedly defended over the course of his highly productive scholarly career a worldview (*Weltanschauung*) that he called 'critical personalism,' the foundation of which was the irreducible distinction between persons and things. In a lecture he gave at the Seventh International Conference for Psychotechnics, held in Moscow in September, 1931,[7] Stern voiced his concern that, in their research-based interventions in the lives of individuals, psychologists were not sufficiently mindful of this distinction. The passage below offers the reader a clear sense for his thinking in this context:

> The fact that the objectives of psychometric work are determined from without is true of many other practical sciences as well, for example the applied natural sciences. But here again, one must guard against extending the parallels too far. The chemist who prepares explosives receives his objective from others: at times its purpose will be to cause explosions in mines; at other times, such as in war, the material will be used to blow up bridges. The immediate goal of his work, to produce explosives, is and remains, in and of itself, neutral with respect to the goals that it will serve, and so has no single meaning in and of itself. Drawing parallels between psycho-technics and the technical or economic special sciences would be similarly inadequate. This is so because the psycho-technician does not work on machines or on wares (in short: on 'things'), but instead on human beings, and human beings are and remain under all conditions centers of their own meaning and values. They remain 'persons' even when they are studied and treated from the standpoint of a transpersonal goal. (Stern, 1933, pp. 54–55, parentheses in original)

Unmindful of Stern's caveat, the collective voice of the putatively authoritative mainstream of scientific psychology directs the subjects of psychological research, as well as the consumers of the knowledge claims that mainstream thinkers commonly attach to that research, to view themselves and others just as the researchers view them. On that view, one's own and everyone else's identity is defined and fixed by the person *kind(s)* into which each individual has been categorized. This entails implicit acceptance of the notion that one's own and everyone else's psychological doings really are causally determined in just the ways that the scientific authorities within mainstream psychology claim that they are, and the notion that anyone—oneself or another—could properly be regarded as unique, or inherently valuable, or in any way at all determinative of one's own doings—which is the view advanced by Stern's critical personalism and by humanistic thinkers more generally—must finally be seen as quaintly naive or, worse, non- or even anti-scientific.[8]

In short, the currently prevailing ethos is such that one must simply accept that psychological scientists can rightly move from statistical evidence entitling claims of the sort 'I know "this" about *your kind*!' to claims, based on the same evidence, of the sort 'I know "this" about *you*!' To the extent that the considerations emphasized in the present volume can make clear the erroneous conceptual nature and practically untoward consequences of this move, space can perhaps be cleared for a renewed consideration of ideas that are not only much sounder methodologically from a strictly epistemic standpoint but also more compatible theoretically with the views of non-mechanistic thinkers such as those cited above.

At the very least, it has perhaps by this point become abundantly clear to the reader that the statisticism that for much of the twentieth century and now well into the twenty-first has thoroughly saturated traditional mainstream thinking about the proper conduct of scientific psychological research is neither epistemically tenable *as psychology* nor socio-ethically unproblematic. The heretofore dominant view can no longer be regarded as a science of persons superior to any that could issue from views that do not conform to the tenets of population-level statistical thinking, if only because adherence to the tenets of such thinking does not lead to a science of *persons* at all. In order for a truly scientific *psychology* to survive, change is imperative.

Notes

1. Possibilities in this regard will be considered in Chapter 7. For the present, I will only reemphasize that in psychology, knowledge about individuals is necessary whether one's objectives are 'nomothetic' or 'idiographic' or both.
2. In some studies, the comparison condition might be an alternative treatment instead of a no treatment control.
3. In personal correspondence from a colleague who is currently a practicing clinician (I myself am not), I have been told that decisive steps in this direction have been taken in Great Britain, and there has been discussion of jurisdictions in Canada following Britain's lead (Tasco, Town, Abbas, & Clarke, 2018). Further, my colleague has indicated that in Veterans Administration (VA) hospitals in the United States, there has been a major shift, supported by at least one of the 2018 candidates for the presidency of the APA (Hollon, 2017) toward evidence supported treatments (ESTs), especially for post-traumatic stress disorder (PTSD). This shift is narrowing the choices that veterans will have in seeking treatment for PTSD to treatments that are considered within the VA as ESTs for PTSD.
4. For a much broader but still decidedly mainstream view of the uses of testing in industrial and organizational psychology, see Landy and Conte (2010).
5. This is a view that Stern (1933) named the 'harmony argument,' i.e., the argument, against which he expressed critical reservations, that 'in the service of other goals, psycho-technical tests in and of themselves work to the advantage of those to whom they are applied' (Stern, 1933, p. 55).
6. The conceptually problematic nature of claims to cause–effect knowledge even when those claims are based on the findings of treatment group experiments was discussed in Chapter 4 (see also Lamiell, 2016).
7. As Stern himself indicated, that lecture, which was published two years after the 1931 Moscow conference, was 'in a certain sense an expansion and extension of a lecture titled "Personality Research and the Methods of Testing" that was given at the Fourth International Conference for Psychotechnics, held in Paris in 1927' (Stern, 1933, p. 52). That lecture was published two years later (Stern, 1929).
8. Consider, for example, the remark by Nunnally in his 1967 book titled *Psychometric Theory*, commenting on Gordon Allport's pleas for more

idiographic inquiry in personality psychology: 'The idiographists may be entirely correct, but if they are it is a sad day for psychology. Idiography is an antiscience point of view' (Nunnally, 1967, p. 472).

References

Allport, G. W. (1968). *The person in psychology: Selected essays by Gordon W. Allport*. Boston: Beacon Press.

APA Presidential Task Force on Evidence-Based Practice. (2006). Evidence-based practice in psychology. *American Psychologist, 61,* 271–285.

Bakan, D. (1955). The general and the aggregate: A methodological distinction. *Perceptual and Motor Skills, 5,* 211–212.

Baker, T. B., McFall, R. M., & Shoham, V. (2008). The current state and future of clinical psychology. Toward a scientifically principled approach. *Psychological Science in the Public Interest, 9,* 67–103. https://doi.org/doi.org/10.1111/j.1539-6053.2009.01036x.

Banicki, K. (2018). Psychology, conceptual confusion, and disquieting situationism: Response to Lamiell. *Theory and Psychology, 28,* 255–260. https://doi.org/10.1177/0959354318759609.

Danziger, K. (1987). Statistical method and the historical development of research practice in American psychology. In L. Krueger, G. Gigerenzer, & M. S. Morgan (Eds.), *The probabilistic revolution, Vol. 2: Ideas in the sciences* (pp. 35–47). Cambridge, MA: MIT Press.

Hake, A. (2001). *Was sagen gruppenstatistische Kennwerte über den Einzelfall aus? Ein Text-und Übungsbuch* [What do aggregate statistics reveal about the single case? A text and workbook]. Landau: Verlag Empirische Pädagogik.

Hanson, F. A. (1993). *Testing testing: Social consequences of the examined life*. Berkely, CA: University of California Press.

Hogan, R., Hogan, J., & Roberts, B. W. (1996). Personality measurement and employment decisions: Questions and answers. *American Psychologist, 51,* 469–477.

Hollon, S. D. (2017). Statement of candidate for APA president. *Monitor on Psychology, 48,* 66.

Kahneman, D., & Tversky, A. (1982). *Judgment under uncertainty: Heuristics and biases*. Cambridge and New York: Cambridge University Press.

Kazdin, A. E. (2008). Evidence-based treatment and practice: New opportunities to bridge clinical research and practice, enhance the knowledge base, and improve patient care. *American Psychologist, 63,* 146–159. https://doi.org/1037/0003-066X.63.3.146.

Kerlinger, F. N. (1979). *Behavioral research: A conceptual approach*. New York: Holt, Rinehart, & Winston.

