Development and Psychopathology, 9 (1997), 193-229

The Hubble hypothesis and the developmentalist’s dilemma

JOHN E. RICHTERS
National Institute of Mental Health

Abstract
Developmental psychopathology stands poised at the close of the 20th century on the horns of a major scientific dilemma. The essence of this dilemma lies in the contrast between its heuristically rich open system concepts on the one hand, and the closed system paradigm it adopted from mainstream psychology for investigating those models on the other. Many of the research methods, assessment strategies, and data analytic models of psychology’s paradigm are predicated on closed system assumptions and explanatory models. Thus, they are fundamentally inadequate for studying humans, who are unparalleled among open systems in their wide ranging capacities for equifinal and multifinal functioning. Developmental psychopathology faces two challenges in successfully negotiating the developmentalist’s dilemma. The first lies in recognizing how the current paradigm encourages research practices that are antithetical to developmental principles, yet continue to flourish. I argue that the developmentalist’s dilemma is sustained by long standing, mutually enabling weaknesses in the paradigm’s discovery methods and scientific standards. These interdependent weaknesses function like a distorted lens on the research process by variously sustaining the illusion of theoretical progress, obscuring the need for fundamental reforms, and both constraining and misguiding reform efforts. An understanding of how these influences arise and take their toll provides a foundation and rationale for engaging the second challenge. The essence of this challenge will be finding ways to resolve the developmentalist’s dilemma outside the constraints of the existing paradigm by developing indigenous research strategies, methods, and standards with fidelity to the complexity of developmental phenomena.

Introduction

Within minutes of receiving their first data transmission from the Hubble Space Telescope in 1990, earthbound astronomers knew that something had gone terribly wrong. The long-awaited launching of the Hubble had taken place several months earlier following decades of planning, instrumentation design, and construction. This was the eagerly anticipated, first, and flagship mission of the National Aeronautics and Space Administration’s Great Observatories program, and astronomy’s expectations were high; the Hubble mission had been touted in advance as the greatest stride in astronomy since Galileo’s telescope spied the moons of Jupiter (New York *Times*, 1990). Its scientific objective was literally to look backward in time and provide astronomers with high-resolution images of the universe never before seen-images of galaxies that ceased existing billions of years ago. To capture these images, the Hubble had been equipped with the most sophisticated and expensive optical system ever developed and launched into orbit 200 miles above the distorting influence of the earth’s atmosphere. Despite its technological sophistication the Hubble was transmitting blurred images of the universe back to earth that were no clearer than those produced by ground-based telescopes. As diagnostic tests would soon reveal, the source of the problem was a spherical aberration deep in the Hubble’s intricate optical

The views expressed in this paper are those of the author and are not to be interpreted as official views of the U.S. Department of Health and Human Services, the National Institutes of Health, or the National Institute of Mental Health.

Address correspondence and reprint requests to: Dr. J. E. Richters, Child & Adolescent Disorder Branch, Room l8C-17, NIMH, 5600 Fishers Lane, Rockville, MD 20857. E-mail: jrichter@NIH.gov.

system. Human error had resulted in the Hubble's 94-in. primary mirror being ground to the wrong prescription, rendering it nearsighted. The Hubble, in effect, had been launched into orbit on the most significant and expensive exploratory mission in the history of astronomy with the wrong pair of glasses.

The exacting details of the spherical aberration diagnosis provided the necessary foundation for an intensive program of recovery efforts that culminated 42 months later in a complete restoration of the Hubble's vision. The Hubble recovery effort is best remembered for its televised images of the final dramatic moments in which astronauts floating in space lowered a corrective optics package into Hubble while circling the earth at 20,000 miles/hr. The defining scientific moment of the Hubble rescue, however, took place far from public view when astronomers first recognized that something was wrong. Although it passed routinely and without fanfare, the moment was decisive to the ultimate success of the Hubble rescue in the sense that a misjudgment at that early juncture could have resulted in a radically different course of action and outcome. A misinterpretation of the Hubble data as accurate, for example, could have resulted in a protracted period of scientific activity focused on the wrong phenomena. Astronomers, in effect, would have been studying illusory phenomena actually created by the Hubble's lens flaw in the mistaken belief that they were studying inherent complexities of the structures and motions of distant galaxies. Alternatively, the Hubble mission might have been compromised indefinitely if the distortion had been recognized but inadequately understood, or adequately understood but deemed not correctable at its source. Any number of missteps in the process of recognizing and diagnosing the Hubble's vision problem could have resulted in a prolonged diversion of scientific resources away from the main objectives of the Hubble exploratory mission.

It is difficult to imagine that astronomers were ever at serious risk for these alternative scenarios. The Hubble scientists were operating on the basis of a well established, advanced knowledge base and their scientific questions were carefully anchored in and designed to extend that base. In addition, benefitting from advances in physics, telemetry, optics, and communications, they were relying on sophisticated instruments capable of yielding sufficiently precise and detailed observations to address those questions. The specifications for Hubble's optical system, for example, called for its primary mirror to be ground to a smoothness which, if scaled to the width of the continental United States, would result in no mountain or valley varying from the mean surface by more than 2 1/2 in. It is hardly surprising that scientists working at this level of detail would be exquisitely sensitive to the presence and meaning of discrepancies between their expectations and observations.

Psychology's Progress as a Science

It would be an understatement to say that the science of psychology does not yet possess anything of comparable precision in its theories, research instruments, or methods. Nor is it yet able to proceed with surefooted confidence in its ability to discriminate successfully between facts and artifacts, flawed theory and flawed data, real and illusory phenomena. To be sure, psychology has accomplished much during its first full century as a science. Its research activities have uncovered a wealth of previously unknown facts about myriad aspects of human sensation, perception, learning, language, cognition, emotion, motivation, personality, and social functioning and development (Gerstein, Luce, Smelser, & Sperlich, 1988). Moreover, these facts, in combination with the language, concepts, and theories of psychology have had an immeasurable transforming influence on how individuals in Western society think about themselves and others. Despite these gains, it is generally acknowledged that psychology has fallen far short of its own expectations as well as society's for scientific progress in the 20th century. In commenting on widespread public perceptions of psychology's practical scientific contributions to society, psychologist-researcher Ronald Fox reluctantly offered the following assessment in his 1994 presidential address before the American Psychological Association (Fox, 1996, p. 779):

> We have not succeeded in demonstrating the public benefit that is derived from investments in psychological research. Social and behavioral sciences research funding has been under attack for several years now. From my discussions with legislators and government officials, it is clear that one of the reasons for the attacks is the belief among many of those who allocate the funds that considerable money has been invested in our research with little to show for it in terms of practical help for society's problems.

Fox and many other psychologists find such assessments unduly harsh and overstated because they give insufficient weight to psychology's genuine contributions. Nonetheless, most also admit that those contributions pale in comparison to the expectations and enthusiasm of the field following World War II (Sarason, 1981). Even more important, there have been growing concerns within psychology's scientific community about its difficulties in establishing sustained theoretical progress in many areas of research. In a sobering assessment almost two decades ago, Meehl characterized the natural history of theories in so-called soft psychology as a sequence beginning with initial enthusiasm, followed by repeated attempts to demonstrate a theory's merits, disillusionment about negative findings, growing bafflement about inconsistent and unreplicated results, and multiple resort to ad hoc excuses. In the end. according to Meehl, researchers eventually lose interest and move on to other ideas as their theories "... suffer the fate that General MacArthur ascribed to old generals-They never die. they just slowly fade away" (Meehl, 1978, p. 807). Indeed. ideas and theories that are widely accepted and embraced by one generation of psychologists are routinely discarded or ignored by the next. Even within a given generation, dramatic findings that promise to move the knowledge base forward frequently have a half-life of only a few years before they fade from memory. Despite its accomplishments, psychology has not yet produced anything comparable to the established theories and laws of the natural sciences in terms of their explanatory power or capacity to yield precise and reliable predictions.

Observations about psychology's comparatively slow progress are often met with reminders about its scientific youth and the complexities of its mission. Not only is psychology considerably younger as a formal discipline than most established sciences, but many scholars agree that psychological phenomena are inherently more complex and challenging to study than most phenomena in the physical and biological sciences. Psychological functioning and behavior is a product of the most complex and sophisticated organ known to science; the human brain is actually designed by nature to write and revise its software across the life span (Lykken, 1991). Moreover, for a variety of obvious practical and ethical reasons psychology is disadvantaged in comparison to many of the natural sciences (astronomy not included) by not being able to experimentally manipulate some of its most important phenomena. This is especially true of research domains once classified by Cronbach as "correlational" psychology, not necessarily in the statistical sense, but because they must rely on observing, organizing, and analyzing phenomena as they are found in nature (Cronbach, 1957). Furthermore, psychology has been disadvantaged by not being able to capitalize on the advances and technological accomplishments of physical and biological sciences, and has thus been hampered for much of its history by limitations in instrumentation and measurement technology. Finally. humans are reactive in powerful, complex, and not well understood ways to the knowledge or suspicion that they are being studied. In contrast to the objects of interest to the physical sciences, humans sometimes misunderstand, lie, forget, distort, confabulate, mischaracterize, misjudge, project. and deny in reaction to being questioned and/or observed by researchers.

Persistent paradigm problems

There have been other more troubling voices over the years calling attention to fundamental inadequacies in psychology's scientific approach to these challenges. In his recent attempt to answer the frequently asked question "What's wrong with psychology anyway?" David Lykken concluded " ... psychology isn't doing very well as a scientific discipline, and something seems to be wrong somewhere . . . due in part to a defect in our research tradition . . . " (Lykken 1991. p. 4). Lykken is one of many distinguished scholars who have argued that psychology's basic problem for more than a century has been the absence of a coherent, agreed upon paradigm[1] for defining the boundaries, objectives, phenomena, and methodology of its science. As philosopher of science Ernest Nagel noted almost 40 years ago, the social and behavioral sciences generally lack anything approaching the almost complete unanimity commonly found in the natural sciences about " . . . what are matters of established fact, what are the reasonably satisfactory explanations (if any) for the assumed facts, and what are some of the valid procedures in sound inquiry" (Nagel, 1961, p. 448).

Thomas Kuhn, in his classic analysis of scientific revolutions, described the natural history of sciences in terms of a cyclical pattern of stages or periods (Kuhn, 1970). Each science begins in a preparadigm period characterized by the proliferation of competing schools of thought, a diversity of fact gathering activities, and minimal cumulative theoretical progress. Eventually, when a particular viewpoint or school ascends to dominance, it serves as the shared paradigm for a period of normal science characterized by the relative absence (or silence) of competing schools, intensive coordinated research activity focused on solving puzzles defined by the paradigm, and resulting theoretical progress. Initially, as anomalies emerge during periods of normal science, they are either shelved for a period or dismissed as unimportant as scientists concentrate on puzzles that can be solved successfully within the paradigm. Over time, however, as the balance between successful problem solving and important unexplained anomalies shifts. insecurity eventually sets in about the viability of the paradigm. This, in turn, gives rise to a loosening of the paradigm's restrictions, an active consideration of alternatives, and the reemergence of competing schools of thought. When a replacement paradigm is finally adopted, the process begins again with a new period of normal science.

Although Kuhn's thesis about the revolutionary nature of paradigm replacement was vigorously challenged and eventually softened by Kuhn himself (Kuhn, 1977), his basic distinction between periods of normal and crisis science is a useful one. So also is his description of the initial preparadigm period in fledgling fields of science, such as chemistry before Boyle, astronomy before Copernicus, electricity before Franklin, and physical optics before Newton. Prior to the Copernican revolution, for example, astronomers found it necessary to make so many ad hoc adjustments to the Ptolemaic system that its complexity eventually far exceeded its accuracy. Similarly, prior to Newton, there was no generally accepted view about the nature of light. There were numerous competing schools of thought based on one or another variant of Epicurean, Aristotelian, or Platonic theory, each able to explain particular aspects of light phenomena but not others. Each also claimed that the phenomena it could best explain were preeminently significant, offering ad hoc explanations for-or dismissing the theoretical significance of-those it could not (Kuhn, 1970).

The parallels between these early sciences and late 20th century psychology have not escaped the attention of scholars. Kuhn himself was drawn initially to his study of scientific paradigms by a curiosity about why, in contrast to fields such as physics, astronomy, and biology, disciplines like psychology seemed perpetually engaged in controversies about the " . . . fundamentals of their science" (Kuhn, 1970, p. viii). Nagel in a strikingly similar but earlier observation about psychology, noted with puzzlement that in most fields of science disagreements over fundamentals tend to occur not in the mainstream but on the

1. I use the term "paradigm" throughout this paper as a generic reference to the scientific practices and procedures characteristic of psychology's mainstream research community.

frontiers of scientific activity. Kuhn's subsequent explanation, of course, was that psychology is still in its initial preparadigm period of development. Now, almost 30 years later, psychology's reputation as a preparadigm science is so taken for granted by most scholars that the claim is commonly asserted as self-evident without elaboration (Bevan & Kessel, 1994; Horgan, 1996; Koch & Leary, 1992; Lakatos, 1977; McGuire, 1983; Robinson, 1984).

One of the defining characteristics of preparadigm science is a proliferation of ideas about exactly what is wrong and how to make it right. Psychology has been no exception. Critics dating back to its founding have differed considerably both in their explanations for its difficulties and in their proposed solutions, The common core complaint underlying many criticisms has been that psychology's founders erred strategically by modeling their science after the 19th century natural sciences. The narrow emphasis of that paradigm on passive physical objects, Humean notions of causality, reductionistic explanations, and mechanistic laws, according to critics, rendered it unsuitable for a coherent science of psychology. In contrast to the objects of interest to classical physics, human beings are living, active, interactive, reactive, and adaptive organisms. Thus, a science of psychology requires a fundamentally different approach to scientific investigation based on what Sigmund Koch would later call "indigenous methodologies" capable of reflecting and illuminating the richness of human phenomena (Bevan & Kessel, 1994; Koch, 1959, 1961).

Ironically, psychology's embrace of this paradigm took place at about the same time that the physical sciences were abandoning its limitations. The classical physics model is now recognized by the natural sciences themselves as narrowly useful only in the study of closed physical systems and under a limited range of circumstances (Brandt, 1973). In modern physics, for example, the physical world is no longer viewed as similar at all levels; there is a widespread recognition of the need to employ different conceptual strategies and explanatory models at different levels of analysis; and the earlier, narrow emphasis on efficient causality has given way to meson fields, gravitational fields, and magnetic fields, none of which requires or is expected to be accounted for by traditional models of Humean causality.

The more immediate concern of psychology's founders, though, was to establish its independence from philosophy and secure its status as a genuine field of science. And, in pursuing this objective, they were influenced strongly by a sentiment expressed by John Stuart Mill, who asserted that "... the backward state of the Moral Sciences can only be remedied by applying to them the methods of Physical Science, duly extended and generalized" (quoted in Leahey, 1987, p. 4). Thus, rather than risking an investment in the development of new methodologies, psychology embraced explanatory models, discovery methods, and research strategies that had already proven successful in the established 19th century natural sciences. The predictable and enduring consequence of this decision, according to critics, has been to limit psychology's coherence, credibility, and potential as a science (Bevan, 1991; Bevan & Kessel, 1994; Smith & Torrey, 1996).

