



National Institute of Mental Health

Roundtable Discussion
Promissory Notes and Prevailing Norms in Social and Behavioral Sciences Research

December 5th, 1997

Renaissance Hotel, Washington D.C

Invited Roundtable Participants

Mark Appelbaum, Ph.D.
Department of Psychology (4314A McGill Hall)
University of California at San Diego
La Jolla, California 92093-0109

Robert B. Cairns, Ph.D.
Center for Developmental Science
CB# 8115, 521 S. Greensboro St.
University of North Carolina - Chapel Hill
Chapel Hill, NC 27599-8115

Reuven Dar, Ph.D.
Department of Psychology
Tel Aviv University
Tel Aviv 69978 Israel

Nathan Fox, Ph.D.
Department of Human Development (Benjamin Building)
University of Maryland
College Park, Maryland 20742

Stephen P. Hinshaw, Ph.D.
Department of Psychology (3409 Tolman Hall)
University of California @ Berkeley
Berkeley, California 94720-1650

Kimberly Hoagwood, Ph.D.

Michael Maltz, Ph.D.
Department of Criminal Justice (M/C 141)
The University of Illinois at Chicago
1007 West Harrison Street
Chicago, Illinois 60607-7137

William McGuire, Ph.D.
Department of Psychology (311 K)
Yale University (Box 208205)
New Haven, Connecticut 06520-8205

Leonard Mitnick, Ph.D.
Division of Mental Disorders, Behavioral Research and AIDS
National Institute of Mental Health
5600 Fishers Lane, Room 18C-17
Rockville, Maryland 20857

John E. Richters, Ph.D.
Developmental Psychopathology Research Branch
National Institute of Mental Health
5600 Fishers Lane, Room 18C-17
Rockville, Maryland 20857

Daniel N. Robinson, Ph.D.
Department of Psychology
Georgetown University
306B White-Gravenor (Box 571001)

Division of Basic and Clinical Neuroscience
Research

National Institute of Mental Health
5600 Fishers lane, Room 10C-06
Rockville, Maryland 20857

Peter S. Jensen, M.D.
Developmental Psychopathology Research
Branch

National Institute of Mental Health
5600 Fishers Lane, Room 18C-17
Rockville, Maryland 20857

Jerome Kagan, Ph.D.
Department of Psychology (William James Hall
1346)

Harvard University
33 Kirkland Street
Cambridge, Massachusetts 02138

Geoffrey R. Loftus, Ph.D.
Department of Psychology
University of Washington

Seattle, Washington 98195-1525

Washington, D.C. 20057-1001

Kenneth Rubin, Ph.D.
Department of Human Development (Benjamin Building)

University of Maryland
College Park Maryland 20742

Ellen Stover, Ph.D.
Division of Mental Disorders, Behavioral Research and
AIDS

National Institute of Mental Health
5600 Fishers Lane, Room 18C-17
Rockville, Maryland 20857

Edison J. Trickett, Ph.D.
Department of Psychology (Zoology-Psychology
Building)

University of Maryland
College Park, Maryland 20742-4411

Welcome and Overview

Dr. Richters: Good morning, welcome, and thank you all for coming. I'm deeply indebted to each of you contributing so generously and thoughtfully to the intensive background dialogue leading to today's unusual undertaking. Thanks also for re-arranging your schedules on such short notice to be here. I'm looking forward to working with you on the difficult and exciting challenges ahead. I'm sorry to report that two of our scheduled participants, Jerry Kagan and Ruvi Dar, will not be able to join us today. Jerry is suffering from a bout of the flu and Ruvi has been grounded in Israel by the general strike. Despite their physical absence, though, Jerry and Ruvi remain very much part of our group and are represented today by the comments and ideas they contributed to our background dialogue.

Most workshops convened by the National Institute's of Health are devoted to the puzzle-solving activities of normal science, where the puzzles themselves and the strategies available for solving them are determined largely in advance by the shared paradigmatic assumptions, frameworks, and priorities of the scientific community's research paradigm. They are designed to facilitate what Thomas Kuhn referred to as *elucidating topological detail within a map whose main outlines are available in advance*. And apparently for good reason. Historical studies by Kuhn and others reveal that science moves fastest and penetrates most deeply when its practitioners work within well-defined and deeply ingrained traditions and employ the concepts, theories, methods, and tools of a shared paradigm. No paradigm is perfect and none is capable of identifying, let alone solving, all of the problems relevant to a given domain of inquiry. Thus, the essential day-to-day business of normal science is not to question the limits or adequacy of a given paradigm, but rather to exploit the presumed virtues for which it was adopted. As Kuhn cautioned in his discussion of paradigms, *re-tooling, in science as in manufacture, as an extravagance to be reserved for the occasion that demands it*.

Well, as the marketing people say --- *this is not your father's Oldsmobile*. We are breaking with tradition today by stepping outside the map to initiate and pursue a long-overdue dialogue about paradigm reform and scientific retooling. Our warrant for prosecuting this agenda is a Kuhnian *occasion that demands it*--- is a protracted paradigm crisis, the neglect of which has hurt us terribly and the resolution of which will determine the viability and fate of the social and behavioral sciences in the 21st century. Since the details of the crisis are well known within and outside our ranks, a brief sketch of its main outlines will suffice as a framework for our dialogue today. They include, (a) widespread dissatisfaction with the meager theoretical progress and practical yield of more than a century of social and behavioral sciences research in many substantive domains, (b) long-neglected yet widely recognized deficiencies in the epistemological assumptions, discovery practices and justification standards of the dominant paradigm on which the social and behavioral sciences have relied --- and rely--- to conceptualize, interpret, and guide their empirical research, (c) a broadly based consensus among leading scholars and scientists about the need for fundamental paradigm reforms, and (d) institutional incentive structures that not only encourage and reinforce the status quo but discourage constructive reform efforts.