Lamiell, J. T. (1998). "Nomothetic" and "idiographic": Contrasting Windelband's understanding with contemporary usage. *Theory and Psychology, 8,* 23–38.

Lamiell, J. T. (2003). *Beyond individual and group differences: Human individuality, scientific psychology, and William Stern's critical personalism*. Thousand Oaks, CA: Sage.

Lamiell, J. T. (2013). Statisticism in personality psychologists' use of trait constructs: What is it? How was it contracted? Is there a cure? *New Ideas in Psychology, 31,* 65–71. https://doi.org/10.1016/j.newideapsyh.2011.02.009.

Lamiell, J. T. (2016). On the concept of "effects" in contemporary psychological experimentation: A case study in the need for conceptual clarity and discursive precision. In R. Harré & F. M. Moghaddam (Eds.), *Questioning causality: Scientific explorations of cause and consequence across social contexts* (pp. 83–102). Santa Barbara, CA: Praeger.

Landy, F. J., & Conte, J. M. (2010). *Work in the 21st century: An introduction to industrial-organizational psychology* (3rd ed.). Hoboken, NJ: Wiley-Blackwell.

Lilienfeld, S. O., Ritschel, L. A., Lynn, S. J., Cautin, R. L., & Latzman, R. D. (2013). Why many clinical psychologists are resistant to evidence-based practice: Root causes and constructive remedies. *Clinical Psychology Review, 33,* 883–900.

Meehl, P. E. (1973). Why I do not attend case conferences. In P. E. Meehl (Ed.), *Psychodiagnosis: Selected papers* (pp. 225–302). Minneapolis: University of Minnesota Press.

Nunnally, J. C. (1967). *Psychometric theory*. New York: McGraw-Hill.

Porter, T. M. (1986). *The rise of statistical thinking: 1820–1900*. Princeton, NJ: Princeton University Press.

Rychlak, J. W. (1988). *The psychology of rigorous humanism* (2nd ed.). New York: New York University Press.

Stern, W. (1917/2010). Psychology and personalism (J. T. Lamiell, Trans.). *New Ideas in Psychology, 28,* 110–134. https://doi.org/10.1016/j.newideapsych.2009.02.005.

Stern, W. (1929). Persönlichkeitsforschung und Testmethode [Personality research and the methods of testing]. *Jahrbuch der Charakterologie, 6,* 63–72.

Stern, W. (1933). Der personale Faktor in Psychotechnik und praktischer Psychologie [The personal factor in psychotechnics and applied psychologie]. *Zeitschrift für angewandte Psychologie, 44,* 52–63.

Tasco, G. A., Town, J. M., Abbas, A., & Clarke, J. (2018). Will publicly funded psychotherapy in Canada be evidence based? A review of what makes psychotherapy work and a proposal. *Canadian Psychology/Psychologie Canadienne, 59,* 293–300.

Valsiner, J. (1986). Between groups and individuals: Psychologists' and laypersons' interpretations of correlational findings. In J. Valsiner (Ed.), *The individual subject and scientific psychology* (pp. 113–151). New York: Plenum.

Venn, J. (1888). *The logic of chance*. London and New York: Macmillan.

Wendt, D. C., & Slife, B. D. (2007). Is evidence-based practice diverse enough? Philosophy of science considerations. *American Psychologist, 62,* 613–614. https://doi.org/10.1037/0003-066X62.6.613.

Windelband, W. (1894/1998). History and natural science (J. T. Lamiell, Trans.). *Theory and Psychology, 8,* 6–22.

Check for updates

7

In Quest of Meaningful Change

The time is long overdue for a clear and unequivocal acknowledgment within scientific psychology's mainstream of the conceptual confusions that have long infected prevailing understandings of the knowledge claims that population-level statistical investigations can and cannot warrant. That acknowledgment should be followed promptly by a discipline-wide shift to investigative methods that are formally compatible with psychology's original mission: achieving scientific accounts of the psychological doings of individuals.

In this final chapter, I discuss a variety of matters about which maximum clarity will be essential if the obdurate resistance to the needed change that has thus far prevailed within psychology's mainstream is ever finally to be overcome. Following a brief review of the major historical developments, elaborated in previous chapters, that landed mainstream psychology in its current epistemic predicament, attention is directed to a promising alternative investigative framework, called 'observation oriented modeling' (OOM), that is currently being developed by James W. Grice. I then reiterate certain points on which widespread confusion continues to obstruct efforts toward change. The chapter concludes with a call for an expanded vision of psychological

J. T. Lamiell, *Psychology's Misuse of Statistics and Persistent Dismissal of its Critics*, Palgrave Studies in the Theory and History of Psychology,
https://doi.org/10.1007/978-3-030-12131-0_7

science that would revive the appreciation that once existed for the discipline's dual nature as both a natural science and a human science.

A Synopsis of Historical Developments Leading to Mainstream Psychology's Current Epistemic Predicament

It was noted in Chapter 1 that concerns about the issues of focal concern in this work were concisely but explicitly raised within the mainstream of scientific psychology no later than mid-twentieth century. It was then that David Bakan issued an unobtrusive but pointed call for a clear distinction 'between general-type and aggregate-type propositions' (Bakan, 1955, p. 211). He noted that '(t)he use of methods which are appropriate to the one type (of proposition) in the establishment and confirmation of the other (type), leads to error' (Bakan, 1955, p. 211, parentheses added). In a subsequent publication, Bakan (1966) correctly pointed out that general-type propositions were those at which the late nineteenth century investigative efforts of the original experimental psychologists (Wundt, Ebbinghaus, and their contemporaries) were aiming, while the later-emerging work of the differential/correlational psychologists had to be understood as suited strictly to the establishment of aggregate-type propositions.

It had happened, however, that for decades prior to Bakan's writings on this topic, differential/correlational psychologists thinking along lines championed by such prominent early twentieth-century figures as E. L. Thorndike and Hugo Münsterberg—but not by their contemporary and differential psychology's equally prominent founder, William Stern—had already been misinterpreting their aggregate statistical research findings as justifying not only claims to knowledge of aggregates, but also claims to knowledge about the individuals within those aggregates (refer to Chapter 3). What is more, during this same period of time mainstream experimental psychologists were effecting what would become a virtually complete transformation of their subdiscipline from one committed to the $N=1$ Leipzig model of investigation, a model that *was* formally suited to the validation of general-type propositions and that did *not* entail the use of aggregate statistical methods, into one committed primarily to a treatment group model of experimentation reliant upon the same aggregate statistical methods that had been introduced to psychology by the differential/correlational psychologists (refer to Chapter 4).[1] Of crucial significance in this regard was the widespread belief that this transformation of experimental psychology was 'epistemically seamless.' That is, the consensus—but mistaken—view was that general-type propositions could be validated *by means of* discovering of aggregate statistical regularities (refer to Chapter 4).

The epistemic commitments that thus came to prevail within each of scientific psychology's 'two disciplines,' and that continued to prevail through the merger of the two that Cronbach (1957) urged, could not accommodate the concerns raised by Bakan (1955, 1966). However, since neither could Bakan's argument be refuted, mainstream psychology's only recourse was to simply ignore it, and that is exactly what happened.

Inevitably, the repressed would return several years after Bakan's 1966 publication, when Fred N. Kerlinger furrowed his brow over research psychologists' 'troublesomely paradoxical' practice of using aggregate statistical regularities as the basis for claims to knowledge about the psychological doings of individuals (Kerlinger, 1979)—precisely the conceptual problem to which Bakan had pointed. Once again, however, the problem was simply ignored, even in subsequent writings by Kerlinger himself (refer again to Chapter 1), and the 'troublesomely paradoxical' practice continued unabated.