Although many of psychology's critics have agreed in general about the inadequacies of its mechanistic foundations, there has been little consensus about a remedy. Within each succeeding generation psychologists have initiated isolated efforts to champion particular research methods or strategies aimed at overcoming one or more of the paradigm's perceived inadequacies. Occasionally, such efforts have resulted in the emergence of specialized subdisciplines and/or new scientific journals devoted to particular methods or substantive issues. In the main, however, such efforts have not translated into anything approximating the kinds of fundamental reforms that would be required to replace psychology's paradigm with one based on more indigenous methodologies and research strategies with greater fidelity to human psychological phenomena (Koch, 1959, 1961). On the contrary, psychology has become increasingly fragmented into narrowly defined subdisciplines, each pursuing its objectives in relative isolation from the others. As Copernicus described the state of astronomy before the Copernican revolution, "... it is as though an artist were to gather the hands, feet, head, and other members for his images from diverse models, each part expertly drawn, but not related to a single body, and since they in no way match each other, the result would be monster rather than man" (quoted in Kuhn, 1957, p. 83). Moreover, late 20th century psychology continues to be characterized by widespread dissatisfaction with the existing paradigm yet little consensus about how to proceed (Lykken, 1991).

Developmental Psychopathology

Some of the most integrative and progressive reform efforts in recent decades have arisen not with the field of psychology per se, but in the hybrid discipline of developmental psychopathology (Cicchetti, 1990, 1993; Cicchetti & Cohen, 1995a, 1995b; Rutter & Garmezy, 1983; Sroufe & Rutter, 1984). Historically, developmental scientists have played a prominent role in discussion of psychology's paradigm (Baldwin, 1895; Bronfenbrenner, 1979; Hinde, 1992; Kagan, 1992; Lewin, 1931a, 1931b; Magnusson, 1985; Overton & Horowitz, 199 1; Overton, 1984a; Sameroff & Chandler, 1975). Developmental psychopathology evolved as a discipline, in fact, out of a long history of progressive efforts to overcome the limited disciplinary boundaries and narrowly focused paradigms of traditional social, psychological, and biological sciences (Cicchetti, 1984). In terms of breadth and depth, it stands in sharp contrast to the increasingly narrow foci of psychology's specialty areas and subdisciplines. Its broadly defined conceptual framework accommodates, benefits from, encourages, facilitates, and contributes to integrative research across the full array of social and biological sciences in the service of a developmental understanding of human development and functioning.

Also reflecting its progressiveness, developmental psychopathology's emphases on open system concepts and organismic, holistic, organizational, and transactional approaches to research parallel the dramatic 20th century ascendancy of these frameworks throughout the natural sciences. Those who align themselves with the developmental psychopathology perspective tend to share the objective of gaining a process-level understanding of how, why, and in what ways individual differences in normal and abnormal social, emotional, cognitive, and behavioral development emerge, interact, and develop across the life span. Most also recognize the necessity of employing organizational principles and systems concepts such as equifinality, multifinality, epigenesis, and transactional influences in the pursuit of these objectives.

The first challenge in any revolution or reform movement lies in building consensus about the need for change. The second lies in finding ways to translate that consensus into action. Viewed from this perspective, developmental psychopathology has been remarkably successful in establishing the necessary foundation for reform. The heuristic value of its interdisciplinary, integrative framework is now widely recognized throughout the human sciences. During the past two decades alone it has fueled the establishment of new scientific journals, special sections in numerous existing scientific journals, and a plethora of book chapters, edited volumes, and books. In addition, its basic principles were embraced by the 1989 Institute of Medicine report *Research on Children and Adolescents with Mental, Behavioral, and Developmental Disorders* (Institute of Medicine, 1989), and in the more recent National Institute of Mental Health's 1990 *National Plan for Research on Child and Adolescent Mental Disorders* (National Advisory Mental Health Council, 1990). For all of these reasons, developmental psychopathology seems poised to define a new center of gravity for many areas of psychological science in the 21st century.

Developmentalist's dilemma

The major dilemma facing developmental psychopathology in realizing its potential lies in the contrast between its heuristically rich open system concepts and developmental models on the one hand, and the closed system paradigm on which it relies for investigating those models on the other. At present, most empirical research in developmental psychopathology is based largely on the paradigm it adopted from mainstream psychology. The research methods, assessment strategies, and data analytic models of this paradigm, however, remain wedded to the closed system assumptions and explanatory models of 19th century science. It therefore emphasizes the study of variables over individuals, the modeling of complex interactions among variables, and the search for invariant causal explanations to account for phenotypically similar patterns of overt functioning across individuals. These strategies make perfect sense for the study of within-class structurally homogenous objects functioning in closed physical systems. But they make little sense as strategies for studying human functioning and development, given that humans are unparalleled among open systems in their wide ranging capacities for equifinal and multifinal functioning (Allport, 1960; Cairns, 1986; Lewin, 193 la; Lykken, 1991; Valsiner, 1986a; von Bertalanffy, 1968).

Kurt Lewin criticized mainstream psychology's research paradigm on these grounds over 60 years ago, prior to the formal introduction of open system concepts and principles such as equifinality and multifinality (Lewin, 193la, 1931b). Research in psychology was already characterized at the time by common practices of aggregating subjects on the basis of phenotypic similarities, assessing psychological and environmental variables in isolation from each other and from their contexts, and submitting the data to group-level statistical analysis. Lewin warned that averages obtained in this way on large samples of children rendered virtually hopeless the possibility of a meaningful understanding of individual differences in human functioning: "The concepts of average child and average environment have no utility whatever for the investigation of dynamics . . . An inference from the average to the particular case is ... impossible" (Lewin, 1931b, p. 95). Gordon Allport made similar arguments three decades later with specific reference to open systems in his classic article entitled *The Open System in Personality Theory* (Allport, 1960). And, with increasing frequency since that time, senior developmental scholars have raised similar concerns, arguing for the need to develop research methods and data analytic strategies with greater fidelity to developmental phenomena (Cairns, 1986; Emde, 1994; Hinde, 1992; Hinde & Dennis, 1986; Kagan, 1992; Magnusson, 1985; Radke-Yarrow & Sherman, 1990; Sameroff & Chandler, 1975; Valsiner, 1986a).

Despite this long-standing recognition, an alternative paradigm for studying developmental phenomena has yet to emerge and gain widespread acceptance in the scientific community. Consequently, developmental psychopathologists necessarily rely for their empirical research on the established research methods and strategies of psychology's existing paradigm. This confronts them with the challenges of studying an organismic model of development with research methods and strategies based on mechanistic model. Inevitably, the incompatibilities between these models translate into a major dilemma as researchers must choose between adhering to developmental principles and frameworks *or* conforming to research practices engendered and reinforced by the existing paradigm.

A review of contemporary developmental psychopathology research reveals that the developmentalist's dilemma is often resolved in favor of existing methods. Developmental principles embraced in the Introduction sections of empirical manuscripts are routinely ignored and/or violated in Methods and Results sections, only to resurface in Discussion sections. There is often clear evidence of developmental sophistication in the thinking leading to and following empirical research. But much of this sophistication is sacrificed in the research process itself as developmental principles and frameworks are transduced into closed system research strategies. Thus, with the exception of new theoretical constructs, greater emphasis on psychometric sophistication, and more elaborate data analytic models, the contemporary empirical literature of developmental psychopathology continues to reflect many of the same practices lamented by Lewin over 60 years ago. Subjects who are recruited on the basis of background factors and/or phenotypic similarities in behavior are presumed a priori to be homogenous with respect to the causal factors underlying their functioning; isolated parcels of disembodied information are collected about each subject; and data analytic strategies that presuppose within-sample structural homogeneity are employed in the service of identifying a unitary causal structure to account for individual differences in the phenomena of interest.

Although these widespread practices are contraindicated by the impressive human capacities for equifinal and multifinal functioning, they are perpetuated by the illusion of progress created by the paradigm's weak scientific standards. Those standards have been criticized harshly by numerous statisticians, psychologists, and philosophers of science as pseudoscientific and misleading (Meehl, 1990). Although these flawed standards are seldom discussed in the developmental literature, they play a powerful role in blinding researchers to the indispensable negative feedback about their theories and ideas that is the self-correcting essence of scientific progress. Those flawed standards, in turn, are virtually required by inadequacies in the paradigm's closed system methods and practices. I argue below that an examination of these interdependent weaknesses reveals much about why the paradigm requires fundamental reforms and why those reforms have not been undertaken.

Overview

The remainder of this paper is devoted to an examination of the developmentalist's dilemma and a consideration of its implications for paradigm reform in developmental psychopathology research. My point of departure is the premise that mutually enabling weaknesses in the discovery and justification processes (Reichenbach, 1938) function like a distorted lens on the research process by variously sustaining the illusion of theoretical progress, obscuring the need for fundamental reforms, and both constraining and misguiding reform efforts. These undermining influences are certainly not unique to the discipline of developmental psychopathology; they are evident across the broad array of social, behavioral, and psychological sciences that rely for their empirical research on psychology's paradigm. Unlike many other disciplines, however, developmental psychopathology faces the genuine dilemma of having explicitly embraced open system principles that are fundamentally incompatible with the paradigm's closed system assumptions. Developmental psychopathology is actually at odds with itself in the empirical arena by attempting to study a complex open system with strategies and methods of scientific discovery based on mechanistic assumptions.

Developmental psychopathology faces two challenges in successfully negotiating the developmentalist's dilemma. The first lies in recognizing how and under what conditions the current paradigm encourages research practices that are antithetical to sound science and the developmental agenda. The second challenge lies in developing and adopting strategies for overcoming these inadequacies and for translating open system principles and developmental frameworks into equally sophisticated research designs, methods, and practices.

I begin by focusing on some of the fundamental inadequacies of the existing discovery paradigm for studying human functioning from an open system perspective. I then present a basic critique of the justification paradigm, summarizing the interrelated statistical, logical, and conceptual weaknesses for which it has been so harshly criticized. I make no attempt to cover the already extensive critical literatures concerning either set of paradigm weaknesses. My purpose rather is to establish enough of a foundation for examining how they enable and obscure each other's undermining influences in the practical context of day to day research. I conclude by considering the implications of this analysis for undertaking a program of constructive reforms.

Psychology's Paradigmatic Lenses

The problem of infidelity between developmental psychopathology's heuristics and traditional research methods can be conceptualized within Reichenbach's classic framework as arising in the scientific *context of discovery*

(Reichenbach, 1938). These are the activities of a science that occupy the initial phases of the research process, ranging from basic decisions about what kinds of ideas and phenomena are considered worth studying, to decisions about how they are studied and analyzed. The context of discovery, in brief, concerns how scientists go about generating their ideas and evidence. The *context of justification,* in contrast, concerns how scientists go about evaluating those ideas. Thus, justification activities occur later in the research process and include the procedures and decision processes through which a science evaluates the merits of its ideas.

Historically, discovery and justification issues have been discussed as separate paradigm problems in psychology, by different authors, for different audiences, and often in different publication outlets. Discovery issues always have figured prominently in the mainstream literature of psychology because of their obvious direct relevance to substantive questions about the conceptual frameworks, theories, and research methods in particular research domains. These issues of method are almost never discussed, however, with reference to the justification process. Typically, they are viewed as problems that emerge and can be remedied in the context of discovery.

Justification discussions have tended to accumulate outside the literatures of particular substantive domains in more general forums. Like discovery issues, however, psychology's scientific standards tend to be discussed as abstract conceptual problems that emerge and are to be resolved in the context of justification. They are almost never discussed with reference to specific processes of discovery. Even those who have discussed both types of issues have tended to discuss them as relatively independent problems. Meehl's classic 1978 critique of the slow progress of so-called soft psychology, for example, is half devoted to cogent summary discussions of 15 complex discovery issues and problems facing psychology and half to issues of justification. Neither set of issues, however, is discussed with reference to or within the context of the other (Meehl, 1978). In a similarly thoughtful, more recent analysis of psychology's paradigm problems Lykken also discussed discovery and justification issues in relative isolation from one another (Lykken, 1991).

Although discovery and justification activities are conceptually different and temporally separated in the research process, they are not independent. They are interdependent in the sense that decisions and activities in each context inevitably anticipate those in the other. Discovery decisions about how to conceptualize phenomena, frame research questions, design studies, choose variables, assess subjects, and analyze data, necessarily anticipate the logic, standards, and procedures by which they will be evaluated later in the context of justification. Similarly, the coherence of the justification process hinges decisively on its fidelity and goodness of fit with the assumptions, strategies, and procedures of the discovery process. While this interdependence is true in all areas of science, it is an especially powerful and negative one in psychology because it exists as a dynamic between *weaknesses* in the discovery and justification processes.

In the latest in his series of penetrating critiques of psychology's justification practices, Paul Meehl commented, "This is my last attempt to call attention to a methodological problem of our field that I insist is not minor but of grave import" (Meehl, 1990, p. 108). Meehl's lament is a common one among psychology's discovery and justification critics, who have expressed varying degrees of frustration, disbelief, puzzlement, disappointment, and concern about why long-standing weaknesses in the paradigm have not given rise to concentrated reform efforts. I argue below that the answer lies in the mutually enabling interdependence between weaknesses in the paradigm's discovery and justification processes. Each obscures the inadequacies of the other and imposes constraints on attempted solutions. Combined, these weaknesses function like a distorted lens in the research enterprise by sustaining the illusion of progress, obscuring the need for fundamental paradigm reforms, and both constraining and misdirecting isolated reform efforts. This is a longstanding pressing problem throughout the psychological sciences. But it constitutes a genuine dilemma for the discipline of developmental psychopathology because of its explicit commitment to open system concepts that are incompatible with psychology's closed system paradigm.

Context of discovery

Mechanistic and organismic world views. Every scientific discovery process is framed and shaped by a series of initial decisions about which phenomena are worth studying, what their essential natures are, how they should be conceptualized, and what strategies, procedures, and methods should be used to study them. Many of these decisions, however, are based on background assumptions that are so implicit that the researcher is often unaware that he or she is making consequential choices. They seem more like obvious choices than like important decisions over which the investigator has control. Pepper characterized scholarly approaches to knowledge in terms of major world hypotheses, each of which is associated with a different set of core philosophical assumptions about the nature of the physical world and thus how it can and should be studied (Pepper, 1942). The mechanism and organicism world views are particularly relevant for present purposes because of their central significance to most critiques of psychology's discovery paradigm.

During the 19th century the dominant world view in the physical sciences was mechanism. Newtonian physics viewed the universe as consisting of passive physical objects that interacted with one another in clockwork fashion, with energy as the causal link between objects. The task of science was therefore to study individual objects, identify regularities among patterns of actions and interactions between them, and determine the general laws that accounted for them. To understand the complex workings of any given whole or system or machine, according to this view, one need only break it down into its individual, uniform components, examine the characteristics of each in isolation from the other, and add them together. The whole, from a mechanistic viewpoint, is nothing more than an additive sum of its individual parts.

Organicism, which has gained increasing dominance during the 20th century, takes as its root metaphor the integrated, living organism. It emphasizes holism and emergence beyond mere additive complexity; active, teleological predilection beyond passivity and reactivity; dynamic beyond static systems; and structural differentiation and integration over time. Although organicism allows for and can accommodate mechanistic phenomena, scholars have long debated whether the reverse is true-that is, whether a mechanistic model is inherently incapable of accommodating or explaining many of the essential characteristics of human functioning such as growth, adaptiveness, development, intentionality, structural reorganization, and other characteristics of living organisms (Nagel, 1961). It is not necessary to take a stand on this issue one way or another, however, to appreciate the practical barriers to studying organismic phenomena within a mechanistic framework.