Awakened to an initial understanding of that problem by students in an undergraduate course in the psychology of personality, I myself broached the topic in a 1981 *American Psychologist* article (Lamiell, 1981). In that work, I highlighted the inappropriateness of drawing inferences about the psychological characteristics of individuals on the basis of aggregate statistical research findings defined primarily in terms of the reliability and validity coefficients associated with various tests of individual differences in personality traits (refer to Chapter 2). A quarter of a century later, Harré (2006) would ruefully observe that the argument I had launched in the 1981 article and then expanded in numerous subsequent publications had remained unheeded—but still unrefuted—within the mainstream.

Sadly, the situation has remained unchanged to this day. Moreover, as invalid understandings of the knowledge secured by aggregate statistical methods have infected the thinking not only of psychologists engaged in basic research, but also the understandings of applied psychologists implementing, or endorsing the implementation of, interventions in the lives of individuals, the problem must be recognized as having a socio-ethical facet as well (refer to Chapter 6).

It is in consideration of this long and deeply problematic history that I have found warrant for characterizing mainstream psychology as 'incorrigible.' It is a discipline that, for over six decades now, has proven itself unable or unwilling to acknowledge the fundamental invalidity of interpreting statistical knowledge of variables with respect to which individuals have been differentiated (whether by tests for existing differences or by experimental treatments designed to produce the differences, or by some combination of both) as knowledge about the individuals who have been differentiated in terms of those variables.

Statistical knowledge of variables with respect to which individuals have been differentiated is knowledge that is defined only for populations. To see this is to see that research devoted to the generation of such knowledge is, in effect, a species of demography. Because investigators identifying themselves as psychologists often do harbor substantive theoretical interests in psychological phenomena, the field can justifiably be characterized as *psycho*-demography, but it is demography all the same, and, as such, really has nothing of scientific authority to say about the individuals within the populations that are studied. A discipline that has nothing of scientific authority to say about individuals cannot be a scientific *psychology*, whatever else and however meritorious in its own right that discipline might be.

As I have stressed repeatedly, both in this work and in many other writings, the problem here is not some gnat-like 'paradox' that will one day be swatted away without requiring any fundamental changes in investigative practices. The problem is a deep, abiding, and very serious epistemic flaw in mainstream thinking that has existed for decades and that cannot be eradicated from within the framework of the established methodological canon. It is nothing less than that 'great failure of understanding' about which the German mathematician and philosopher Moritz Wilhelm Drobisch wrote in 1867 (cf. Porter, 1986), and overcoming that great failure is going to require a commensurately great reorientation of mainstream investigative sensibilities.

If my own first-hand struggles with conventional thinking on this matter have thus far been stonewalled by the mainstream, they have at least helped me to see more clearly the stones in the wall, i.e., what seem to me to be the major impediments to change. It must be emphasized that those impediments are not simply—or even primarily—of a technical nature and hence surmountable simply by resorting to alternative investigative methods. Unquestionably, alternative methods are needed, and this point will be addressed further in the pages to follow. In the main, however, the challenges are *conceptual* in nature, and, consistent with Wilhelm Wundt's prediction in 1913 (Wundt, 1913/2013), psychology's eventual divorce from philosophy has not only greatly diminished—arguably to near extinction—mainstream psychologists' interest in conceptual questions (a development mourned by Machado and Silva, 2007), but has also compromised mainstream thinkers' ability to deal with such questions when they are raised by scholars who do appreciate their importance.

Be this as it may, the effort to get the discipline back on sound epistemic footing *as psychology* must proceed, and in doing so should strive to eliminate as many conceptual impediments as manifest themselves along the way. The remainder of this final chapter has been written in this spirit.

On the Nature and Value of Psycho-demographic Inquiry

In this work, I have argued that most of mainstream psychological research—correlational, experimental, and hybrid—has long since become, effectively, a species of demography that can reasonably be called 'psycho-demography.' This label is warranted because although the knowledge actually gained through most of that research is essentially demographic in nature—it is knowledge of statistical relationships between variables marking *between-individual differences* that are definable *only* for populations—it is conducted and published by persons who identify as psychologists and who have theoretical interests in what I have called the psychological 'doings' of *individuals*—sensations, perceptions, judgments, emotions, cognitions, memories, and behaviors.[2]

Though the point was stated previously, it is sufficiently important to warrant re-emphasis here: the argument is *not* that psycho-demographic inquiry is without merit. On the contrary, the knowledge gained through such inquiry can be very useful, both in its own right as a guide to public policy, and indirectly as a source of *hypotheses about* psychological phenomena that might subsequently be tested using methods that are logically suited to gaining individual-level knowledge. A study recently published by Johnson, Markowitz, Hill, and Phillips (2016) can serve to illustrate each of these possibilities.

Psycho-Demographic Inquiry as a Guide to Public Policy

The research discussed in the article just cited was conducted by Johnson, Markowitz, Hill, and Phillips as affiliates of the Center for Research on Children in the United States (CROCUS), administered through the McCourt School of Public Policy at Georgetown University in Washington, DC. The study focused on children who either had just completed or would soon begin a pre-kindergarten (pre-K) program at one of 45 schools in the Tulsa, OK Public Schools system.[3] The investigators aimed to determine (1) if a measure of the 'impact' *(sic)* of pre-K programs on various indicators of children's cognitive functioning varied across the 45 schools, and, if so, (2) if 'impact' variability could be found to covary with a measure of the quality of instructional support designed into the respective pre-K programs.[4]

The 'impact' of a given pre-K program was expressed quantitatively by comparing the average level of cognitive functioning displayed by children enrolled in that program to the average level of cognitive functioning displayed by children who were about to (but had not yet) entered the program. In a total sample constituted of 1195 'treated'[5] children and 1417 comparison children (Johnson et al., 2016, p. 2148), sophisticated statistical analyses indicated (1) that there was, indeed, variation in the measure of pre-K program 'impact' across the 45 schools included in the study, and that (2) that variation did, indeed, covary significantly with the measure of the quality of the respective programs.[6] Further statistical analyses were conducted in order to estimate, in raw score points defined on each of the scales used to assess the children's cognitive functioning, the average increase or decrease in the impact of a program per standard deviation of difference in program instructional quality, relative to the average 'impact' of schools with average instructional quality.[7] More specifically, those analyses revealed that each standard deviation increase in program 'impact' (itself a function of program quality) was associated with an average increase in cognitive functioning of 12, 14, or 23%, depending upon the particular cognitive function involved.[8]

Reflecting their affiliation with a university-based school of public policy (refer above), the authors concluded the report of their findings as follows:

> As federal and state attention to expanding pre-K availability intensifies, it is increasingly important to develop our toolbox of 'what works' in pre-K programs. … As the first study of its kind, more research is needed before policy recommendations based on these results can be made. Nevertheless, our findings contribute to ongoing conversations about how best to promote early learning for public pre-K students. … If our findings are replicated, they suggest that increasing pre-K instructional quality is one way to boost the positive impacts of pre-K exposure on children's cognitive school readiness. (Johnson et al., 2016, pp. 2155–2156)

Obviously, the emphasis here was, exactly as it ought to have been, on research of the sort discussed as a means of informing public policy through careful empirical investigation. In this case, the domain of substantive interest was early childhood education. The overriding question was: What worked relatively well in the Tulsa pre-K programs and might therefore, pending the outcomes of additional studies, warrant the investment of community resources, both in Tulsa and, possibly, elsewhere, going forward? The answer to the question of 'what worked relatively well' was achieved through the statistical comparison of program averages on various criterion variables reflecting levels of children's cognitive functioning.

The question in this study was not—nor ought it to have been—about the 'impact' of a given pre-K program on the cognitive functioning of individual children, and, indeed, no answer to that question was or could possibly have been revealed by the statistical analyses that produced the study's findings. Those findings constitute knowledge about what transpired, on average, within populations of pre-K children exposed to different programs. That is what makes the study essentially demographic in nature, and that study and others like it can be valuable because demographic knowledge is, very often, exactly the kind of knowledge that the makers of public policy need.

The dependent variables in the study were defined on scales obviously designed to tap content of a psychological nature: a child's level of cognitive functioning. While this fact qualifies the research as psycho-demography, it does not qualify the research as psychology. The aggregate statistical analyses that produced the study's findings did not reveal *anything* about *any* child's cognitive functioning. Arguably, the dependent variable assessments that were made of a given child provided psychological knowledge of that child, but that knowledge necessarily existed prior to and entirely separate from the determination of the aggregate statistical relationships that define the study's findings. Such aggregate statistical relationships constitute demographic knowledge even when that knowledge is secured by aggregating individual-level observations of psychological phenomena.

Yet, at least three of the four investigators involved in the study being discussed here (Johnson, Markowitz, and Phillips) identify themselves as psychologists. Those three participated in the conduct of the Tulsa study not only as affiliates of the Georgetown University McCourt School of Public Policy, but also as members of that same university's Department of Psychology. All were prominent in that department's Ph.D. program in 'developmental science.'[9] There is, therefore, ample reason to suppose that each of the three has theoretical interests in the cognitive functioning of individual children, and might therefore have an interest in testing hypotheses in that psychological domain prompted by the findings revealed by the aggregate statistical methods they so skillfully employed for psycho-demographic purposes.[10] I turn now to a description of one way in which they might proceed.

Testing Hypotheses About Psychological Functioning Prompted by Psycho-demographic Inquiry

It was mentioned above that Johnson et al. (2016) derived raw score estimates of the average difference in program 'impact' on a given cognitive function per standard deviation of difference in pre-K program quality, relative to average 'impact' on that function of a pre-K program of average quality. The authors made clear their appreciation for the fact that, as averages, those findings did not reveal individual-level results. However, those averages do suggest testable hypotheses—even if only actuarially- rather than theoretically based—about what the findings could prove to be in individual cases.

For example, knowing that the average impact of a pre-K program of average quality was 3.22 points on the letter-word identification task, and that each standard deviation of difference in program quality was associated with a difference in average performance on that task of 0.44 points (Johnson et al., 2016, p. 2151), it might be hypothesized that a child who attended a pre-K program the assessed quality of which was one standard deviation above the mean would score somewhere within a narrow interval of scores centered around the value 3.66, while a child who attended a pre-K program of assessed quality one standard deviation below the mean would score somewhere within a narrow interval of scores centered around the value 2.78.[11] In any case, an hypothesis built upon this line of reasoning could be tested on each one of the 1195 children who had attended one of the 45 pre-K programs included in the Johnson et al. (2016) study, and the findings revealed by those tests—hypothesis confirmed or not—would reflect the *generality* of the 'impact' findings reported by Johnson et al. (2016), in the sense of 'general' understood by scientific psychology's original experimentalists (refer to Chapter 1). On the view being advanced here, this would qualify the research as psychological because (a) it would be an investigation of a psychological phenomenon, level of cognitive functioning, and (b) it would generate individual-level findings. By contrast, it is perhaps clearer at this point that nothing of this sort could possibly be accomplished by aggregate-level null hypothesis significance testing.

Observation-Oriented Modeling as an Alternative Framework for Psychological Research

The idea at the core of the individual-level analyses just described is one that James W. Grice (b. 1964) has, over the past decade, developed into an insightful and very promising approach to analyzing data from psychological experiments (see, e.g., Grice, 2011, 2014, 2015). He calls his approach 'observation oriented modeling' (OOM), and he has developed software to facilitate its adoption and use. As an alternative to conventional aggregate-level statistical investigations, OOM merits careful consideration, particularly by those reluctant to embrace non-quantitative methods.

One of the many salutary features of OOM is that it forces an investigator to reflect carefully on, and then express as precisely as possible, his/her theoretical understanding of the psychological processes at play in determining his/her empirical findings. For this purpose, Grice advocates the use of pictograms constructed of stick figures and geometric forms (see, e.g., Grice, 2015, p. 3). A given pictogram provides a kind of visual 'snapshot' of the structures and processes, or causes and effects, that are theoretically presumed to be at play in producing the patterns to be found in one's data. Analysis of those data is then aimed at determining the frequency with which the theoretically expected outcomes are—and are not—empirically realized in individual subjects.

Grice (2015) discusses a study in the social psychology of rejection to illustrate OOM. The procedure used in that study called for a male college student to be told that he would be interacting online with another male student located elsewhere on campus. The participant was asked to provide a short biographical sketch to share with his counterpart and was then given a corresponding biographical sketch of that counterpart. The counterpart's biographical sketch described an individual who is kind and inquisitive and likely to be pleasing to interact with in an informal social setting. After reading the counterpart's biographical sketch, the participant made ready for the online interaction, only to be informed by the experimenter that the counterpart had suddenly decided not to participate in the online discussion after all, and was withdrawing from the experiment.

It was assumed that the counterpart's abrupt withdrawal from the experiment would engender feelings of rejection in the participant. This was termed the 'rejection' condition of the experiment. In a comparison condition, the 'no-rejection' condition, the need to suddenly abort the expected interaction was attributed to a failure of the counterpart's computer.

Each of these two conditions called for the participant to rate the counterpart, presumably on the basis of the counterpart's biographical sketch, on several qualities, one of which was 'popularity.' The ratings were to be expressed on a 6-point scale, where, in the case of 'popularity,' a rating of 1 would indicate that the participant was viewing the counterpart as 'very unpopular,' and a rating of 6 would indicate that the participant was viewing the counterpart as 'very popular.' Presumably, the negative feelings experienced by a participant in the 'rejection' condition would cause him to rate the counterpart negatively, i.e., as unpopular, while the absence of such feelings in a participant in the 'no rejection' condition would cause him to rate the counterpart positively, i.e., as popular.

Grice (2015) then presented the results of data analyses carried out for three different samples, each of which was comprised of 160 subjects. Discussing first the results obtained by analyzing the data conventionally, using *t*-tests for the difference between means of independent groups, Grice reported a statistically significant difference ($p < .05$) in the hypothesized direction for each of the three samples. In the first sample, the ratings made by the subjects in the 'rejection' condition were, on average, .3 of a scale point lower than the average of the ratings made by the subjects in the 'no-rejection' condition. In the second and third samples, the average ratings made by the subjects in the 'rejection' condition were, respectively, .6 and .3 of a scale point lower than the average ratings made by the subjects in the 'no-rejection' condition. For the three samples, the obtained t-values for the differences between the treatment group means (computed as 'rejection' group mean minus 'no rejection' group mean) were −2.10, −2.06, and −2.10, respectively. By conventional thinking, these findings would be interpreted as empirical confirmation of the hypothesized psychological effect of rejection on an individual's judgment of a rejecting person.

Grice (2015) then discussed the findings of the same three studies obtained when the data were analyzed using the OOM method. To illustrate how hypothesis testing in OOM can vary in stringency depending upon the level of precision at which an investigator is prepared to articulate his/her theoretical expectations, Grice (2015) set forth two possible data patterns. For the experiment under discussion here, a very stringent test would be against the hypothesis that a participant who had experienced the 'rejection' condition would judge the rejecting counterpart as negatively as possible and therefore assign a 'popularity' rating of 1 to that counterpart, while a participant who had experienced the 'no-rejection' condition would judge the counterpart as favorably as possible and therefore assign a popularity rating of 6 to the counterpart. An alternative and decidedly less stringent test of the hypothesis would be to consider as confirmatory a rating of 1, 2, or 3, i.e., anywhere on the negative side of the rating scale, by a participant in the 'rejection' condition, and a rating of 4, 5, or 6, i.e., anywhere on the positive side of the rating scale, by a participant in the 'no-rejection' condition.