Scientific hard cores. The importance of worldviews lies in the powerful central role they play in guiding, facilitating, and/or constraining the activities of a science. Overton, elaborating on the significance of worldview assumptions and commitments, has argued that they function as an essential part of the *hard core* that serves as the foundation for scientific programs (Lakatos, 1977; Laudan, 1977; Overton, 1984a). The defining characteristic of hard cores is that they are shielded from potential falsification by the research program's *positive heuristic* and *belt of auxiliary hypotheses* (Lakatos, 1977). The positive heuristic includes suggestions and guidelines for examining, challenging, and changing only the refutable (i.e., non-hard-core) features of the research program and for modifying and refining the protective belt of auxiliary hypotheses. Thus, only the particular ideas, theories, hypotheses, methods, and research strategies of a research program are subject to scrutiny, challenge, and potential falsification. The hard core itself is taken for granted and remains relatively insulated. Within the hard core of every science, for example, are the assumptions that its phenomena are rule governed and that the relevant laws are, in principle, discoverable. Thus, the failures of a science may be attributed to errors in particular theories, methods, procedures, and strategies, or even to intrinsic complexities in its phenomena, but not to weaknesses in its hard-core assumptions. The power of the hard core was best expressed by Laudan's characterization of research traditions as "... a set of ontological and methodological do's and don't's ... To attempt what is forbidden by the metaphysics and methodology of a research tradition is to put oneself outside the tradition and to repudiate it" (Laudan, 1977, p. 80).

The basic conceptual and philosophical criticisms of a strict mechanistic approach to studying human functioning have been discussed extensively in the literature of psychology (e.g., Altman & Rogoff, 1987; Lewin, 193la, 193lb; Nagel, 1961; Overton & Horowitz, 1991; Pepper, 1942). So also have the comparative merits of transactional, dialectic, and organizational variants of organismic approaches (Lerner & Kaufman, 1985; Overton & Horowitz, 1991). I concentrate below on some of the practical limitations of the mechanistic model and advantages of the organismic model for thinking about and conducting day to day research in psychological science, My point of departure is a question raised by Kagan et al. almost 20 years ago in reaction to the statement, "The interactions between biological and environmental forces determine the psychological growth of the organism: What in heaven's name does that fourteen word sentence mean?" (Kagan et al., 1978, p. 44). Indeed, descriptions of this type are often so general that they could mean almost anything and therefore communicate very little. I will attempt to elaborate on some specific implications of an organismic approach to human functioning.

Closed versus open systems. One of the essential differences between mechanistic and organismic models lies in the distinction between closed and open physical systems (von Bertalanffy, 1968). All nonliving physical matter falls within the category of closed systems, which are arrangements of physical entities connected in ways that constitute a unity or whole, and are completely isolated from their external environment. That is, the entities in a closed system interact only with one another and are not subject to outside interactions or influences of any kind (Popper, 1991). Because they exchange neither energy nor matter with the external environment, closed systems are subject to the second law of thermodynamics. At rest, they tend toward the state of highest probability (equilibrium), maximal disorder, and the disappearance of existing differentiations in structure. Strictly speaking, the only truly closed physical system is the universe itself, because it presumably has no external environment. As a practical matter, however, nonorganic physical systems can be conceptualized along a continuum of closedness as a function of their relative degrees of isolation from external factors (e.g., temperature, force). Although only quasi-closed in this sense, they are nonetheless commonly referred to as closed systems.

The metaphor most closely associated with closed systems is the machine, characterized by fixed patterns of mechanical reactions and interactions among its parts, governed by universal laws, producing fixed end states. To the extent that a closed system is completely isolated from external factors, its end states are determined by initial conditions. The final concentrations resulting from chemical equilibrium in a closed system, for example, are completely determined by initial conditions, such that a change or modification (whether manipulated or observed) in those conditions or intervening processes will produce predictable changes in the final concentrations (von Bertalanffy, 1968). Similarly, with adequate knowledge of a completely closed system, it is possible to infer backward from a given state to its initial conditions.

Equifinality in open systems. All living organisms fall into the category of open systems, which exchange energy and matter with the environment and are not therefore subject to entropy. Beyond thermodynamic openness, living organisms are distinguished from closed systems by unique functioning characteristics such as growth, adaptiveness,

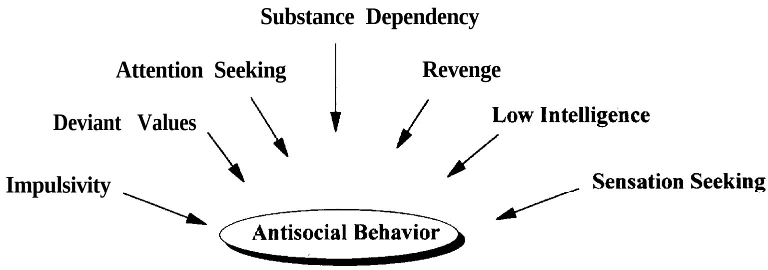

Figure 1. Equifinality in human functioning as an illustration of different personality structures and processes that might give rise equifinally to persistent antisocial behavior in different individuals, depending on a matrix of other internal and external factors.

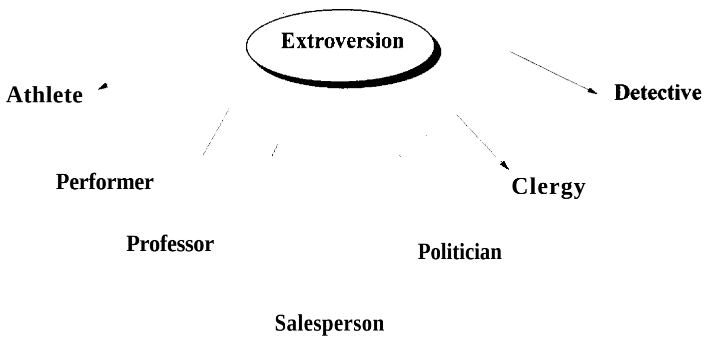

Figure 2. Multifinality in human functioning as an illustration of different overt behavior patterns that might arise multifinally from a common characteristic such as extroversion, depending on a matrix of other factors internal and external to an individual.

regeneration, emergence, differentiation, dynamic self-stabilization, and self-reorganization. Organisms are also distinguished from closed systems by the functioning capacities of equifinality and multifinality. Equifinality is the ability to reach similar outcomes or end states from different starting points and through different processes (Figure 1). Conversely, multifinality (Figure 2) is the ability to reach different outcomes from similar starting points and/or through similar processes (von Bertalanffy, 1968). Thus, in contrast to closed systems, the initial conditions of open systems imply nothing about their end states and their end states imply nothing about their initial conditions.

Equifinality was observed and labeled originally in the biological sciences following discoveries by Spemann and others that cellular tissues can be transplanted successfully from one area of presumptive growth to another (Spemann, 1938). During critical periods of growth and maturation, for example, tissues from the neural plate of amphibia can be transplanted successfully to areas of limb growth, where they take on characteristics of muscle or skin tissue instead of brain tissue. Similarly, in embryology, it was demonstrated that living organisms (end states) could develop from ova that were whole, divided, or fused (Waddington, 1957). Even in ontogeny, equifinality is evident in the fact that organisms often overcome-make up for-temporary delays in physical and physiological maturation by nonetheless reaching similar end states. The underlying principle of equifinality in biological systems is teleological in the sense that the end states of the organism are understood to be genetically precoded. Moreover, because the elements of a living organism are reciprocally related and mutually regulating, their complex organization allows for the achievement of different end states from different starting points and through different processes.

Equifinality is not restricted to teleological biological processes. As von Bertalanffy noted, it is also found in psychological and sociological systems. In an early and elegant demonstration of fail-safe behavioral systems. for example, psychobiologists demonstrated that the ability of female rats to retrieve their pups a-as not dependent on the stimuli of any single sensory system (Beach & Jaynes, 1956; Herrenkohl & Rosenberg, 1977). Regardless of whether their tactile, olfactory, auditory, or visual sensory channels were destroyed, and even when rendered anosmic, their ability to retrieve their pups was preserved by the ability to function equifinally.

It is in this broader sense that equifinality is often used in developmental psychopathology to refer to different structural/causal processes underlying similar overt patterns or syndromes of cognition, emotion, behavior, and/or psychopathology (Cicchetti & Richters, 1993; Cicchetti & Rogosch, 1996; Richters & Cicchetti, 1993a, 1993b). In this broader usage, equifinality refers to a defining characteristic of human functioning so ubiquitous and widely recognized that it is simply taken for granted in social discourse and behavior.

Individuals vary considerably in their beliefs, values, attitudes. preferences, temperaments, intelligence(s), personality characteristics, experiential histories. current circumstances, and anticipated futures. These differences, in turn, can give rise to very similar patterns of overt behavior and functioning for different reasons (equifinality) as well as very different overt behaviors for similar reasons (multifinality). There are virtually unlimited different ways in which any given feeling, desire, or objective (e.g., anger. love. empathy, fear, greed, anxiety, sadness. happiness, survival) may be expressed or acted upon (multifinality). Conversely. very different feelings, desires, and objectives may be expressed through similar overt patterns of functioning (equifinality). For example. individuals may contribute generously to charities primarily for tax write-off purposes, to enhance their reputations in the community (for good purposes or bad), because they are deeply committed to a charity's objectives, or more generally to the principle of helping others (equifinality). Not everyone who is committed to helping others, however, contributes generously to charities. Depending on a matrix of other internal and external factors, they may instead participate in church activities, volunteer their time to charitable organizations, run for elective office, join lobbying efforts, or merely act compassionately toward others in their daily lives (multifinality).

Equifinality and multifinality also have significant implications for thinking about the so-called mental disorders of childhood. Consider, for example, the common childhood syndrome of attention deficit hyperactivity disorder (ADHD), codified in the Diagnostic and Statistical Manual of Mental Disorders, 4th edition (American Psychiatric Association, 1994). The diagnosis of ADHD is based strictly on reports about a child's patterns of inattention, hyperactivity, distractability, and so forth, not on a process-level assay of the causes or structures responsible for those behaviors. Thus, although children who meet ADHD diagnostic criteria share in common a general constellation of overt functioning deficits, there is no reason a priori to believe that they stem from a common etiology or underlying structure. Moreover, there are sound conceptual, theoretical, and clinical reasons for believing that the population of children who meet ADHD criteria is heterogeneous with respect to the relevant causal factors. Different subsets of children diagnosed with ADHD may have different kinds of inherited, biologically acquired, and/or environmentally acquired diatheses giving rise to the overt syndrome. Others may be responding primarily to environmental stressors in the absence of an underlying diathesis. And yet for others the symptoms may be products of diathesis/ circumstance interactions.

Equifinality and structure. Note that the extension of equifinality beyond biological structures per se to human psychological and social functioning introduces an additional element of complexity. In biology, there are considerable limits to the extent that different end states can be accomplished by different structures. The functions of the heart, for example, cannot be performed by the liver.

Moreover, hearts, livers, and kidneys perform the same functions in all humans. In psychological and social functioning, however, there are very few constraints on the ability of different processes and structures to compensate for and/or facilitate each other in the service of end states.

The digital computer provides a useful analogy for thinking about equifinality in terms of the structures underlying overt human behavior (Elasser, 1966; Lykken, 1991). All functions carried out by a digital computer stem from the same unchanging hardware configuration. Each time a software program is loaded to perform a particular task such as bookkeeping, word processing, statistical analyses, or graphics, however, the computer undergoes elaborate structural modifications. It would be impossible therefore to understand the computer's overt functioning in any of these domains by studying the hardware itself. Each program constitutes a different underlying structure which operates in different ways and according to different algorithms in the service of accomplishing different tasks. Moreover, the same tasks may be accomplished in different ways on the same computer (e.g., SPSS vs. SAS statistical packages). Thus, understanding the computer's overt functioning r quires a consideration of the software from which it arises. Human brain structures, of course, constitute the biological equivalent of computer hardware. Moreover, unless we are Cartesian dualists, we must assume that all human psychological functioning and behavior is ultimately a product of the brain.

Nonetheless, for good reasons, most of our theories do not focus directly on biochemical or electrophysiological processes of the brain. Rather, they concern higher level processes and structures of personality, temperament, intelligence, and so forth that constitute the structural equivalents of software programs. These are the emergent characteristics of the human nervous system's extraordinary plasticity and capacity for structural modifiability at every level of organization from molecular to molar. These higher order structures, in turn, emerge and develop through each individual's unique combination of inheritance, learning, and experience. It is precisely this modifiability that allows individuals to achieve similar overt functions through different processes (equifinality) and different overt functions through similar processes (multifinality). Obviously, then, there may be structurally different variants of many of these higher order characteristics, as well as qualitatively different configurations among them, to consider in approaching the study of overt functioning.[2]

Structure in closed versus open systems. The basic business of every science is to describe the nature of its entities and to develop theories about their structures from which their functioning can be deduced. The coherence of these activities rests on the success of a science in organizing its entities into classes that are sufficiently homogenous in theory-relevant structure that they can be considered interchangeable for theoretical purposes (Elasser, 1966). Quantum theory, for example, absolutely depends on the assumption of within-class homogeneity. In fact, it is impossible to formulate the mathematics of quantum mechanics without the axiom that the members of each class of particle (e.g., electrons, protons) are rigorously alike and interchangeable (Elasser, 1966). Similarly, classical physics requires that its objects are sufficiently homogenous in this respect that within-class irregularities can be averaged out statistically for purposes of prediction and explanation. It is also this structural homogeneity that makes it coherent to pursue a single theory to account for the actions and interactions among phenomena within a closed system.

The extraordinary human capacities for equifinal and multifinal functioning, however,

2. The limitations of the computer software analogy are apparent here. Although different word-processing programs (e.g., Word and WordPerfect) may function similarly based on different underlying structures (e.g., algorithms, subroutines, etc.), they are nonetheless designed to accomplish the same objectives. In the case of human functioning. however, similar patterns of surface functioning may not only stem from different underlying structures (like the computer), but may be in the service of different goals or purposes (unlike the computer).

render the structural homogeneity assumption untenable. Very similar patterns of overt functioning may be caused by qualitatively different underlying structures both within the same individual at different points in time, and across different individuals at the same time (equifinality). Conversely, different patterns of overt functioning may stem from very similar processes within the same individual over time, and across different individuals at the same time (multifinality) (Cicchetti & Rogosch, 1996; von Bertalanffy, 1968). These structural differences in turn, have tremendous implications for the psychological stimulus properties of stressors, events, and experiences, the causal relevance of which may vary qualitatively across individuals. In a closed system consisting of within-class structurally homogeneous objects there is warrant for assuming that the influence of a given variable will vary only quantitatively across members of the class. In the realm of human functioning, however, there is no basis for this assumption. The meaning, salience, and causal impact of a given event or experience can vary *qualitatively* across individuals as a function of their differences in genetic endowment, experiential history, resulting internal structures, and current circumstances. As the Stoic philosopher Epictetus observed "Men are disturbed not by things, but by the views which they take of them." From an open system standpoint, these different "views" are a function of qualitative differences in internal structures.

The distinction between closed and open systems reveals the sense in which psychology's approach to discovery can be characterized and criticized as mechanistic. It is certainly not the case that most psychologists endorse a mechanistic model of human psychological functioning. Most, in fact, would reject it vigorously as an anachronism. Nonetheless, the mechanistic model is built into psychology's paradigm in the sense that each major step in the discovery process makes sense only as part of an overall strategy for studying objects that are within-class structurally homogeneous in relatively closed physical systems. That is, the discovery process itself is framed as a search for universal laws to account for individual differences in the development and functioning of individuals who are presumed to be structurally homogenous. I argue below that the homogeneity assumption is evident throughout the discovery process in the selection and recruitment of subjects, the selection and assessment of variables, and in the analysis and interpretation of data (see Interdependence Problem section). Although this assumption is seldom justified and is often contraindicated by common sense, it has been protected for decades by an overly forgiving justification process.