In any case, hypothesis testing in OOM requires the investigator to make explicit what should be observed of any given participant in order for that participant to be regarded as having confirmed—or not—the hypothesis. The data are then tallied on a case-by-case basis. Table 7.1 presents the final tallies reported by Grice (2015) for each of the three samples. There it can be seen that in Sample 1, 64 of the 80 participants in the 'no-rejection' condition did assign the counterpart a rating on the positive half of the rating scale: 16 participants assigned a rating of 4, 32 participants assigned a rating of 5, and 16 participants assigned a rating of 6. Viewed alone, those results are favorable to the hypothesis. However, and contrary to the hypothesis, 16 of the 80 subjects in the 'no-rejection' condition assigned the counterpart a rating on the *negative*

Table 7.1 Results of OOM analyses reported by Grice (2015)

Frequencies of scale ratings across experimental conditions

Sample 1

	Popularity rating					
	1	2	3	4	5	6
Rejection	-	-	-	64	16	-
No-rejection	-	8	8	16	32	16

Sample 2

	Popularity rating					
	1	2	3	4	5	6
Rejection	16	16	-	-	-	48
No-rejection	-	8	8	8	24	32

Sample 3

	Popularity rating					
	1	2	3	4	5	6
Rejection	-	-	16	40	16	8
No-rejection	-	-	8	40	16	16

half of the rating scale, with eight of them assigning a rating of 2 and eight others a rating of 3. What is even less favorable to the hypothesis is that none of the 80 participants exposed to the ‘rejection’ condition assigned to the counterpart a rating on the negative half of the rating scale. Instead, 64 of those 80 participants assigned to the counterpart a rating of 4 and 16 a rating of 5.

Even when evaluated against the less stringent of the two hypotheses discussed above, where a rating of 1, 2, *or* 3 by a ‘rejected’ participant and a rating of 4, 5, *or* 6 by a ‘non-rejected’ participant counts as a ‘hit,’ the finding in Sample 1 was that the hypothesis was confirmed in only 40% of the cases. The hypothesis fared slightly better in Samples 2 and 3, where, respectively, 60 and 55% of the cases qualified as ‘hits,’ i.e., conformed to theoretically based expectation. Nevertheless, as was true for Sample 1, Samples 2 and 3 also produced abundant evidence counter to the hypothesis. The finding in Sample 2 that fully 48 of the 80 'rejected' participants assigned the most favorable rating possible, 6, to the counterpart, while only 32 of those 'rejected' participants behaved as theoretically anticipated, runs strongly counter to the hypothesis. Similarly in Sample 3: while 72 out of the 80 'non-rejected' participants behaved as expected, assigning the counterpart a rating of 4, 5, or 6, a rating that high was also assigned to the counterpart by fully 64 of the 80 *'rejected'* participants.

The findings revealed by OOM, as summarized in Table 7.1, can hardly be regarded as favorable to the hypothesized psychological effect of rejection on individuals' judgments of a rejecting person, a conclusion quite at odds with the one that would conventionally be drawn given the results of the *t*-tests discussed previously. For reasons elaborated throughout the course of this book, those *t*-tests provided no knowledge whatsoever of what was transpiring at the level of individual subjects. When that knowledge is acquired by means of an appropriate method such as OOM, the empirical realities that can be obscured by generalizations of the sort commonly regarded within psychology's mainstream as justifiable by aggregate statistical analyses become starkly apparent.[12]

In the course of developing OOM, Grice has noted that certain statistical computations can prove helpful as aids to the interpretation of findings. The simplest of them is what he calls 'percent correct classification,' or PCC, which is equivalent to what I termed above 'proportion of hits.' Grice also writes of a '*c*-value,' (or *chance*-value) which can help to place a given PCC value into larger context. The *c*-value expresses the proportion of instances in 1000 random pairings of research participants with dependent variable observations that yielded a PCC index greater than the one actually obtained. As an index of proportion, the *c*-value can range from zero to 1, with a relatively low value indicating that an obtained PCC value was infrequently exceeded by results obtained with the randomized pairings. For Sample 1 of the rejection study discussed above, Grice (2015) reported a *c*-value of 1.0, indicating that in all 1000 of the randomized pairings of the 160 research participants with the popularity ratings, the obtained PCC index of 40% was exceeded. Grice noted that "the observed PCC index was therefore not only low, but values at least that high were entirely ordinary as well" (Grice, 2015, p. 8).

All of the foregoing said, Grice has repeatedly emphasized that such statistical computations are of strictly secondary importance (cf. Grice, 2011, 2015; Grice, Barrett, Schlimgen, & Abramson, 2012; Grice et al. 2017). As the name 'observation oriented modeling' suggests, what is always of primary importance is what can be learned by a careful visual examination of the data, case by individual case, in the light of the iconic or 'integrated' model of the presumed structures and processes, or causes and effects, that one supposes, theoretically, are producing those data. As Grice put matters in his discussion of the research just described:

> The methods shown in this paper represent a return to the person or persons in psychology. Because these methods are primarily visual in nature and do not rely on the computation of parametric statistics; outliers or assumptions of normality, homogeneity, etc., are never a concern. The Percent Correct Classification index is a simple frequency, and therefore an aggregate statistic, but it is always interpreted in light of a pattern, … and the complete set of observations. The simple 'eye test' or more severe 'inter-ocular traumatic test'[13] … is taken seriously in OOM as there is simply no substitute for examining the data, particularly in light of an integrated model. (Grice, 2015, p. 11)

For the reader seeking a method for conducting and analyzing the results of psychological experiments in a way that avoids the conceptual pitfalls that have plagued mainstream psychological research for decades, OOM offers one very promising alternative.

Some Additional Points of Needed Conceptual Clarity

There Is a Role for Quantitative Methods in Psychological Research

Lest any confusion on this point be left to linger, it should be emphasized that the argument set forth in this book is *not* an argument against the use of quantitative methods in psychological research. Rather, it is an argument directed specifically and quite pointedly against the use of *aggregate* statistical investigative methods to gain knowledge about the psychological doings of individuals. It is the aggregate nature of those methods, and not their quantitative nature per se, that is problematic.

In research that I directed many years ago, extensive and fruitful use was made of quantitative methods in testing hypotheses about the nature of the judgment process by which individuals formulate and express subjective ratings of their own and one another's personality characteristics (see, e.g., Lamiell & Durbeck, 1987; Lamiell, Foss, Larsen, & Hempel, 1983). That research can serve, alongside the OOM approach discussed above, as another alternative model for psychological research that is free of the conceptual pitfalls inherent within the paradigm long-dominant within the mainstream.

Another salutary development altogether hospitable to the use of quantitative methods in studies of individuals is the recent (2015) founding of a publication titled *Journal for Person-Oriented Research.* That journal, published in Sweden,[14] is linked with an organization called the Society for Person-Oriented Research. The mission statement of the new journal includes the following passage:

> Person-oriented research refers to theoretical, methodological, and empirical research that is guided by a research paradigm in which the individual is at focus and seen as a functioning totality. This paradigm implies that theories and findings should be interpretable at the level of the individual and that patterns of individuals' characteristics are of key interest. Hence, a standard variable-oriented approach with the variable as the basic conceptual and analytic unit, and analyzing data using group statistics, for example, correlational analysis, falls normally outside the journal's scope. (*Journal for Person-Oriented Research*, statement of Aims and Scope, 2015, Volume 1, issue 1–2)

This statement, and, indeed, the very existence of the *Journal for Person-Oriented Research*, makes clear that there already exists a recognition on the part of some psychologists—currently outside the mainstream—that conventional aggregate-level, variable-oriented inquiry simply does not serve the scientific objective of advancing our understanding of the psychological doings of real individual persons, and that something on the order of a paradigm shift is not only necessary but is also possible. Readers who perhaps are now persuaded of the merits of this view but uncertain about what to do next, especially within the context of a disciplinary ethos that remains largely inhospitable to nonquantitative methods of inquiry, should know that viable alternatives to the long-dominant paradigm do exist, and that models of the use of those alternatives are available *now.* Many examples can be found in the aforementioned journal.

In the same mission statement cited above, it is indicated that out of respect for 'the standard scientific criteria of objectivity and replicability of research findings, … many qualitative research approaches also fall outside the … scope' of the *Journal for Person-Oriented Research.* For reasons to be discussed below, my own view is that this restriction is regrettable, because it reflects an excessively narrow vision on what can and cannot qualify as 'scientific psychology.' As a lead-into that discussion, however, I direct attention briefly to a related yet distinct conceptual stumbling block.

Misconceptions Endure in Psychologists' Understandings of 'Nomothetic' and 'Idiographic'

In an article published 20 years ago (Lamiell, 1998), I wrote of the conceptual chasm that existed between the meanings of the concepts 'nomothetic' and 'idiographic' intended by the inventor of the terms (see Windelband, 1894/1998) and the understandings of those same concepts that had taken firm hold among psychologists by the middle of the twentieth century. The hope for that article was that it would correct the mistaken notions that (a) the study individuals is, per se, idiographic inquiry, and that (b) the quest for nomothetic knowledge of psychological doings mandates group-level inquiry.[15] Unfortunately, the aims of my 1998 article have not been widely realized among mainstream thinkers. The mistaken notions just stated persist, and they can be found not only in the writings of psychologists who firmly endorse the aggregate methods of investigation, but also in the writings of those urging a shift to more individual-level inquiry.

For example, in the article by Lilienfeld, Ritschel, Lynn, Cautin, and Latzman (2013) discussed earlier, one finds the claim that evidence-based practice (EBP), 'relies primarily on nomothetic findings, which strive to extract universal or quasi-universal laws that apply to all or most individuals within the population' (Lilienfeld et al., 2013, p. 891). In Chapter 6, it was explained both that and why null hypothesis significance testing, which is standard procedure in the randomized control trials (RCTs) to which proponents of EBP appeal, does *not* warrant knowledge claims that 'apply to all or most individuals.' The results of null hypothesis significance testing in RCTs do not warrant *any* claims to knowledge about *any* individuals at all. Precisely because this is true, the findings of RCTs cannot properly be regarded as *nomothetic* in the sense of 'nomothetic' intended by Windelband.

Re-emphasizing a point made earlier, Windelband (1894/1998) used the term 'nomothetic' to refer to knowledge of general laws in the sense of 'general' understood by psychology's original experimentalists, i.e., in the sense of *common to all* of the individuals investigated (refer to Chapter 1). It is not logically possible to adduce evidence that such commonality does—or, for that matter, does not—exist without studying individuals, and this just *means* that the study of individuals is a necessary feature of the quest for *nomothetic* knowledge in psychology.[16]

Recognizing that 'it is the individual organism that is the principle unit of analysis in the science of psychology,' Barlow and Nock (2009, p. 19) state, correctly, that the study of individual organisms in the 'hallowed tradition' of Wundt, Ebbinghaus and others among their contemporaries is fully suited to 'establishing causal relations among variables' (p. 19; refer also to Chapter 4 of this volume). Unfortunately, Barlow and Nock (2009) then err in identifying the quest for such knowledge as 'the idiographic approach' (p. 19). For reasons just stated, the study of individuals in the search for generalizable causal relations among variables is the very essence of *nomothetic* inquiry on the understanding of such inquiry articulated by Windelband (1894/1998). For Windelband, the need for idiographic knowledge exists precisely because not all that we would wish to know about the psychological doings of individuals *can* be framed in terms of laws reflecting causal relationships that are general in the sense of common to all. Apart from and irreducible to knowledge of '*was immer ist*'—what always is—true of individuals, one will inevitably encounter in the study of any individual evidence of '*was einmal war*'—what once has been—true for *that* and possibly *only* that individual. It was specifically for reference to knowledge of this latter sort that Windelband invented the term 'idiographic' (cf. Windelband, 1894/1998, p. 13).

From the foregoing, it should be apparent that in order for the concepts 'nomothetic' and 'idiographic' to be applied properly in psychology, it must be understood that *the study of individuals is necessary whether one's knowledge objectives are nomothetic or idiographic or both.* Achieving clarity on this point will also make more plausible to mainstream thinkers the notion that both kinds of knowledge can and should have a place in a genuinely *scientific* psychology. It is to this final point that the discussion now turns.

Renewing a Broadened Vision of Scientific Psychology

Looking Further at What Windelband Said

When Windelband introduced the neologisms 'nomothetic' and 'idiographic,' his overriding concern was for situating psychology properly among the scientific disciplines. The psychology of which he was writing was the original Leipzig-model experimental psychology, and he regarded its proper classification among the sciences as problematic because it seemed to straddle the dividing line between the *natural* sciences—*die Naturwissenschaften*—on the one side, and the *human* sciences—*die Geisteswissenschaften*—on the other. He wrote that 'to judge (psychology) by its subject (matter),' which he understood to be human mental life, 'it can only be characterized as a human science, … but its entire procedure, its methodological arsenal, is from beginning to end that of the natural sciences' (Windelband, 1894/1998, p. 11, parentheses added).[17]

Elaborating on these points, Windelband likened psychology to the natural sciences in that, like the natural scientists, psychologists were seeking knowledge of such general lawfulness—*allgemeine Gesetzmäßigkeit*—as could be discerned in the phenomena investigated. Throughout the natural sciences, Windelband wrote, 'it is always laws of occurrences which (are sought), whether the occurrence concerns a movement of bodies, a transformation of matter, an unfolding of organic life or (as in psychology) a process of ideation, feeling and willing' (Windelband, 1894/1998, p. 12, parentheses added).

It was in juxtaposition with these observations that Windelband drew the distinction he was aiming to highlight:

> In contrast to the foregoing, the many empirical disciplines which one otherwise properly labels as humanities[18] are directed decidedly to the complete and exhaustive portrayal of a particular more or less protracted occurrence of a unique, temporally circumscribed reality … One deals with an isolated event or an interconnected sequence of acts and fates, with the essence and life of a single man or an entire folk … But always the goal of knowledge is that of reproducing and rendering intelligible a creation of human life in its factuality. Clearly, the entire province of the historical disciplines is implied here. (Windelband, 1894/1998, p. 12)

Summing up his observations on the knowledge objectives of the natural and human sciences, Windelband wrote:

> So we may say that the empirical sciences seek in the knowledge of reality either the general in the form of the natural law or the particular in its historically determined form *(Gestalt).* They consider in one part the ever-enduring form, in the other part the unique content, determined within itself, of an actual happening. The one comprises sciences of law, the other sciences of events; the former teaches what always is *(was immer ist)*, the latter what once was *(was einmal war)*. If one may resort to neologisms, it can be said that scientific thought is in the one case nomothetic, in the other idiographic. (Windelband, 1894/1998, p. 13)

The reader will perhaps have noticed that nowhere in this passage is there the slightest indication that empirical findings constituted of statistical relationships between variables defined only for populations could possibly qualify as nomothetic knowledge in psychology. The very idea would have made Windelband chuckle, and would be merely comical today had it not led so many psychologists so far astray for so long.