Context of justification

Traditionally, scientific justification has been the chief concern of logicians, philosophers, and historians of science, notably because of its significance to the so-called problem of demarcation between science and pseudoscience (Lakatos, 1977). From a philosophy of science standpoint, it is irrelevant how scientists arrive at their ideas or observations within the context of discovery. Standardized research methods and procedures are indispensable, of course, for the coherence of a scientific community's activities and communications. But they are of little concern to the philosopher of science because the merits of all ideas, regardless of how they are arrived at in the context of discovery, must be evaluated finally in the context of justification. The scientific enterprise is distinguished from other ways of fixing beliefs about the world, such as various forms of dogmatism and pseudoscience, by the rigor with which it questions and evaluates the ideas it discovers. Thus, criticisms about the scientific standards of a discipline are tantamount to questions about its legitimacy as a field of science.

It is therefore especially noteworthy and troubling that psychology stands alone among recognized sciences in being criticized harshly for its justification practices, notably for weaknesses in the logic, methods, and standards by which it evaluates theoretical claims. The distinguished Nobel laureate physicist Richard Feynman was quite explicit in characterizing psychology as a "Cargo Cult" pseudoscience on these grounds by comparing its scientific practices to the superstitious rituals of South Sea islanders following the second World War. Having witnessed military aircraft landing on their island with materials and supplies during the war, they engaged in elaborate rituals using bars of bamboo in place of antennas, and building wooden huts to look like control towers, all in a vain effort to attract the planes again. Like the South Sea islanders, according to Feynman, many psychologists and other social scientists follow only the apparent precepts and forms of scientific investigation. But, because they are missing something essential, they suffer the same fate: "... their planes don't land" (Feynman, 1986, p. 179).

As psychology's justification processes have come into question over the decades, they have been discussed most often by and for those most familiar with the highly specialized languages and concepts of logic, philosophy of science, and statistics. Although the essential criticisms are rather straightforward and compelling, their impact potential has been limited by three related factors. First, the issues often have been discussed in technical terms that make them seem variously inaccessible, esoteric, and/or of little practical consequence. It is not uncommon, for example, to encounter complex formulas of symbolic logic and difficult statistical notations in justification discussions (Meehl, 1990; Rozeboom, 1982). To the uninitiated, justification issues often seem indistinguishable from technical discussions about the relative merits of applying a Bonferroni versus Scheffe statistical correction procedure under a given set of circumstances. Second, with some notable and welcome recent exceptions, justification issues often have accumulated outside the mainstream literature of psychology in specialized books, chapters, and journals. This probably has reinforced the misperception that, whatever the technical merits of justification arguments, they have little practical import for the day to day activities of the practicing scientist. As Meehl insisted almost 20 years ago, however, "I am not making some nit-picking statistician's correction. I am saying that the whole business is so radically defective as to be scientifically pointless" (Meehl, 1978, p. 26). Third, justification problems characteristically have been discussed in relative isolation from specific issues of discovery, further obscuring their practical implications.

Scientific conventions of theory evaluation. In most developed sciences, theories are evaluated by their explanatory power and ability to make predictions of phenomena that are both precise and novel. Precision typically is indexed by the degree of accuracy with which a theory can make point predictions of phenomena within a very narrow range of specificity. Relatedly, novelty is indexed by a theory's ability to anticipate and predict facts and observations that would be unlikely absent the theory. In Salmon's classic phrase, a theory is considered impressive to the extent that its accurately predicted observations would be otherwise explainable only as a *damn strange coincidence* (Salmon, 1984). Following from this principle, when different theories offer competing explanations for the same phenomena, the one that generates the most accurate predictions and anticipates the most novel outcomes generally is considered-at least for the time-the better theory. Thus, in general, theories achieve their status by having survived what Popper referred to as grave risks of refutation in the empirical arena (Popper, 1959).

Theory testing in psychology. In most areas of psychology neither precision nor novelty are required to justify theoretical claims. Instead, they are evaluated through formalized procedures of hypothesis testing, the outcomes of which are determined by a null-hypothesis test of statistical significance. Theories in most nonexperimental areas of psychology are of a sufficiently general nature that they predict and purport to explain that theory-relevant variables will not be random with respect to each other and will be positively or negatively correlated in the population of interest. That is. they yield directional rather than point predictions of group differences or correlation values. such that a theory positing a causal influence of variable *A* on variable *B* will be tested by determining whether the variables are at all correlated in the direction (above or below 0) predicted. Thus, the observed correlation between *A* and *B* will be tested against the null hypothesis (*Ho*) of no relationship-literally, a covariation value of zero. By convention, if the computed correlation is greater than zero, *Ho* is rejected in favor of the investigator's theory. Conversely, if the correlation is not greater than zero the investigator failed to reject *Ho* and instead rejects the theory.

Null-hypothesis significance tests. An absolute comparison of the *AB* correlation with zero would be straightforward if it were based on all subjects in the population of interest and if the variables were assessed with perfect precision. The fact that it is based on only a sample drawn from that population, however, raises the possibility that it may be an artifact of sampling bias. That is, even in a population in which two variables are randomly distributed with respect to each other, any particular sample drawn from that population may nonetheless reflect a correlation greater than zero. Thus, to guard against the risk of erroneously concluding in favor of the theory on the basis of such an artifactual correlation, a test of statistical significance is conducted using *Ho* as the standard of comparison. The basis for this test, simply stated, is the fact that repeated samples drawn from a population in which variables are truly random with respect to each other will nonetheless yield a normal distribution of correlation coefficients above and below zero. This hypothetical distribution provides the basis for estimating the probability of detecting a nonzero correlation in a sample drawn from a zero-correlation population. To protect against making such a Type 1 error-rejecting *Ho* when it is true-a relatively low probability value (usually $p < .05$ or .01) is adopted. Thus, the sample based correlation is considered *statistically significant*-and therefore supportive of the theory-if the probability of detecting a correlation of similar or greater magnitude from a zero-correlation population would be very unlikely.

It has been estimated that over 90% of empirical articles published in psychology's leading scientific journals rely on null-hypothesis testing as the sole means for inferring conclusions from research data (Loftus, 1993). Estes recently reported the proportion of empirical studies presenting significance tests in current issues of leading psychology journals as follows: *Memory and Cognition* (100%), *British Quarterly Journal of Experimental Psychology* (100%) *Journal of Abnormal Psychology* (100%) and *Journal of Personality and Social Psychology (88%)* (Estes, in press). Despite its ubiquity, the ways in which null-hypothesis testing is relied upon in psychology has been criticized by numerous statisticians, philosophers of science, and psychologists as the equivalent of no scientific standard at all for evaluating substantive theories. It warrants emphasis that null-hypothesis significance testing per se is recognized as a valuable, necessary, and elegant tool for answering particular types of questions under certain specified conditions (Hagen, 1997). These virtues do not, however, extend to most of the applications for which psychology relies on statistical significance testing. Specifically, it is psychology's widespread reliance on significance testing as a *theoretical* tool that has been harshly criticized (Cohen, 1990, 1994; Gigerenzer et al., 1989; Glymour, 1980; Loftus, 1991; Lykken, 1968, 1991; Meehl, 1978, 1990; Morrison & Henkel, 1970).

Null hypothesis is quasi always false. The first criticism is that in the social, behavioral, and biological sciences, the null hypothesis is almost never true (Bakan, 1966; Edwards, 1965)! Or, in Meehl's terms, it is "... quasi-always false . due to the empirical fact that everything is correlated with everything else, more or less" (Meehl, 1990, p. 123). This phenomenon has been referred to both as the *crud factor* in psychology (Lykken, 1991), and, less pejoratively, as psychology's *ambient correlational background* (Meehl, 1990). The rather obvious general explanation is that it probably reflects complex, multidetermined causal relationships among genetic linkage as

3. These criticisms typically have not been extended to technological uses of statistical significance testing such as clinical trials, which more closely approximate the applications and conditions for which it was developed (Meehl. 1990).

well as autocatalytic processes of affect, cognition, and behavior both within and between individuals (Meehl, 1990). Whatever its sources, the troubling empirical fact is that nonzero correlations (both positive and negative) have been demonstrated routinely between randomly chosen variables in large data sets (Evans & McConnell, 1941; Meehl, 1967, 1990; Lykken, 1991; Thurstone, 1938). Thus, a theory's directional predictions stand a 50/50 chance of being confirmed even if the theory has absolutely no merit, assuming sufficient statistical power to meet criteria for statistical significance!

Meehl's paradox. The combination of directional predictions and the so-called crud factor creates an additional problem, dubbed *Meehl's Paradox* (Meehl, 1990; Leventhal, 1992; Lykken, 1991). In developed sciences, which emphasize point predictions within very narrow boundaries of error, increases in measurement precision create more difficult hurdles for theories by increasing sensitivity to deviations between theoretically predicted and observed values. In psychology, where the predictions are only directional, increases in precision actually do the reverse. That is, they *lower* the evidentiary hurdle by increasing the likelihood of falsely rejecting H_0 (Type 1 error) and capitalizing on ambient background correlations that are irrelevant to the theory. The assumption-supported by considerable data-that H_0 is quasi always false means that most variables will be correlated to some degree for reasons that have nothing to do with the theory being explored. In many areas of physical science this is a relatively minor consideration for three reasons. First, background correlations are much less likely to exist. Second, the physical sciences are much better able to isolate many of their basic phenomena in laboratory settings and manipulate them in ways that reasonably ensure the absence of artifactual correlations. Third, and most important, the physical scientist typically is making a rather precise point prediction within a very narrow range of error from his/her theory. Thus, to the extent that measurement errors can be reduced through greater precision, it becomes easier to detect true (nonartifactual) differences or discrepancies between predictions and observations.

Everything is reversed in psychological research. Background correlations are ubiquitous in psychology generally, and particularly in uncontrolled nonlaboratory settings. Moreover, in contrast to the physical scientist's point prediction, a directional prediction allows for any value (or group difference) other than absolute zero as confirmatory evidence as long as it is in the right direction. Therefore, an increase in measurement precision will increase rather than decrease the likelihood of capitalizing on theory-irrelevant associations and therefore lower rather than raise the evidentiary hurdle a theory must jump in the justification process. Relatedly, the statistical power for detecting a correlation as statistically significant also increases with sample size. Thus, literally any deviation from H_0 in the population will result in its rejection if the sample is large enough; the larger the sample the higher the risk of committing a Type 1 error! As I argue below, the frequently recommended replacement standard of effect sizes is often an inadequate solution to this problem (see Justification in *context* below).

Statistical versus theoretical hypotheses. A somewhat separate problem is that relying on a test of significance to evaluate the merits of a theory confuses the researcher's statistical hypothesis with his/her theoretical hypothesis, and relatedly conflates the very different concepts of statistical significance and theoretical support. The H_0 and alternative hypothesis in a test of statistical significance (e.g., above example) are statistical hypotheses concerning a population parameter. That is, the purpose of the significance test is solely to aid the investigator in drawing an inference about an unobserved population parameter from the actual value(s) computed on an observed sample. The test has nothing whatever to do with the merits of any particular theoretical explanation for *why* the value exists. The theoretical hypothesis, on the other hand, concerns a relation between the theory and patterns in the sample data. Thus, whereas the statistical significance test is a way of using the sample to draw an inference about the *population,* the correlation is the investigator's method of drawing an inference from the sample to the *theory.* The significance test, in short, is not in any way a test of the merits of the substantive theory. Meehl explains this difference succinctly by telling students-" .. assume you had the parameter; what would you know, and-how confidently?' (Meehl, 1990, p. 117).

Misinterpretations of p values. Even those who recognize that the significance test refers only a link between the sample and population nonetheless often misinterpret even its statistical meaning, usually in one of three ways. First, the *p* of a significance test is often interpreted as the probability that the observed sample-based correlation was due to chance. Second, it is often referred to as the probability of detecting a correlation of similar or greater magnitude from another sample drawn from the same population. The third, most egregious, and pervasive misinterpretation is that $1 - p$ value (e.g., $1 - .05 = .95$) is the probability that the alternative (i.e., investigator's) hypothesis is correct. Oakes recently estimated that approximately 95% of academic psychologists believe this latter interpretation. None of these three interpretations is accurate! The *p* value is nothing more than the probability of detecting a correlation (or group difference) of the same or greater magnitude in a sample randomly drawn from a zero-correlation (or zero group difference) population. It is hardly surprising that the statistical significance test is so commonly misinterpreted and misused in psychological research. Erroneous interpretations have been taught in some of psychology's classic, most widely used and respected textbooks on measurement and statistics (Anastasi, 1958, p. 11; Ferguson, 1959, p. 133; Guilford, 1942, pp. 156-166; Nunnally, 1975, pp. 194-196). Moreover, for more than a decade, the 2nd edition of the official *Publication Manual of the American Psychological Association* misled psychologists by informing them that data trends failing to meet the usual levels of significance " are best interpreted as caused by chance and are best reported as such" (American Psychological Association, 1974, p. 19).

Logical fallacy. Weak uses of significance tests to evaluate theories are further compounded by a conceptual and logical error in reasoning that has nothing to do with statistics. The error stems in the first place from psychology's emphasis on empirical corroboration to judge the merits of theories. In the developed sciences, theories are evaluated by submitting them to grave risks of refutation in the empirical arena (Popper, 1959). The logic of refutation conforms to two forms of argument that are deductively valid from a premise. If, for example, theory *T* holds that variable X causes variable Y under specified conditions, the premise is that *T* entails or implies the existence of *0* under those conditions *(T:0).* Following deductively from this premise, *modus ponens* holds that if theory *T* is true, observation 0 must be true *(T :. 0).* Also following deductively from the original premise, *modus tollens* holds that if observation 0 is not true, theory *T* is not true (*T* :. *O*). Psychology's reliance on significance testing to evaluate theories relies instead on an *invalid* form of argument known as affirming the consequent (*0* :. *T*). That is, it relies on the fallacy of concluding from *(T :. 0)* that, if observation *0* is true, theory *T* is true. Obviously, this interpretation rests on the faulty assumption that the presence of children's antisocial behavior necessarily implies a history of exposure to marital distress. Since numerous plausible competing theories can be invoked to account for and successfully predict the same directional association,

4. Although sufficient for present purposes, this is a gross oversimplification of theory testing conventions for two reasons. First, the core of a theory is challenged in the face of disconfirming evidence only if the implicit *ceteris paribus* (i.e.. all things being equal) clause can be assumed to hold--which is virtually impossible to establish in most domains of psychology citing Meehl (1990). Second, theories are not rejected on the basis of a single test of experiment. but on the basis of their long-term performance over a range of tests within the context of a research program (Lakatos, 1977).

the correlation itself implies nothing about the merits of a particular theory. Again, referring back to the distinction between statistical and theoretical hypotheses, a rejection of H_0 is technically (but only when used properly) a correct application of modus tollens concerning the statistical hypothesis of no difference. The conclusion that a confirmed prediction, directional or point, in any way corroborates the theory purporting to explain the association is a simple error in logic (Rorer, 1991).

Summary

Alone, any one of these weaknesses would be sufficiently consequential to compromise the integrity of psychology's justification proc cesses. Collectively, they have been characterized variously as feckless, impoverished, weak, illogical, misleading, a terrible mistake, and even one of the worst things to happen in the history of psychology (Meehl, 1978). Jacob Cohen, one of social and behavioral sciences' leading statisticians, characterized mainstream research psychologists as being ". . mesmerized by a single all-purpose, mechanized 'objective' ritual in which we convert numbers into other numbers . . We have come to neglect close scrutiny of where the numbers come from" (Cohen, 1990, p. 1310). Echoing similar concerns, criminologist Michael Maltz recently criticized the widespread misuse of null-hypothesis testing in psychology and allied disciplines as being ". . in all too many cases a meaningless exercise" (Maltz, 1994). Geoffrey Loftus, upon being appointed as editor of the journal Memory *and Cognition,* characterized the practices as "...at best severely limited, and at worst highly misleading" (Loftus, 1993, p. l), a form of "tyranny" and a "virtually barren technique" that lies ".. at the heart of much malaise in psychological research" (Loftus, 1991, p. 1). Imre Lakatos, one of the 20th century's leading philosophers of science, was even more stinging in his assessment, criticizing psychology as a largely ad hoc enterprise resulting in a research literature of disparate and often conflicting findings "cancerous growth' amounting to "intellectual pollution" (Lakatos, 1978, p. 89). Surprisingly, there has been little evidence in psychology's contemporary scientific journals or ongoing research activities that the majority of research psychologists are even aware of these criticisms of its standards.

The Interdependence Problem

The ways in which discovery and justification weaknesses interact to undermine the coherence and potential of research are most easily understood in the context of day to day research. In the discussion below 1 consider the case of a prototypical study in which the investigator is interested in testing a theory that assigns a causal role to marital distress in the development of children's antisocial behavior. Thus, the theory both predicts and purports to explain an association between these two variables. In the more common multivariate case a theory might predict such an effect only under certain mediating conditions. Although the issues and arguments below are easier to illustrate with the more simple bivariate case, they apply with equal force to the multivariate case. The research begins in the context of discovery with decisions about study design, sample selection, data collection, and data analysis. Following this, the interpretability and inferential relevance of these data to the investigator's theory will be evaluated in the context of justification.

Study design

Note that the initial decision to conduct research in the service of testing a theory already reflects an influence of the anticipated justification process. That process is narrowly defined in terms of a hypotheticodeductive model of theory evaluation in which the investigator begins with a theory, deduces a testable hypothesis, and collects theory-relevant data to be analyzed and summarized in the service of testing the theory. The paradigm makes a clear distinction between the functions of discovery and justification, and values discovery activities only to the extent that they are conducted in the service of testing a theory. An investigator interested instead in a more creative endeavor of exploratory, descriptive research regarding some aspect of marital distress, children's antisocial behavior, or their links, would suffer from the outset. The likelihood of research funding from most sources for such descriptive/exploratory work is relatively low because the same peer review justification standards used to evaluate the empirical worth of a study for later journal publication will be applied to the initial funding proposal as well. It often will be evaluated as weak and/or nonscientific to the extent that it is not designed in the service of a hypothetico-deductive theory test. Realizing this, the investigator will embed-and sometimes mask-any exploratory/descriptive goals within the framework of a traditional test of a theory. The sensitivity of the justification process to statistical issues of psychometrics, variable to subject ratios and statistical power, however, will impose considerable constraints on the nature and depth of whatever exploratory research is conducted within the theory-testing framework of the study design.

Sample selection

The investigator's first practical task will be to recruit a sample of subjects from the population of interest. In the study of a closed system an investigator samples the objects of interest on the basis of their known or presumed within-class structural homogeneity. Most samples in psychological research, however, are composed either of self-selected individuals of unknown heterogeneity in their underlying structures or randomly selected individuals recruited from a specified population of interest to the researcher. Ironically, although random samples are valued for purposes of generalization, they are almost by definition more likely than a typical convenience sample to be structurally heterogeneous with respect to a given phenomenon.

The prototypical risk factor study, however, will proceed on the implicit assumption that, with respect to the theory-relevant structures under study, individuals recruited into the sample are homogeneous. That is, it will be assumed a priori that they are interchangeable members of a class who differ quantitatively but not qualitatively with respect to theory-relevant structures and variables. If they are further subclassified at some point in the study, the reclassification typically will be based again on their phenotypic similarities, not with reference to possible differences in the presumed underlying structures responsible for those similarities.

There may be specific empirical or theoretical reasons for distinguishing subtypes of subjects on the basis of socioeconomic status, age, sex, or background characteristics. But in the absence of a specific rationale for such distinctions the default assumption will be that members of the sample are structurally interchangeable. Most of the emphasis in recruitment will be placed on securing a sample with sufficient phenotypic variability in the variables of interest to warrant a reasonable test of the study hypothesis. Once subjects are recruited on the basis of phenotypic similarities, every subsequent step in the discovery and justification processes will be based necessarily on the simplifying assumption of structural homogeneity across subjects. If this assumption is rejected, there is no rationale or justification for including them in the same sample for purposes of exploring a model or theory about a unitary underlying structure.

Data collection

The structural homogeneity assumption also simplifies the data collection process. That is, it justifies (a) assessing the same variables with respect to each subject, (b) assessing each variable in isolation from the others, and (c) employing the same standardized data collection procedures across all subjects. For reasons explained below, it also *requires* practices (a) and (c) but not (b). These strategies are consistent with the study of a closed system because within-class structural homogeneity implies that any influence of a variable or assessment procedure will be uniform, differing only quantitatively across members of the class. To the extent that members of the class/sample are structurally heterogeneous, however, this threefold standardization strategy may have opposite effects.

Variable selection. Structural heterogeneity, for example, entails the considerable possibility that a variable such as marital distress may be a causal factor for antisocial behavior in some children, an effect for others, a spurious correlate for others, and unrelated for yet others. Variables such as peer rejection, attributional bias, family disharmony, parenting practices, parent symptomatology, and so forth, are all obvious examples of variables that may play any of these qualitatively different roles with respect to a given outcome variable of interest as a function of structural differences across members of a sample. Unless the study design includes a deliberate attempt to identify such differences and classify them into different samples, the coherence of assessing exposure to marital distress across all subjects absolutely requires the assumption of structural homogeneity (see also Inference subsection). Note that the justification process is again necessarily anticipated at this point in the selection of a restricted set of variables to assess. The statistical requirements of the justification process will restrict this set to a sufficiently low ratio of variables to subjects to minimize the risk of committing a Type 1 error. In light of the possibility of structural heterogeneity within the sample, the risk of committing Type 2 errors is at least as great. But these latter risks will not figure prominently in the investigator's choice of variables because structural homogeneity is assumed. The emphasis, therefore, will be on assessing only those variables that are directly relevant to understanding the possible causal role of marital distress on antisocial behavior, which is assumed to hold or not hold uniformly across the entire sample.

Standardization. Similarly, the principle of using standardized measurement and assessment procedures in research is necessary for obvious reasons. In a sample of structurally homogenous individuals this is easily accomplished by adhering rigorously to the same methods and procedures for each subject. But, to the extent that members of the sample differ qualitatively in their measurement-relevant structures, it may be necessary to administer different methods and procedures to different subjects in order to achieve true standardization. This may seem counterintuitive at first, but it is merely an extension of the commonsense principle that requires altering assessment procedures to accommodate individual differences across subjects in language, age, grade level, cultural context, and so forth. Obviously it would make no sense to require the standardized administration of oral interviews in English to subjects who are non-English speaking, deaf, or too young to comprehend the meaning of the interview. These factors reflect important structural differences across individuals that are easily recognized on the basis of overt functioning and are therefore routinely accommodated in the standardization process. Equally significant structural differences across subjects that are not easily recognizable on the surface may have major influences on their interpretations of questions, response patterns, thresholds for responding, attributions about the purpose of questions, etc. To the extent that they are not obvious on the basis of overt functioning, however, the standardization process will proceed on the simplifying assumption that any such differences are essentially random and will therefore cancel each other out across the sample.

Data analysis

When assessments are completed on all subjects, the resulting data typically will be examined for patterns of covariation between and among variables across individuals in the sample. Those patterns in turn will be interpreted with reference to the presumed common underlying structure(s). Again, on the assumption of within-class theory-relevant structural homogeneity, the attempt to account for patterns of covariation with a single theory makes perfect sense. To the extent that members of the sample are structurally heterogeneous, however, a sample-based analysis of covariation makes no sense for addressing theoretical questions about causal structure. To understand why, it is necessary to distinguish carefully between the descriptive and inferential meanings of a correlation coefficient.

Description. Assume, for example, that the investigator has computed a correlation of $r = .34$ across the sample between individual differences in marital distress and children's levels of antisocial behavior. What could the correlation mean descriptively? One possibility is that the variables covary slightly but relatively uniformly across all children in the sample. Another is that they covary considerably for a small subset of children and not at all for others. There are numerous other descriptive possibilities depending of course on how the two variables are distributed across the sample in both univariate and bivariate (and multivariate, in the more complex case) space, each of which would (or should) in one way or another place absolute constraints on any inferential interpretation. Some of these patterns (e.g., outliers) can be detected by specialized statistical techniques, whereas others require visual inspection of the bivariate or multivariate scatterplot. Let us assume, however, a textbook situation in which the covariation pattern is fairly evenly distributed across cases. Descriptively, we would be correct in saying that increments in levels of marital distress are associated with increments in levels of child antisocial behavior. The crucial question, of course, is what might the correlation mean inferentially?

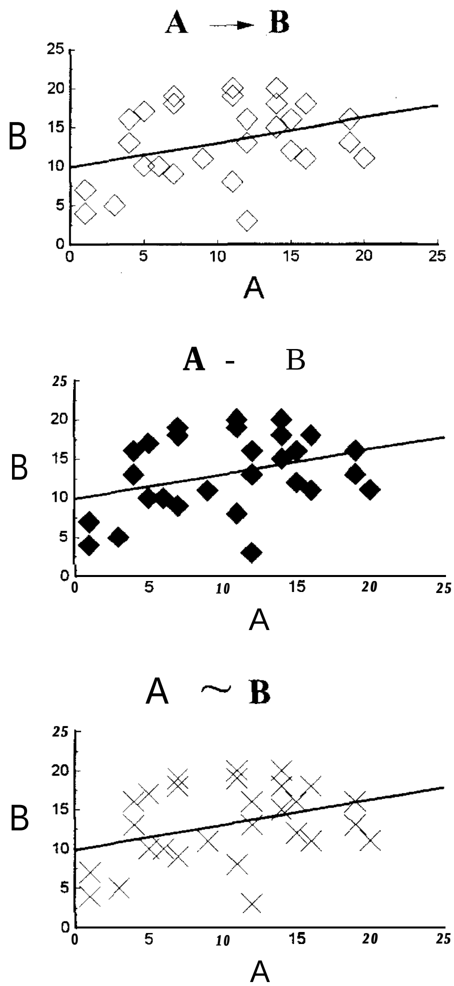

Figure 3. Three competing models of explanation for a correlation of .34 between factors A and B in a simple closed system.

Inference. In a closed system the inferential possibilities for a .34 correlation between variables A and B would be a small influence of A on B, a small influence of B on A, or an influence of some unmeasured variable or variables (including measurement artifacts) on both A and B (Figure 3).[5]

Assuming that the values of these variables were observed rather than experimentally manipulated, any one of these models might be operating in a closed system. Importantly, on the assumption of structural homogeneity among the objects across which the variables were assessed, these are competing models in the sense that only one can be correct. In an open system, however, equifinality raises the possibility that different structural models might be operating for different children or subsets of children in the sample. The relevant causal structures for some children may include a diathesis for impulsivity (inherited or acquired), poor self-esteem, a violent home environment and/or community, and minimal support or nurturance from caregivers. For others, the relevant structures may be positive characteristics such as high intelligence and resourcefulness in conjunction with conditions of poverty, poor parent monitoring, and/or a criminogenic neighborhood. And yet others for whom none of these factors is relevant may have been socialized into a deviant value

5. In principle, of course, evidence that B *preceded* A would rule out the A → B model. In practice, however, this is exceedingly more difficult to establish in longitudinal, sample-based individual differences research than most empirical reports would suggest.

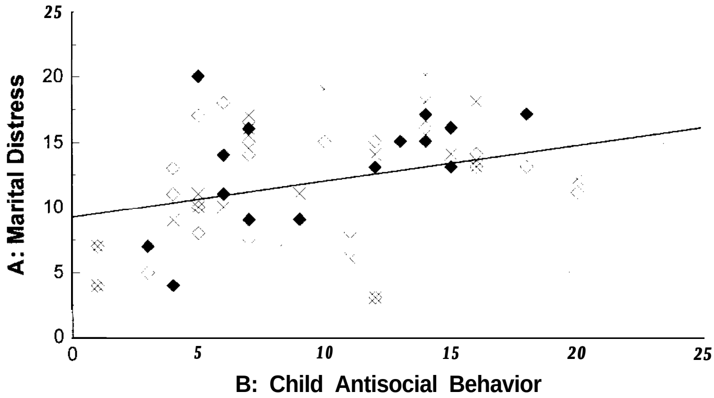

Figure 4. A hypothetical data set in which the .34 correlation between marital distress and child antisocial behavior arising from theory-relevant structural heterogeneity within the samnle. The symbols reflect children for whom marital distress is a cause, effect, or spurious correlate of antisocial behavior.

system, or may be coping defensively in a hostile, dangerous environment.

Obviously, any number of qualitatively different causal structures may give rise to similar patterns of antisocial behavior in different children (Figure 4). Moreover, one can easily imagine how marital distress might be a cause, consequence, or spurious correlate within structurally different variants of these subtypes. Thus, in addition to the three possibilities (cause, consequence, spurious correlate) within a closed system, there are potentially unlimited combinations of possibilities for the role of marital distress in a structurally heterogeneous sample of children. Marital distress may be a powerful influence on antisocial behavior for a subset of children, a consequence of antisocial behavior for another, and yet a spurious correlate for others. Artifacts resulting from unrecognized heterogeneity may work against the investigator in two ways. First, they may engulf and mask a theoretically important causal link between marital distress and antisocial behavior that holds for only a subset of children. Second, for the same reasons, they may give rise to an illusory correlation across the sample even if marital distress plays no causal role for even a subset of children. Thus, inferences about the theoretical meaning of correlations both observed and not observed in structurally heterogeneous samples may be equally treacherous!

Justification

Recall (see Statistical versus theoretical hypotheses subsection) that the investigator enters the justification process with two hypotheses, one statistical, the other theoretical. From the investigator's standpoint the important statistical question is-what is the probability that the sample-based correlation of .34 between marital distress and child antisocial behavior exists in the population of interest. Following convention, the observed correlation will be tested against the H_0 of no association in the population by conducting a statistical significance test. Assuming a sufficient sample size to achieve statistical significance, the investigator will reject H_0 and thus fail to reject the statistical hypothesis of .34. This outcome, however, does not address the investigator's question about the population value. The significance level is an index of the probability of detecting a correlation of .34 or greater in a sample randomly drawn from a population in which the variables are unrelated *(H_0).* Although it implies *nothing* about what the correlation value might be in the population, this is most often how the investigator will interpret a statistically significant correlation. For purposes of argument, however, let us assume following Meehl that the investigator has computed the correlation not on a sample but on the entire population. Thus, the statistical question of generalizability disappears and the only question remaining is what does the correlation mean with reference to the investigator's theory?

The purpose of the study of course was to evaluate a theory-based hypothesis about the influence of marital distress on children's antisocial behavior. There is no statistical test for evaluating a substantive hypothesis per se. Thus, the correlation itself (with the requirement of statistical significance if it is sample based) typically will be accepted as support for the investigator's theory. Again, although this inference is based on a logical fallacy of affirming the consequent (i.e., $T \supset 0 \therefore 0 \supset T$), it will be accepted as long as the theory is intuitively plausible and does not contradict either a widely accepted premise or an established empirical fact. If the existing knowledge base is sufficiently advanced the investigator might be required to demonstrate that the correlation still holds while controlling statistically for other variables. Moreover, in a highly developed research domain the evidentiary burden may be increased such that the investigator also must demonstrate that his or her theory accounts as well or better for the observed correlation(s) than an alternative theoretical model. Such a direct test of competing theoretical models on the same data set is quite consistent with conventions of theory testing in the established sciences.