Note may also have been taken that in the passage quoted earlier from p. 12 of Windelband's (1894/1998) text, he indicated clearly that although idiographic knowledge *could* be knowledge of 'a single man'—i.e., an individual person, as would certainly be the case in psychology—such knowledge could also be knowledge of 'an entire folk,' as might be the case in the discipline of anthropology. It is the nature of the knowledge, and not the entity to which that knowledge applies, that determines whether it is, or is not, to be regarded as idiographic.[19]

Broader Implications

With all of the foregoing in place, the major point to which I wish to draw attention in concluding this work is Windelband's clear recognition that *both* the natural sciences (*Naturwissenschaften*) *and* the human sciences (*Geisteswissenschaften*) qualify as *empirical sciences (die Erfahrungswissenschaften). 'Erfahrung'* is the German word for 'experience.' Hence, the *Erfahrungswissenschaften* are disciplines focused on different domains of empirical encounters, and are to be contrasted with the rational sciences (*die rationalen Wissenschaften*), logic and mathematics, which have no one specific domain of empirical content.

Windelband was concerned with the proper classification of psychology within this larger framework. To him, it was altogether clear that psychology qualified as an empirical science, with its specific empirical content being the domain of observable indications of human mental life. While that content marked psychology as a human science, and therefore as a discipline that would have need for idiographic knowledge, its quest for nomothetic knowledge, i.e., knowledge of such general laws of mental life as studies might reveal, also marked it as a discipline with some knowledge objectives that are formally similar to those of the natural sciences. The point here is not to advocate for a view of psychology as either a *Geisteswissenschaft* or as a *Naturwissenschaft*, but to note, and to consider the implications of the fact, that in either case, psychology is a *Wissenschaft*, i.e., a *science*.

Unfortunately, the view came to dominate within the mainstream that psychology could qualify as a science only if it modeled itself on the *natural* sciences (cf. Gantt & Williams, 2018). Indeed, the disciplines referred to in the German language as *Geisteswissenschaften* are typically referred to in English as 'humanities.' Obviously, the English expression does not contain the equivalent of the German word for science, '*Wisssenschaft,*' and, consistent with this fact, those disciplines are not commonly thought of within the (English-dominated) mainstream of psychology as 'sciences' at all. However, with Schiff (2017) and others (cf. Schiff, 2018) I believe that psychology would be enriched by a revival of a broader vision of the discipline as a science, a vision that, as a matter of historical fact, once prevailed.

The German word *Wissenschaft* is a compound of the words *'Wissen,'* which means knowledge (as a verb, '*wissen*' means 'to know'), and a derivative of the verb '*schaffen*,' which means 'to do' or 'to make.' Hence, to 'do science' is to *schaffen Wissen*—to make knowledge—with no a priori restrictions on just *how* one is to go about this task, or on *what* the nature of the resulting knowledge must be. Some methods, chiefly experimentation and mathematization, have proven well-suited to the natural sciences, where the goal is to *explain.* Other methods, chiefly qualitative in nature, seem better suited to the human sciences, where the goal is to *understand* (cf. Dilthey, 1894). If psychology is both a natural science and a human science, then it is reasonable to regard it as a discipline in need of *both* quantitative *and* qualitative methods in the '*schaffen*-ing' of *both* nomothetic *and* idiographic knowledge about individuals. Due to the long-standing insistence that psychology could qualify as scientific only to the extent that it '*schaffens*' its *Wissen* by means of the quantitative and experimental methods on the model of natural science (cf. Williams & Robinson, 2016), the discipline has long been vastly more constrained than it could be, needs to be, and, in fact, once was.

A Concluding Comment

Mark Twain (1835–1910) once observed that, often, the troubles that folks encounter arise less from what they don't know than from what they know for certain that 'just ain't so.' This is perhaps the best explanation for mainstream psychology's incorrigibility heretofore in the face repeated admonitions that statistical knowledge about populations cannot be interpreted validly as knowledge of the individuals within those populations. Mainstream convictions to the contrary have been as firm and as widespread as they have been mistaken, and, as a consequence, the critiques of long-standing interpretive practices have been neither refuted nor respected. They have simply been ignored, and the false interpretive practices have continued unchanged.[20]

My hope for this volume is that, at long last, it will prompt the needed paradigmatic changes in this so egregiously unhealthy intellectual state of affairs, that the fundamental difference between psycho-demographic and psychological studies will finally be recognized, and that the psychological studies can be rededicated to their original task: advancing scientific accounts of the psychological 'doings'—sensations, perceptions, judgments, emotions, cognitions, memories, and behaviors—of individuals.

Notes

1. Behaviorists in the intellectual patrimony of J. B. Watson (1878–1958) and B. F. Skinner (1904–1990) were the one notable group of experimental psychologists who eschewed this transition.
2. Proctor and Xiong (2018) have challenged the aptness of my characterization of mainstream experimental psychology as psycho-demography (Lamiell, 2018a). For my rejoinder to their arguments, see Lamiell (2018b).
3. The study was part of a much larger program of research being conducted in the Tulsa school system, selected because of (a) that system's national reputation for funding its programs, and (b) the researchers' interest in what could be discovered about the effectiveness of early childhood educational programs administered under relatively favorable circumstances.

4. ‘Impact’ is the term that the authors used throughout their article to characterize the statistical relationships between a measure of classroom instructional quality and various indicators of children’s cognitive functioning. The use of that term reflects the investigators’ belief that the quasi-experimental design of their study entitled them to infer the causal influence of instructional quality on children’s respective levels of cognitive functioning. In application to statistical relationships between variables defined only for populations, causal inferences are inappropriate—even when those relationships are determined experimentally or quasi-experimentally—for reasons discussed in Chapter 4. However, this point is tangential to our present concerns, and so will not be pursued further here.
5. The authors’ use of this term is another clear reflection of their regard for their study as quasi-experimental, in nature.
6. The quality measure was based on in-classroom observations of ‘proximal interactions between teacher and children’ (Johnson et al., 2016, p. 2148).
7. A great many additional statistical analyses were carried out in the course of this study to investigate ancillary questions. The interested reader can learn about these by consulting the published research report.
8. Children’s cognitive functioning was assessed using three subtests of the Woodcock-Johnson achievement tests (Woodcock, McGrew, & Mather, 2001). Those subtests were described by Johnson et al. (2016) as follows: ‘The Letter-Word Identification subtest measures decoding skills by asking children to identify letters and whole words. The spelling subtest assesses prewriting and spelling skills by requiring children to draw lines, trace and produce letters, and spell words correctly. The applied problems subtest evaluates pre-numeracy skills by asking children to analyze and solve math problems with relatively simple calculations’ (Johnson et al., 2016, p. 2148).
9. Two of the three (Johnson and Phillips) were then and still are faculty members in that program. Markowitz was, at the time of the study, an advanced graduate student in the program, and after completing her psychology Ph.D. in developmental science, moved into a position as Research Professor in the Educational Policy Works Center at the University of Virginia.
10. Actually, this statement is not highly speculative on my part. Until my retirement in December of 2017, I was myself a faculty member in the Georgetown University Department of Psychology, and served as chair

of that department from 2009 to 2015. I had a great many interactions with Johnson, Markowitz, and Phillips during that time (and well before then in the case of Phillips, who chaired the Department from 2000 to 2006). As a result, I know first-hand that each of the three has theoretical interests in the psychological development of children.