But the logic of covariance procedures and competing model justification strategies rests on the assumption of within-sample structural homogeneity. Moreover, this is a substantive theoretical requirement not to be confused with a statistical requirement concerning the distributional properties of a variable. Many widely used statistical procedures are known to be quite robust to violations of their distributional assumptions. Thus, statistical requirements are commonly viewed as idealizations that can be ignored with impunity for most practical purposes, and researchers are able to proceed on the basis of simplifying assumptions that may or may not be true. Theoretical interpretations of sample-based data, however, are *not* robust to violations of the structural homogeneity assumption. They absolutely depend on the assumption that all subjects share the same theory-relevant structures, and that theory-irrelevant differences among them are relatively small and nonsystematic (Valsiner, 1986b). Moreover, theoretical inferences are not merely weakened or qualified but *invalidated* by violations of the homogeneity assumption.

Thus, to justify interpretations of correlations, partial correlations, semipartial correlations, interaction terms, and the like, an investigator must assume that a single explanatory model is sufficient ultimately to account for individual differences in the outcome measure across all members of the sample. To the extent that subjects differ *qualitatively* and not merely quantitatively with respect to the structure(s) of theoretical interest, however, everything changes. Not only is there no justification for employing any of these strategies on a structurally heterogeneous sample, but if they are applied the results are likely to be misleading in unpredictable ways. By definition, *any* deviation from structural homogeneity means that a minimum of two explanatory models will be required to account for the data. Moreover, artifactual data patterns introduced by undetected structural heterogeneity may favor or disfavor one or both theories regardless of their merits! Thus, even if both theories are absolutely wrong, a correlation of .34 between marital distress and antisocial behavior may nonetheless arise from artifactual (e.g., spurious or reverse causal) associations resulting from the undetected presence of structurally different subsets of subjects in the sample. Conversely, if one of the theories is correct, these same artifacts can as easily mask the causally relevant association that holds for a subset of cases. Given the small association indexed by a .34 correlation (i.e., 12% shared variance), it would take only a few structurally different cases in a typical sample to produce either of these untoward and undetectable effects. Note that the two data columns from the investigator's bivariate scatterplot presented in Figure 5 each include cases for which marital distress is a cause, consequence, and spurious correlate. The slope of the regression line passing through these columns reflects the optimal least-squares trajectory based on an average of their scores on each variable.

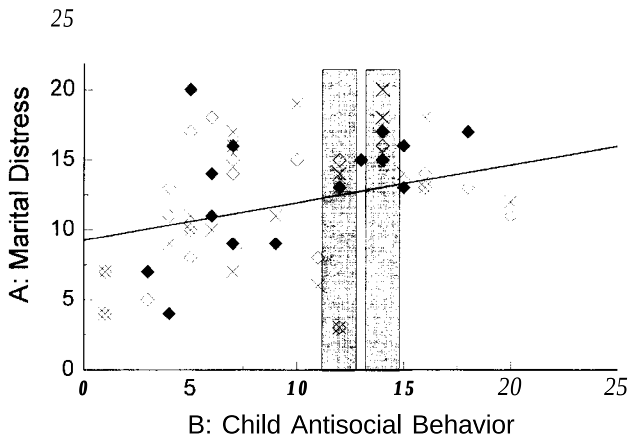

Figure 5. An enlarged illustration of structural heterogeneity in two data columns from the bivariate scatterplot Figure 4.

In computing this solution. however, the algorithm obviously is sensitive only to the quantitative properties of these variables; all scores are treated as qualitatively the same. The logic of the investigator's interpretation of this correlation, however, is supremely dependent on the assumption that the data points do not represent qualitative differences. Obviously, the least-squares solution based on children characterized by the marital distress as causal structure is far different from the solution based on all column scores. Thus, the .34 correlation is a terribly misleading index of the association in which the investigator is really interested. Suppose, however, that the investigator took seriously the possibility of structural heterogeneity as reflected in the scatterplot data in Figure 5. A moment's reflection will reveal that the likelihood of being able to discriminate between structurally different cases with such similar scores in the middle of a distribution is exceedingly low.

Constrained Solutions

Historically, the tendency of critics has been to consider psychology's discovery and justification weaknesses in relative isolation from one another. Consequently, proposed solutions to paradigm problems in one context often have suffered from either ignoring or being overly constrained by unrecognized weaknesses in the other. An understanding of why these approaches are inadequate provides a foundation and rationale for reaching beyond the constraints of the current paradigm for solutions to the developmentalist's dilemma.

Justification in context

Isolated considerations of the justification problem have resulted in calls to abandon null-hypothesis testing in favor of more rigorous standards for evaluating theoretical propositions. The most commonly recommended replacement standard has been to require more precise and "risky" theoretical predictions, usually indexed by narrower confidence intervals or larger effect sizes. These recommendations seem reasonable as solutions to an isolated problem of justification. Within the practical constraints of the current discovery paradigm, however, it is difficult to imagine how these standards might be implemented. The relatively primitive state of measurement technology in combination with the very general nature of most theoretical propositions pose formidable obstacles to implementing more rigorous standards. It isn't obvious, for example. whether a true improvement in most theories would result in an increase or decrease in the expected effect sizes of particular factors. In addition, artifacts resulting from correlated methods variance and various sources of sampling and measurement bias may overwhelm any theoretically expected correlation in many domains. Relatedly, the sheer number of variables that would be necessary to compare and test competing theoretical models on the same data set would impose formidable sample size requirements in order to maintain sufficient statistical power for minimizing Type 1 errors and maximizing sensitivity to effects.

When these recommendations are reconsidered outside the artificial constraints from a developmental perspective, however, they are not merely impractical but often misguided. Until there is a reasonable basis for the assumption that subjects in a sample are homogenous with respect to their theory-relevant structures, there is no justification for drawing any conclusions-however tentative -about a unitary underlying structure. Interpretations of regression coefficients, partial correlations, semipartial correlations, and tests of competing theoretical models are no more justified and are potentially as misleading as interpretations of simple correlations when based on structurally heterogeneous samples.

Discovery in context

Similarly, efforts to wrestle with discovery issues have been undermined by the mutually enabling interdependence between the paradigm's justification weaknesses and its hardcore structural homogeneity assumption. For example, even though the questionableness of the homogeneity assumption is easy to understand in the abstract, its practical consequences are obscured by the illusion of progress sustained by flawed justification standards. The effect of this illusion is to predefine the researcher's task as one of choosing an overt syndrome of interest, identifying the right combination of putative causal factors, and modeling complex interactions among variables in the service of characterizing the unitary underlying causal structure. Presumed homogeneity also justifies a focus on maximizing the proportion of variance explained in an outcome of interest by a fixed set of theory-relevant predictors. Within this simplified framework, problems of discovery are seen as resolvable by identifying additional predictor variables and interaction terms, and employing more sophisticated data analytic strategies to model complex interactions.

From a developmental perspective, however, the search for more variables and more sophisticated data modeling techniques may be grasps in the wrong direction. The coherence of these strategies absolutely depends on the assumption of structural homogeneity across individuals. From a developmental standpoint this assumption is often highly questionable, defies common sense, and is contradicted by the human capacities for equifinal and multifinal functioning. It is sustained only by the tranquilizing effect of the paradigm's flawed justification standards, which in turn are sustained and virtually required by deficiencies in the discovery.

Local solutions

It would be wrong to conclude that these interdependent paradigm weaknesses always and necessarily lead to flawed conclusions and misguided research efforts. To be sure, developmental psychopathology research has achieved significant progress in many areas despite its paradigmatic handicaps (Cicchetti & Cohen, 1995a, 1995b). Thoughtful researchers have long recognized the inadequacies of the paradigm and have struggled to wrestle in meaningful ways with their phenomena within its considerable constraints. This is the sub-rosa reality of the research enterprise not reflected in the sterilized, formal reports of research findings published in journal articles. The details and machinations of creative exploratory data analyses, for example, are seldom reflected in the literature. When results are negative they are often suppressed and not reported for sound conceptual and theoretical reasons. When outcomes are positive they are often reported sans details about how variables were combined and recombined, and data analyses cast and recast, until they yielded patterns that made sense to the investigator.

Although many of these practices are technical violations of the statistical standards of the justification process, they are often justified as reasonable adaptations to unreasonable requirements. Ultimately, however, they are inadequate as solutions to the developmentalist's dilemma for two reasons. First, local circumvention strategies only free researchers -and partially at that-from the arbitrary standards of the flawed justification process. The trade-off is that this also leaves researchers with no uniform scientific standard at all, which is the essential self-correcting engine of scientific activity. Second, as indicated earlier, major decisions about study design and data collection are so constrained by the theory-testing and statistical requirements of the paradigm that resulting data are often of limited usefulness for purposes beyond the hypothesis testing objectives for which they were collected (see Study design subsection).

Resolving Developmentalist's Dilemma

It is one thing to understand the inadequacies of the existing paradigm. It is quite another to engage the considerable challenges of developing a suitable alternative for guiding and interpreting developmental psychopathology research. A useful starting point is to consider briefly what the human capacities for equifinal and multifinal functioning do not imply about the challenges ahead. First, they do not imply that individuals will differ structurally to the extent that each must be studied as his/her own unique class. It is certainly the case that individuals *are* literally unique in many important respects. But the coherence of developmental science rests on a reasonable expectation that we are also sufficiently alike in important respects to warrant the search for structural similarities across individuals and for the laws governing the emergence, development, functioning, and functions of those structures. Second, equifinality and multifinality do not imply that all overt functioning patterns will be characterized by theory-relevant structural heterogeneity. There may be many overt characteristics for which the theory-relevant structures are relatively homogenous across individuals, and many others for which most individuals will be characterized by a relatively small number of qualitatively different structures. Third, they do not imply that all structural differences between individuals are relevant to all research questions. There may be many overt syndromes that stem from structures and processes that differ qualitatively across individuals at one level of organization but converge onto a final common pathway at others. Thus, for purposes of research, the alternative to structural homogeneity is neither complete nor unmanageable heterogeneity.

From individual differences to different individuals

The possibility of multiple causal structures underlying similar patterns of overt functioning does not merely add to the complexity of studying causal processes and structures. It radically redefines the discovery challenge in ways that require a fundamental reconceptualization of developmental psychopathology research. The reason, simply, is that the entire individual differences research tradition absolutely depends on the simplifying assumption that overt functioning characteristics (e.g., shyness, sociability, depression, aggression, self-esteem, etc.) are relatively straightforward reflections of unknown, complex, but nonetheless unitary causal structures that are common across individuals. If this assumption is invalid, so also are the assumptions that (a) individual differences in the psychological characteristics of subjects reflect only quantitative differences in their underlying structures, (b) each variable (e.g., internal structure or external factor) under consideration either is or is not causally relevant to the overt characteristic, and therefore (c) its causal role, if any, in producing that characteristic, however direct or indirect, will always be the same. In short, a rejection of the structural homogeneity assumption perforce requires a rejection of research strategies, methods, and assessment procedures that depend for their coherence and justification on its validity.

Inevitably, understanding individual differences in overt functioning will require understanding the qualitatively different structures from which they arise in the population. This, in turn, requires an initial shift from research concerning *individual differences* to research concerning *different individuals.* The unit of analysis in individual differences research, of course. is not an individual or group of individuals. but the sample itself. Data points contributed by individuals within a sample are merely used to construct an average hypothetical individual characterized by the structure shared in common by sample members. This strategy loses its coherence if individuals within the sample do not share a common theory-relevant structure. If there is reason to believe that a given characteristic of functioning may stem from different structures and processes in different individuals, then the immediate challenge is to determine how many qualitatively different structures there are and which individuals are characterized by which structures. The necessary starting point for this search is an initial intensive focus on the causal structures and functioning of particular individuals, only then followed by a search for others characterized by similar theory-relevant structures.

From causal factors to causal structures

The shift in focus from samples to individuals also requires a fundamentally different approach to identifying causal structures and processes. In sample-based research, the search for causal structure is typically approached with bottom-up strategies. That is, models of causal structure are constructed by combining individual factors and processes that have been identified as correlates of overt functioning patterns. But the coherence of this approach depends on the simplifying assumption that a given factor either is or is not causally related to the phenomenon of interest. To the extent that different causal structures may give rise to the same phenomenon across different individuals, the logic underlying bottom-up strategies breaks down. Any causal relation(s) between a variable or set of variables and the overt characteristic of interest may be virtually undetectable in a general search across structurally heterogeneous cases. Thus, the challenge of identifying causal structure at the individual level will require top-down strategies that begin at the highest levels of broadly defined explanatory constructs and work systematically down the hierarchy to specific processes and factors.

A useful framework for thinking about top-down discovery strategies is provided by Mackie's concept of causally sufficient conditions in the naturalistic world (Mackie, 1974; see also Russell, 1974, for an extension of Mackie's work on development). In the study of closed systems in physics, theories frequently can be refined to the point of being able to specify conditions that are both necessary and sufficient for the occurrence of a phenomenon (Elasser, 1981). Many effects in the natural world, however, can be produced by a variety of conditions that are neither necessary nor sufficient for their occurrence. A building fire, for example, may be caused by any number of different conditions, such as an act of arson, an electrical short-circuit, a carelessly discarded cigarette, or a bolt of lightning, etc. Each of these conditions is *un*-necessary but sufficient for producing a building fire under certain circumstances (e.g., the presence of combustible material, oxygen, etc.). Moreover, each condition consists of a different constellation of elements, each of which is causally insufficient but contributes non-redundantly to the causal efficacy of that condition. The elements of the arson condition, for example, might include a flammable liquid, a lighted match, and a person intending to bum the building down. Alone, each element is causally insufficient for producing a fire; it is only in combination that they constitute the arson condition. Thus, they are insufficient, nonredundant components of an unnecessary but sufficient *(inus)* condition (i.e., arson) for producing the building fire effect.

Different causal structures that are capable of producing similar characteristics of overt-functioning can be conceptualized usefully as *inus* conditions. The added complexity of studying *inus* conditions underlying human functioning, of course, is that any given factor may play a qualitatively different causal role across different *inus* conditions. As in our earlier example, factors such as marital distress, parent symptomatology, poor peer relations, school failure and the like may be causal in one or more *inus* conditions for antisocial behavior, consequential in others, and spurious in yet others. Nonetheless, the inus concept is useful for thinking initially about unnecessary but sufficient causal conditions. Consider, for example, the variety of potential inus conditions that might give rise to a pattern of persistent antisocial behavior in childhood. One condition might be a criminogenic neighborhood and/or family environment in which antisocial behavior is modeled, expected, and/or rewarded by major influences in a child's life. This is certainly a plausible model to consider in late 20th century America, where in many major cities the allure of drug-related crime is ever present, and where gangs virtually control the social commerce and economic life of many neighborhoods (Richters, 1996). It is not difficult to imagine that the causal structures responsible for antisocial behavior in these children will differ in important respects from those giving rise to antisocial behavior among children raised in a mainstream culture characterized by fundamentally different influences, socialization pressures, and opportunities. Even within each of these different subcultures there may be children whose antisocial behavior was initiated and/or is sustained primarily by the need for money to support drug or alcohol dependence, an expression of anger against parents or adult authority figures, or a method of attention seeking. Persistent antisocial behavior in other children may stem primarily from inherited or acquired deficits in nervous system functioning that translate into problems of impulse control or difficulties learning from experience either absent or irrespective of these other conditions.

These are not, of course, mutually exclusive sources of influence, and there are no doubt many children whose antisocial behavior arises from complex combinations of these and other sources of influence. The necessary basis for understanding these more complex causal structures, however, may be an understanding of the structures characteristic of children for whom more primary sources of influence can be identified. This does *not* imply the study of extreme cases: it is actually the opposite. It is a call for focusing initially on children for whom the primary reasons or explanations for antisocial behavior can be identified most easily. The value of intensively studying well characterized exemplars of carefully defined structures has been amply demonstrated in the developmental literature (Hinde, 1992; Hinde & Dennis, 1986; Kagan, 1992; Kavale & Forness, 1987).

Referring back to our earlier example, the initial emphasis would be on identifying the most common primary causes of building fires (i.e., *U*nnecessary but sufficient conditions), rather than their individual components (i.e., insufficient, nonredundant elements). Thus, approaching the structural heterogeneity challenge in the domain of children's antisocial behavior would begin by developing an initial taxonomy of putative inus conditions-primary reasons or explanations -for persistent antisocial behavior. The working assumption would be that qualitatively different *inus* conditions would be reflected in the qualitative nature, history, and patterns of antisocial behavior across time and contexts. A formulation of prototype descriptions for candidate *inus* conditions would then provide a foundation and rationale for the intensive study of individuals typified by those conditions. The assumptions underlying this bootstrapping strategy may be fundamentally wrong, and there may be obvious reasons why it is unlikely to succeed. An overemphasis on these possibilities, however, would miss the point. Overcoming the weaknesses of the existing paradigm and developing suitable alternatives will require a willingness to proceed from common sense, to accept the inevitability of false starts and mistakes, and to value and harness the corrective feedback they provide.

The next step, of course, would be to identify individual children with antisocial behavior patterns consistent with a particular inus condition. This would not initially require methods for reliably classifying antisocial children into a candidate *inus* condition. Rather, the initial goal literally would be to identify a single child for whom there are concrete, plausible reasons for believing that he or she is a likely prototype of a particular *inus* condition. Note that the shift in focus from sample to individual fundamentally redefines the challenges and opportunities of the discovery process.

Consider, for example, the contrast between sample-based strategies for studying the causes of antisocial behavior and the more common individual-based strategy employed in practical nonresearch settings. In a sample-based study the researcher will probably never meet or interview the antisocial child and his/her parents. Moreover, on the basis of idiosyncratic, pragmatic considerations often loosely linked to a particular theoretical notion, the research design will focus on collecting a fixed, circumscribed set of information that may or may not be relevant to different children in the sample, thereby necessarily ignoring information that might be decisively important to understanding any particular child. Furthermore, in the service of scientific rigor the researcher will require that all information be collected in a standardized way from the child and his/her family, typically in the form of standardized questionnaires and/or checklists and/or structured interviews. By definition, the standardized process will ignore many characteristics of individuals that may have an important influence on how they understand, interpret, and respond to questions. Also, parcels of information concerning each subject will be collected in isolation from one another with no effort to evaluate them configurally in an effort to determine likely and unlikely explanations for the child's antisocial behavior. In fact, the assumption of structural homogeneity renders this task unnecessary. Because the causes of antisocial behavior are presumed to be qualitatively the same across children in the sample, evidence will be sought through data analytic procedures identifying patterns of covariation between and among variables across the sample.

When the unit of analysis shifts from sample to individual each of these restrictions is necessarily relaxed. The burdens of discovery become not unlike those of a parent, clinician, social worker, or probation officer confronted with an antisocial child. In these settings the discovery process is directed at gaining enough of an understanding of the nature, history, and current context of the child's problems and circumstances to formulate a hypothesis about what seems to be motivating, causing, and influencing the child's antisocial behavior and based on that hypothesis, an initial strategy for intervening. Although the researcher's objective is not to formulate an intervention strategy, the discovery challenge and objective is very much the same. The clinician would probably begin by collecting certain common information about the child, parents, history, and current circumstances with respect to every child. But the clinician will also remain flexible about reaching beyond these initial boundaries for additional information depending on the specifics of each particular case. He or she will also make liberal use of alternative wording, rephrasing, asking questions in different ways, probing, mirroring, and a range of other interview techniques for ensuring that the parents and child understand the intended meaning and nuance of questions asked and that the clinician is interpreting accurately the intended meaning of their answers. The clinician also may decide to seek information from additional informants depending on the nature of the case, looking for points of convergence and divergence, sources of and motivations for bias, and so forth, in the process of trying to make sense of the different streams of information. The clinician also will pay very careful attention to the nature of timing of various kinds of events and circumstances relative to the child's behavior patterns, because they can be decisive in ruling in or out of contention certain potential explanations for the child's behavior. Also, the clinician will want access to all relevant pieces of information about the child, family history, and circumstances *at the same time* in the service of making complex configural judgments about which factors seem to be most important, deserve the most explanatory weight, and have the most clinical significance in understanding the child's behavior. This configural decision is decisive because the meaning and significance to the clinician of any particular piece of information can hinge exquisitely on the rest of the pattern of information in the matrix.

The nature of this undertaking immediately raises three major questions. The first is whether it is necessary or desirable to require configural, clinical, inferential judgments about the meaning of the information concerning a child. After all, there is an impressive body of data extending back decades showing that the clinical judgments are often far inferior to fixed procedures and computer algorithms when it comes to combining data to make predictions (Meehl, 1954). This literature. however, concerns the very specific case of the comparative accuracy of humans and computers in predicting or classifying a phenomenon according to a known formula. The problem posed by structural heterogeneity is that we do not yet know what the formula is or should be. Paul Meehl, in a follow-up article to his classic book *Clinical versus statistical prediction* was quite explicit about conditions under which human judgments are indispensable: "Sometimes there is no formula because the prediction problem is too open-ended . sometimes the very categorizing of the raw observations involves gestated stimulus equivalences for which the laws are unknown, and hence cannot be formulated (although the clinician himself exemplifies these laws and can therefore 'utilize' them); in still other cases there is no formula because nobody has bothered to make one" (Meehl, 1954; Meehl, 1973, p. 83). Meehl also identified several decision-making conditions under which it is necessary to rely on clinical judgment because there is nothing else available. The first is when the decision task is open ended in the sense that the content of the judgment itself must be decided by the clinician. The second is when the decision task requires a recognition of unanalyzed stimulus equivalences, such as perceptual gestalts, psychological similarities in physically dissimilar events. These are recognition tasks for which the clinician implicitly relies on various forms of analogical and primary-process thinking, the algorithms of which are beyond the level of awareness. A third condition is when a pattern of information is relatively rare, or in any event not expected, but its appearance is sufficiently relevant to warrant contramanding a more typical interpretation. Fourth, there are conditions wherein judgments rely on hypothetical mental constructs of such a general nature that they permit a wide range of concrete manifestations that can be recognized but not anticipated in advance. These conditions are likely to be prominent in any initial attempt to recognize behavior and information patterns that are plausible reflections of particular *inus* conditions. Thus, configural human judgments at the point of data assessment may be indispensable in the initial stages of individual-based strategies.

The second issue is whether there is a considerable potential for capitalizing on chance in making configural judgments at the individual level. Suppose, for example, that a factor or experience that seems to the researcher to be central to understanding a particular child's antisocial behavior turns out to be so common among non-antisocial children that it is neither a risk factor nor correlate of antisocial behavior in the population. In sample-based research, on the assumption of structural homogeneity, this would often be sufficient for questioning the causal relevance of a factor. A rejection of the homogeneity assumption, however, invalidates this logic. Again, it is quite possible for a factor to be causally relevant in one inus condition for an outcome and not others, such that it may not be related to antisocial behavior in a structurally heterogeneous population. This does not imply that the clinical judgment is correct, only that it is not invalidated by information based on the general population.

Rethinking justification

This leads naturally to the third and most difficult question of how to guard against illusory explanations when focusing on an individual. It is obviously quite easy in a clinical or research setting as it is in daily life to believe passionately in explanations that may have no merit whatever. Thus, if a clinical judgment about the importance of a factor cannot be evaluated by reference to its overall relation to antisocial behavior in the general population, by what criteria *can* it be evaluated? This is the most important of the three issues because sound justification standards provide the indispensable self-correcting feedback for shaping and insuring the integrity of creativity in the discovery process. In general, judgments made at the individual level must be evaluated by qualitative criteria such as coherence. explanatory power, and the ability to account for facts that are otherwise difficult to explain, especially when applied to individuals with putatively similar causal structures.

The crucial question of how to translate these general principles into concrete criteria and standards for evaluating individual-focused research will be one of the major challenges of emphasizing greater creativity and exploration in the discovery process. This will be especially challenging in developmental psychopathology research because the complexities of its phenomena are such that a meaningful discovery process will require a wide range of research strategies, methods, and procedures. One of the many lessons of the current paradigm is that no single method or research strategy should be held out a priori as inherently superior or inferior to another. The strengths and weaknesses of a given discovery methodology can be evaluated only with specific reference to the phenomena and research questions under study. A guiding principle for thinking about discovery methods and strategies should be methodological pluralism, a willingness to develop and embrace whatever methods and strategies are necessary for understanding the phenomena of interest, including categorical, dimensional, idiographic, nomothetic, variable-based, individual-based, cross-sectional, longitudinal, experimental, quasiexperimental, historical, and ethnographic approaches (Bergman & Magnusson, in press; Bhatara, McMillin, & Krummer, 1995; Gould, 1986, 1987; Magnusson & Allen, 1983; Magnusson & Bergman, 1988; Maton, 1993; McCord, 1993; Shadish, 1995; Shames, 1990).

Methodological pluralism in the discovery process cannot flourish within the constraints of an arbitrarily restrictive and misguided (and misguiding) justification process. Beyond the flaws in its logic and standards, the current justification paradigm constrains the discovery process to a narrowly defined range of activities in the service of ritualized hypothesis testing. The requirements imposed by hypothesis testing, population generalizability, minimizing Type 1 errors, and maintaining low variables-to-subjects ratios often rob the discovery process of its creative potential and forces researchers into premature tests of theories. It does not preclude more meaningful approaches to discovery, but it does engender and reinforce research practices that are often antithetical to sound discovery. In addition, the narrow emphasis on "explained variance" as the chief criterion for evaluating theories often relegates explanatory power to a secondary or tertiary status by emphasizing prediction.

Although the current justification process lacks rigor for testing theories, it is not obvious that a continued emphasis on formal theory testing is either practical or desirable in many areas of developmental psychopathology research. In most domains we are not dealing with formal theories in the conventional ideal sense, but with very general theoretical notions about particular aspects of complex phenomena. The emphasis of the current paradigm on theory testing has led to a narrow focus on maximizing the "explained" variance in outcome measures of interest with shifting sets of predictor variables. In many domains of children's social and emotional development and psychopathology we continually struggle against what seems to be an upper limit of explained variance, which can often be reproduced with very different constellations of variables. In the end, it is seldom clear even within the forgiving standards of the justification paradigm what interpretations of multivariate effects are justified or how they relate to the original phenomena of interest (Cairns, 1986; Kaplan, 1996). A more reasonable and practical approach to scientific rigor under these circumstances would be to emphasize questions that are answerable in straightforward ways and answers that are interpretable and justifiable by reasonable standards. The most constructive direction for reform would be to replace the current emphasis on ambitious claims based on weak evidence with a justification process based on the time-honored principles of strong inference (Platt, 1964).

Conclusion

Differences between mature and fledgling sciences are often less apparent in the errors they commit than in the speed with which they recognize and respond to those errors. By most accounts, astronomy's launching of the myopic Hubble Space Telescope was one of the 20th century's leading scientific blunders. Nonetheless, it took astronomers only moments to recognize that Hubble was transmitting distorted images of the universe back to earth, only days to diagnose precisely, and a matter of months to correct. Psychology launched its equivalent of the myopic Hubble when it founded its science on the closed system assumptions and explanatory models of 19th century classical physics. By itself, this error would have been only a short-term handicap. Like all mistaken ideas and premises in science it would have been remedied eventually by the rigors of scientific method. By relying on null-hypothesis significance testing procedures to evaluate research findings, however, psychology deprived itself of the essential self-correcting engine of scientific method. As a consequence, psychology's reaction time in recognizing and correcting its initial error has been measured so far in decades rather than months.

Psychology's long-standing paradigm problems are especially salient to the newly formalized discipline of developmental psychopathology for three related reasons. First, developmental psychopathology stands poised to define a new center of gravity for integrative, multidisciplinary research in many areas of psychological science because of the tremendous heuristic power of developmental, open system framework. Second, the research strategies, methods, and data analytic procedures of psychology's paradigm on which developmental psychopathology currently depends are fundamentally inadequate in important respects to the study of complex open systems. As a consequence, much of the richness and heuristic power of the developmental framework is lost when empirical research is conceptualized and carried out within the constraints of the current paradigm. Resolving the developmentalist's dilemma will require more than a recognition of the inadequacies of the existing paradigm. It will require intensive efforts to develop indigenous research strategies, methods, and standards with fidelity to the complexity of developmental phenomena.

It would be wrong to conclude that the current paradigm always leads to wrong conclusions and misguided research efforts. Nonetheless, a dispassionate assessment of its mutually enabling discovery and justification weaknesses provides a reasonable basis for concern about the possibility that, in many important areas of research, we may be unknowingly (and, within the constraints of the paradigm, unknowably) studying artifacts rather than facts, flawed theory and flawed data, and illusory phenomena of our own creation. In short, we may be committing Type 3 errors by answering the wrong questions when we should be answering the right questions (Mitroff & Featheringham, 1974). It is generally true in science that as long as the tools of a paradigm prove capable of solving the problems it defines, science functions best by relying on those tools. The reason, as Kuhn noted, is clear: "... in manufacture so in science-retooling is an extravagance to be reserved for the occasion that demands it" (Kuhn, 1970, p. 76). It is difficult to image a more compelling occasion for retooling than the emergence of an exciting new discipline for which the existing paradigm is inadequate and within which the necessary tools do not exist.

Ultimately, the responsibility for constructive reform rests with each individual investigator. The lesson of psychology's history, however, is that individual efforts are unlikely to flourish or succeed in the absence of a more broadly based commitment to reform. The responsibility and power for engendering and maintaining such a commitment, in turn, rests with institutions of scientific leadership and influence. Editorial boards of scientific journals and public and private funding agencies are in unique positions to implement the necessary incentive structures for encouraging efforts toward paradigm reform. Developmental psychopathology's scientific successes in the 21st century will hinge decisively on its willingness and ability to translate sophisticated developmental thinking into equally sophisticated developmental research designs, methods, and practices. The heuristic value of the developmental psychopathology framework is nowhere more evident than in the clarity with which it reveals the nature of psychology's long-standing paradigm problems and in the powerful prescription it provides for a corrective lens.

References

Allport. G. W. (1960). The open system in personality theory. *Journal of Abnormal and Social Psychology, 61. 301-310.*

Altman, I., & Rogoff, B. (1987). World views in psychology: Trait, interactional, organismic, and transactional perspectives. In D. Stokols & I. Altman (Eds.), *Handbook of environmen!al psychology* (pp. 740). New York: Wiley.

American Psychological Association. (1974). *Publication manual of the American Psychological Association* (2nd ed.). Baltimore: Garamond/Pridemark Press.

Anastasi, A. (1958). *Differential psychology* (3rd ed.). New York: Macmillan.

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

Baldwin, J. M. (1895). *Mental development of the child and the race: Methods and processes.* New York: Macmillan.

Beach, F. A., & Jaynes (1956). Studies of maternal retrieving in rats, III: Sensory cues involved in the lactating female's response to her young. *Behaviour, 10, 104-125.*

Bergman, L. R., & Magnusson, D. (in press). A person-oriented approach in research on developmental psychopathology. *Development and Psychopathology.*

Bevan, W. (1991). Contemporary psychology: A tour inside the onion, *American Psychologist, 46, 475-483.*

Bevan, W., & Kessel, F. (1994). Plain truths and home cooking: Thoughts on the making and remaking of psychology. *American Psychologist, 49, 505-509.*

Bhatara, V. S., McMillin, .I M., & Krummer, M. (1995). Aggressive behavior and fire-setting in a 4-year-old boy associated with ingestion of ground beef contaminated with bovine thyroid tissue: A case report and review of neuropsychiatric thyrotoxicosis. *Journal of Child and Adolescent Psychopharmacology 5, 255-271.*

Brandt, L. W. (1973). The physics of the physicist and the physics of the psychologist. *International Journal of Psychology, 8,* 61-72.