11. The size of the interval around a given centering value would depend upon the stringency of the test of the hypothesis desired by the investigator. This point will be elaborated within the context of another research example discussed further on.
12. The reader may recall in this connection my criticism in Chapter 6 of the claim by Lilienfeld et al. (2013) that statistical analyses of differences between treatment group means can justify inferences about what has transpired with 'all or most individuals within the populations (investigated)' (Lilienfeld et al., 2013, p. 891, parentheses added). As the present illustration makes transparent, statements such as this one by Lilienfeld et al. (2013) grossly misrepresent what the statistical analyses commonly undertaken in treatment group experimentation can actually reveal.
13. Here Grice (2015) indicates that he had borrowed this expression from an article by Edwards, Lindman, and Savage (1963).
14. The journal is freely available at: http://www.person-research.org.
15. One of the practically countless places where these notions have been expressed is in the previously cited work by Kerlinger (1979) discussing the 'troublesome paradox' in psychology (refer to Chapter 1). As the reader may recall, that 'paradox' arises because although psychologists' theoretical interests are necessarily in individual-level phenomena, their scientific concern for generality logically has seemed to obligate them to 'hypothesize and test relations at the group or set level' (Kerlinger, 1979, p. 276, emphasis added). It is the 'group or set level' of analysis that Kerlinger (1979) understands to be 'nomothetic' (cf. Kerlinger, 1979, pp. 269–274). In exactly the ways to be explained presently, Kerlinger's (1979) discussion of the nomothetic-idiographic distinction is a textbook example of the utter confusion that dominates mainstream thinking on this matter.
16. It is possible to secure nomothetic knowledge of aggregates, i.e., to discover an empirical regularity that is common to all of the aggregates within a series of investigated aggregates. However, this would not be knowledge of what is common to all of the individuals within any of

the respective aggregates, and so could not properly be regarded as the sort of nomothetic knowledge sought within psychology.

17. For all quotations of Windelband's *History and Natural Science* (1894/1998), the indicated page numbers refer to my 1998 English translation (Windelband, 1894/1998).
18. Throughout my translation of Windelband's 1894 text, I used the term 'humanities,' a shorter and more commonly encountered expression for the 'human sciences,' where Windelband had used the German term *Geisteswissenschafen.*
19. In consideration of these points, this is perhaps as good a place as any to suggest that those who have not read Windelband (1894/1998) should refrain from speaking or writing on the nomothetic-idiographic distinction.
20. The recent response by Proctor and Xiong (2018) to Lamiell (2018a) is one welcome exception to this, for while those authors failed in their attempt to defeat the argument that most contemporary mainstream research in 'psychology' must be understood as a species of demography (cf. Lamiell, 2018b), the attempt was at least made.

References

Bakan, D. (1955). The general and the aggregate: A methodological distinction. *Perceptual and Motor Skills, 5,* 211–212.

Bakan, D. (1966). The test of significance in psychological research. *Psychological Bulletin, 66,* 423–437.

Barlow, D. H., & Nock, M. K. (2009). Why can't we be more idiographic in our research? *Perspectives on Psychological Science, 4,* 19–21.

Cronbach, L. J. (1957). The two disciplines of scientific psychology. *American Psychologist, 12,* 671–684.

Dilthey, W. (1894). Ideen über eine beschreibende und zergliedernde Psychologie [Ideas concerning a descriptive and an analytical psychology]. *Sitzungsberichte der Akademie der Wissenschaften zu Berlin.* Zweiter Halbband, 1309–1407.

Edwards, W., Lindman, H., & Savage, L. (1963). Bayesian statistical inference for psychological research. *Psychological Review, 70,* 193–242. https://doi.org/10.1037/h0044139.

Gantt, E. E., & Williams, R. N. (Eds.). 2018. *On hijacking science: Exploring the nature and consequences of overreach in psychology*. New York: Routledge.

Grice, J. W. (2011). *Observation oriented modeling: Analysis of cause in the behavioral sciences*. New York: Academic Press.

Grice, J. W. (2014). Observation oriented modeling. *Comprehensive Psychology, 3,* ISSN 2165-2228.

Grice, J. W. (2015). From means and variances to persons and patterns. *Frontiers in Psychology, 6,* 1–12.

Grice, J., Barrett, P., Cota, L., Felix, C., Taylor, Z., Garner, S., et al. (2017). Four bad habits of modern psychologists. *Behavioral sciences, 7,* 1–21. https://doi.org/10.3390/bs7030053.

Grice, J. W., Barrett, P. T., Schlimgen, L. A., & Abramson, C. I. (2012). Toward a brighter future for psychology as an observation oriented science. *Behavioral sciences, 2,* 1–22.

Harré, R. (2006). *Key thinkers in psychology*. Thousand Oaks, CA: Sage.

Johnson, A. D., Markowitz, A. J., Hill, C. J., & Phillips, D. A. (2016). Variatioin in impacts of Tulsa pre-K on cognitive development in kindergarten: The role of instructional support. *Developmental Psychology, 52,* 2145–2158.

Journal for Person-Oriented Research, 2015, Volume 1, statement of Aims and Scope. ISSN 2002-0244.

Kerlinger, F. N. (1979). *Behavioral research: A conceptual approach*. New York: Holt, Rinehart, & Winston.

Lamiell, J. T. (1981). Toward an idiothetic psychology of personality. *American Psychologist, 36,* 276–289.

Lamiell, J. T. (1998). "Nomothetic" and "idiographic": Contrasting Windelband's understanding with contemporary usage. *Theory and Psychology, 10,* 715–730.

Lamiell, J. T. (2018a, forthcoming). From psychology to psychodemography: How the adoption of population-level statistical methods transformed psychological science. *American Journal of Psychology, 131,* 471–475.

Lamiell, J. T. (2018b). Rejoinder to Proctor and Xiong. *American Journal of Psychology, 131,* 489–492.

Lamiell, J. T., & Durbeck, P. (1987). Whence cognitive prototypes in impression formation? Some empirical evidence for dialectical reasoning as a generative process. *Journal of Mind and Behavior, 8,* 223–244.

Lamiell, J. T., Foss, M. A., Larsen, R. J., & Hempel, A. (1983). Studies in intuitive personology from an idiothetic point of view: Implications for personality theory. *Journal of Personality, 51,* 438–467.

Lilienfeld, S. O., Ritschel, L. A., Lynn, S. J., Cautin, R. L., & Latzman, R. D. (2013). Why many clinical psychologists are resistant to evidence-based practice: Root causes and constructive remedies. *Clinical Psychology Review, 33,* 883–900.

Machado, A., & Silva, F. J. (2007). Toward a richer view of the scientific method: The role of conceptual analysis. *American Psychologist, 62,* 671–681. https://doi.org/10.1037/0003-066X.62.7.671.

Porter, T. M. (1986). *The rise of statistical thinking: 1820–1900*. Princeton, NJ: Princeton University Press.

Proctor, R. W., & Xiong, A. (2018). Adoption of population-level statistical methods did transform psychological science but for the better: Commentary on Lamiell (2018). *American Journal of Psychology, 131,* 483–487.

Schiff, B. (2017). *A new narrative for psychology*. Oxford, UK: Oxford University Press.

Schiff, B. (Ed.). (2018). *Situating qualitative methods in psychological science.* Abingdon, UK: Routledge.

Williams, R. N., & Robinson, D. N. (Eds.). (2016). *Scientism: The new orthodoxy*. New York: Bloomsbury Publishing.

Windelband, W. (1894/1998). History and natural science (J. T. Lamiell, Trans.). *Theory and Psychology, 8,* 6–22.

Woodcock, R. W., McGrew, K. S., & Mather, N. (2001). *Woodcock-Johnson Tests of Achievement*. Itasca, IL: Riverside.

Wundt, W. (1913/2013). Psychology's struggle for existence (J. T. Lamiell, Trans.). *History of Psychology, 16,* 195–209. https://doi.org/10.1037/0032319.

Index

J. T. Lamiell, *Psychology's Misuse of Statistics and Persistent Dismissal of its Critics*, Palgrave Studies in the Theory and History of Psychology,
https://doi.org/10.1007/978-3-030-12131-0