Bronfenbrenner, U. (1979). *The ecology of human development: Experiments by nature and design.* Cambridge, MA: Harvard University Press.

Cairns, R. B. 1986). Phenomena lost: Issues in the study of development. In .I. Valsiner (Ed.), *The individual subject and scientific psychology* (p p. 97-l12*). New* York: Plenum Press.

Cicchetti, D. (1984). The emergence of developmental psychopathology. *Child Development,* 55, *1-7.*

Cicchetti, D. (1990). A historical perspective on the discipline of developmental psychopathology. In J. Rolf, A. Masten, D. Cicchetti, K. Nuechterlein, & S. Weintraub (Eds.), *Risk and protective factors in the development of psychopathology (pp. 2-28).* New York: Cambridge University Press.

Cicchetti, D. (1993). Developmental psychopathology: Reactions, reflections, and projections. *Developmental Review, 13, 471-502.*

Cicchetti, D., & Cohen, D. J. (Ed%). (1995a). *Developmermd psychopathology: Vol. 1. Theory and methods. New* York: John Wiley & Sons, Inc.

Cicchetti, D., & Cohen, D. J. (Eds). (1995b). *Developmental psychopathology: Vol. 2. Risk, disorder, und adaptation. New* York: John Wiley & Sons, Inc.

Cicchetti, D., & Richters, J. E. (1993). Developmental psychopathology considerations in the study conduct disorder. *Development and Psychopathology, 5, 33 I–344.*

Cicchetti, D., & Rogosch, F. A. (1996). Equifinality and multifinality in developmental psychopathology. *Development and Psychopathology, 8, 597-600.*

Cohen, J. (1990). Things I have learned so far. *American Psychologist, 45, 1304-I 3 12.*

Cohen, J. (1994). The earth is round *(p* < .05). *American Psychologist, 49, 997-1003.*

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

Edwards, W. (1965). A tactical note on the relation between scientific and statistical hypotheses. *Psychological Bulletin, 63, 400-402.*

Elasser, W. M. (1966). *Atom and organism A new approach to theoretical biology.* Princeton, NJ: Princeton University Press.

Elasser, W. M. (1981). Principles of a new biological theory: A summary. *Journal of Theoretical Biology, 89, 131-150.*

Emde, R. N. (1994). Individuality, context, and the search for meaning. *Child Development, 65,* 719-737.

Estes, W. K. (in press). Significance testing in psychological research: Some persisting issues. *Psychological Science.*

Evans, C., & McConnell, T. R. (1941). A new measure of introversion-extroversion. *Journal of Psychology, 12, 111-124.*

Ferguson, L. (1959). *Statistical analysis in psychology and education. New* York: McGraw-Hill.

Feynman, R. (1986). *Surely you're joking, Mr. Feynman! New* York: Bantam Books.

Fox, R. E. (1996). Charlatanism, scientism, and psychology's social contract. *American Psychologist, 51, 777-784.*

Gerstein, D. R., Luce, R. D., Smelser. N. J., & Sperlich, S. (1988). *The behavioral and social sciences: Achievements and opportunities.* Washington, DC: National Academy Press.

Gigerenzer, G., Swijtink, Z., Porter, T., Daston, L., Beatty, J., & Kruger, L. (1989). *The empire of churzce: HOW probability changed science and everyday life.* Cambridge: Cambridge University Press.

Glymour, C. (1980). Hypothetico-deductivism is hopeless. *Philosophy of Science, 47,* 322-325.

Gould, S. J. (1986). Evolution and the triumph of homology, or why history matters. *American Scientist, 74,* 60-69.

Gould, S. J. (1987). Darwinism defined: Sifting fact from theory. *Discover, 8,* 64-70.

Guilford, J. P. (1942). *Fundamental statistics in psychology and education* (3rd ed.). New York: McGraw-Hill.

Herrenkohl, L. R., & Rosenberg, P. A. (1972). Exteroceptive stimulation of maternal behavior in the naive rat. *Physiology & Behavior, 8, 595-598.*

Hinde, R. A. (1992). Developmental psychology in the context of other behavioral sciences. *Developmental Psychology, 28, 1018-1029.*

Hinde, R. A., & Dennis, A. (1986). Categorizing individuals: An alternative to linear analysis. *International Journal Behavioral Development, 9,* 105-l19.

Horgan, J. (1996). Why Freud isn't dead. *American Scientist,* December, 106-l11.

Institute of Medicine. (1989). *Research on children and adolescents with mental, behavioral, and developmental disorders.* Washington, DC: National Academy Press.

Kagan, J. (1992). Yesterday's premises, tomorrow's promises. *Developmental Psychology, 28, 990-997.*

Kagan, J.. Kearsley, R. B., & Zelazo, P. R. (1978). *Infancy: Its place in haman development.* Cambridge, MA: Harvard University Press.

Kaplan, H. B. (1996). Toward an understanding of resilience: A critical review of definitions and models. In M. D. Glantz, J. Johnson, & L. Huffman (Eds.), *Resiliency and development: Positive life adaptations. New* York: Plenum Press.

Kavale, K. A., & Fomess, S. R. (1987). The far side of heterogeneity: A critical analysis of empirical subtyping research in learning disabilities. *Journal of Learning Disabilities, 20, 374-382.*

Koch, S. (1959). *Psychology: A study of science* (Vol. *3). New* York: McGraw-Hill.

Koch, S. (1961). Psychological science versus the science-humanism antinomy: Intimations of a significant science of man. *American Psychologist, 16, 629-639.*

Koch, S., & Leary, D. E. (Eds.). (1992). *A century of psychology as science.* Washington, DC: American Psychological Association Press.

Kuhn, T. S. (1957). *The Copernican revolution.* Chicago: University of Chicago Press.

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

Kuhn, T. S. (1977). *The essential tension.* Chicago: University of Chicago Press.

Lakatos, I. (1977). *The methodology of scientific research programmes :Philosophical papers* (Vol. I). J. Worral & E. G. Zahar (Eds.). Cambridge: Cambridge University Press.

Laudan, L. (1977). *Progress and its problems: Toward a theory of scientific growth.* Berkeley: University of California at Berkeley.

Leahey, T. H. (1987). *A history of psychology: Main carrents in psychological thought* (2nd ed.). New York: Prentice-Hall.

Lerner, R. M., & Kaufman, M. B. (1985). The concept of development in contextualism. *Developmental Review, 5, 309-333.*

Leventhal, L. (1992). Nudging aside Meehl's paradox. *Canadian Psychology, 35, 283-298.*

Lewin, K. (1931 a). The conflict between Aristotelian and Galilean models of thought in contemporary psychology. *Journal of General Psychology, 5, 147-117.*

Lewin, K. (1931b). Environmental forces in child behavior and development. In C. Murchison (Ed.), *A handbook of child psychology* (2nd ed., pp. 590-625). Worster, MA: Clark University Press.

Loftus, G. R. (1991). On the tyranny of hypothesis testing in the social sciences. *Contemporary Psychology, 36, 102-105.*

Loftus, G. R. (1993). Editorial comment. Memo ry & Co*gnition, 21, 1-3.*

Lykken, D. T. (1968). Statistical significance in psychological research. *Psychological Bulletin, 70, 151-159.*

Lykken, D. T. (1991). What's wrong with psychology anyway? In D. Cicchetti & W. M. Grove (Eds.), *Thinking clearly about psychology* (Vol. 1, pp. 3-39). Minneapolis, MN: University of Minnesota Press.

Mackie, J. L. (1974). *The cement of the universe: A study of causation.* Oxford: Oxford University Press.

Magnusson, D. (1985). Implications of an interactional paradigm for research on human development. *Intemational Journal of Behavioral Development, 8, 115-137.*

Magnusson, D., & Allen, V. L. (1983). Implications and applications of an interactional perspective for human development. In D. Magnusson & V. L. Allen (Eds.), *Human development: An interactional perspective* (pp. 369-387). New York: Academic Press.

Magnusson, D., & Bergman, L. (1988). Individual and variable-based approaches to longitudinal research on early risk factors. In M. Rutter (Ed.), *Studies of psychosocial risk: The power of longitudinal data* (pp. 45-61). Cambridge: Cambridge University Press.

Maltz, M. D. (1994). Deviating from the mean: The declining significance of significance. *Journal of Research in Crime and Delinquency, 31, 434-436.*

Maton, K. I. (1993). A bridge between cultures: Linked ethnographic empirical methodology for culture-anchored research. *American Journal of Community Psychology, 21, 701-121.*

McCord, J. (1993). Descriptions and predictions: Three problems for the future of criminological research. *Journal of Research in Crime and Delinquency, 30, 412425.*

McGuire, W. J. (1983). A contextualist theory of knowledge: Its implications for innovation and reform in psychological research. *Advances in Experimental .Social Psychology, 16, 241.*

Meehl, P. E. (1954). *Clinical versus statistical prediction: A theoretical analysis and review of the evidence.* Minneapolis: University of Minnesota Press.

Meehl, P. E. (1967). Theory testing in psychology and physics: A methodological paradox. *Philosophy of Science, 34,* 103-I 15.

Meehl, P. E. (1973). *Psychodiagnosis: Selected papers.* Minneapolis, MN: University of Minnesota Press.

Meehl, P. E. (1978). Theoretical risks and tabular asterisks: Sir Karl, Sir Ronald, and the slow progress of soft psychology. *Journal of Consulting and Clinical* Psychology, 46, l-42.

Meehl, P. E. (1990). Appraising and amending theories: The strategy of Lakatosian defense and two principles that warrant it. *Psychological Inquiry. I,* 108-141.

Mitroff, I., & Featheringham, T. (1974). On systematic problem solving and the error of the third kind. *Behavioral Science, 19, 383-393.*

Morrison, D. E.. & Henkel. R. E. (Eds.). (1970). *The significance test controversy.* Chicago: Aldine.

Nagel, E. (1961). *The structure of science. New* York: Harcourt, Brace. and World.

National Advisory Mental Health Council. (1990). *National Plan for Research on Child and Adolescent Mental Disorders.* Rockville, MD: National Institute of Mental Health.

New York Times. (1990, July 1). Calamities in space. Section 1. *Editorial desk*, p. 16.

Nunnally, J. C. (1975). *Introduction to statistics in psychology and education.* New York: McCraw~Hill.

Overton. W. F. (1984a). World views and their influence on psychological theory and research: Kuhn-Lakatos-Laudan. In H. W. Reese (Ed.), *Advances in child Development and behavior* (Vol. 18, pp. 19 l-225). New York: Academic Press.

Overton, W. F. (1984b). Comments on Beilin's epistemology and Palermo's defense of Kuhn. In H. W. Reese (Ed.). *Advances in child development and behavior* (Vol. 18, pp. 273-276). New York: Academic Press.

Overton, W. F., & Horowitz, H. A. (1991). Developmental psychopathology: Integrations and integrations. In D. Cicchetti & S. Toth (Eds.), *Rochester Symposium on Developmental Psychopathology: Vol. 3. Models and integrations* (pp. 142). Rochester, NY: University of Rochester Press.

Pepper, S. C. (1942). *World hypotheses.* Berkeley: University of California.

Platt, J. R. (1964). Strong inference: Certain systematic methods of scientific thinking may produce much more rapid progress than others. *Science, 146, 347-362.*

Popper, K. R. (1959). *The logic of scientific discovery.* London: Hutchinson.

Popper, K. R. (1991). Of clouds and clocks: An approach to the problem of rationality and the freedom of man. In D. Cicchetti & W. M. Grove (Eds.), *Thinking clearly about psychology.(Vol. 1,* pp. 100-l39). Minneapolis, MN: University of Minnesota Press.

Radke-Yarrow, M., & Sherman, T. (1990). Hard growing: Children who survive. In J. Rolf, A. Masten, D. Cicchetti, K. Nuechterlein, & S. Weintraub (Eds.), *Risk and protective factors in the development of psychopathology* (pp. 97-l19). New York: Cambridge University Press.

Reichenbach, H. (1938). *Experience and prediction.* Chicago: University of Chicago.

Richters, J. E. (1996). Disordered views of antisocial children: A late 20th century perspective. In C. Ferris (Ed.), *Understanding aggressive behavior in children.* New York: New York Academy of Sciences.

Richters, J. E.. & Cicchetti, D. (1993a). Editorial: Toward a developmental perspective on conduct disorder. *Development and Psvchopathology, 5,* 1-4.

Richters, J. E., & Cicchetti, D. (1993b). Mark Twain meets DSM-III-R: Conduct disorder, development. and the concept of harmful dysfunction. *Development and Psychopathology, 5, 5-29.*

Robinson, D. N. (1984). The new philosophy of science: A reply to Manicas and Secord. *American Psyhologist, 39, 920-921.*

Rorer, L. G. (1991). Some myths of science in psychology. In D. Cicchetti & W. M. Grove (Ed.). *Thinking clearly about psychology* (Vol. 1. pp. 61-87). Minneapolis, MN: University of Minnesota Press.

Rozeboom, W. W. (1982). Let's dump hypothctico-deductivism for the right reasons. *Philosophy of Science, 49,* 637-647.

Russell, J. (1990). Causal explanations in cognitive development. In G. Butterworth & P. Bryant (Eds.), *Causes of development: Interdisciplinary perspectives* (pp. 111-134). London: Harvester, Wheatsheaf Hemel Hemstead.

Rutter, M., & Garmery. N. (1983). Developmental psychopathology. In E. M. Hetherington (Ed.), *Carmichaels manual of child psychology: Vol. 4. Social and personality development* (pp. 775-911). New York: Wiley.

Salmon, W. C. *(1984). Scientific explanation and the causal Structure of the world.* Princeton, NJ: Princeton University Press.

Sameroff, A. (1995). Developmental systems: Contexts and evolution. In D. Cicchetti & D. Cohen (Eds.), *Developmental psychopathology: Risk, disorder and aduptation* (pp. 238-294). New York: Wiley-Interscience.

Sameroff, A. J., & Chandler, M. J. (1975). Reproductive risk and the continuum of caretaking casualty. In F. D. Horowitz. M. Hetherington, S. Starr-Salapatek, & G. Siegel (Eds.), *Review of child development research* (pp. 187-244). Chicago: University of Chicago Press.

Sarason, S. B. (1981). *Psychology misdirected.* New York: The Free Press.

Shadish, W. R. (1995). Philosophy of science and the quantitative-qualitative debates: Thirteen common errors. *Education and Program Planning, I8,* 63-75.

Shames, M. L. (1990). On data, methods, and theory: An epistemological evaluation of psychology. *Canadian Psychology, 31,* 229-238.

Smith. P. M., & Torrey. B. B. (1996). The future of the behavioral and social sciences. *Science, 271*, 61, 1-612.

Spemann. H. (1938). *Embryonic development and induction.* New Haven, CT: Yale University Press.

Sroufe. L. A., & Rutter, M. (1984). The domain of developmental psychopathology. *Child Development, 83, 173-189.*

Thurstone. L. L. (1938). *Primary mental abilities.* Chicago: University of Chicago Press.

Valsiner. J. (Ed.). (1986a). *The individual subject and scientific psychology.* New York: Plenum Press.

Valsiner. J. (1986b). Between groups and individuals: Psychologists' and laypersons' interpretations of correlational findings. In J. Valsiner (Ed.), *The individual Subject and scientific psychology* (pp. 113-l52). New York: Plenum Press.

von Bertalanffy. L. (1968). *General system theory.* New York: George Braziller. Inc.

Waddington. C. H. (1957). *The strategy of genes*. London: Allen & Unwin.