Our objective for the next eight hours is to formulate strategies and recommendations for leveraging the resources and influence of the National Institute of Mental Health to foster a climate of constructive reforms where they are needed by freeing investigators in from the oppressive constraints of existing paradigms and facilitating, encouraging, and funding their retooling their efforts. The mechanisms available to us will include (a) workshops and scientific conferences devoted to these themes, (b) Program Announcements and Requests for Applications (with set-aside money) signaling the Institute's commitment to methodological pluralism and the need to develop 'indigenous methodologies' capable of reflecting and illuminating the richness of human phenomena, (c) creating special IRG's for the evaluation of scientifically novel grant applications, (d) summer institutes in the Washington, D.C area devoted to issues of paradigm reform and retooling, and (e) working with journal editors, editorial boards, public and private funding agencies, and professional organizations in the service of reform efforts.

Before we begin though, I'd like to turn the meeting over to Ellen Stover, Director of our new Division of Mental Disorders, Behavioral Research and AIDS, for some opening remarks, followed by Peter Jensen, the Associate Director for Child and Adolescent Research at the Institute, and then Kimberly Hoagwood, Associate Director for Child and Adolescent Translation Research. Following this, I'll elaborate on the details of our agenda and the framework and ground rules for our dialogue, and give everyone an opportunity to share their thoughts, anxieties, concerns, expectations for the day. So let me begin by introducing Ellen Stover.

Opening Comments

Dr. Stover: Good morning and thank you. I have been actually involved now for quite a number of months with Peter and John in the discussions about how to actually convene such an unusual meeting. And, yet, at the end of the day, the end of a process come up with some concrete strategies. I look to the team here to hopefully relieve my own thoughts about that. I think one of the themes that our reorganization put forward, and you heard it in the range of Associate Directors that were mentioned, is that we try to look broadly across cutting kinds of issues. I have not ever been one to shy away from change and tackle rather major public health kinds of issues. I think from that perspective what I would like to see come out of a meeting such as this one, and we will probably undoubtedly have several more, are ways in which we might better utilize the kinds of methods and procedures, tools, knowledge that we have to understand these very complex social behavioral mental problems. Today we are going to zero in, once we get through our intros, into I think I think a very, very fascinating, area that relates to child development. One of the things that I have tried to foster, and you can see from the numbers of people just sitting around the room also, our connections between knowledge from our basic Neuroscience Division and links to our Service and Treatment Division.

Eventually, I would hope that the research that we support, and much of it does, actually, can be used by CDC and other governmental agencies to try to develop programs for people who have these problems. I would like to set a frame for us to think very broadly and with a vision, so that our work has great impact. One of the things Steve Hyman said in his first council meeting was we like to think about supporting studies that make major leaps, in contrast to very tiny increments. That is hard. That is very hard. It is hard to do, particularly, in terms of the kinds of peer review processes we are all used to, the type of funding and supporting systems that we are used to, the ways in which journals review articles, and the way in which our, basically, our whole system operates, so it is complex. The reason I mentioned peer review is that we are in the midst of a major, major change in the behavioral sciences. The review committees at NIMH will be merged. We are looking to a February '99 date. I will

not take time at this meeting to go into any details about that, but basically to say that the kinds of discussions that I expect will go on today I have a feeling are very similar to the ones that we internally are engaging in about how to best configure our science and evaluate our science, so as to have an impact very broadly.

I think I will stop there. I think basically what I came to do, I will stay the day, is set a tone of inspiration and support for this kind of effort, and to absolutely underscore that I want to make sure that we have the best input, the broadest kind of input, and to also welcome the staff around the room to speak up if they wish.

Dr. Richters: Thank you Ellen. Peter Jensen.

Dr. Jensen: Well, I am actually very excited about this, but I am not anxious. I am just not sure whether we will be able to get some good ideas that will actually be doable. Regarding some of the issues that have been raised over the last two decades, there have been mainstream criticisms about the methods and paradigms of our social and behavioral sciences. I think the null hypothesis significance testing issues raised recently exemplify these problems, and they are part of a larger problem and puzzle. So anxiety is not the issue --- certainly not for me. Rather, it *uncertainty* about whether we will be able to come up with some strategies for converging operations to go beyond the scientific approaches we have been using, i.e., those that rely principally on statistical methods that have linked the justification and discovery processes. We need to find some new means to arrive at converging operations to help us know when we know. How *do* we know when we know? This question touches on epistemology and philosophy of science.

So this is actually an exciting time, but I am not sure if we will be able to pull it off as we struggle with this problem and iterate potential solutions with the field. Thus, we expect this to be the first of a number of meetings, to see if we can develop standards that will be useful, and that we can apply this initially in a research domain where the National Institute of Mental Health has primary responsibility, under John Richters' leadership --- conduct disorder and antisocial behavior in childhood and adolescence. We do not want to go out on a limb with some new paradigm that is really a "bust". And so, if we can make it work in one research domain, and if it attracts some interest and excitement, and if it meets good scientific standards for justification, and if it leads to new discoveries in ways that show how we can move rapidly ahead scientifically, then this would be a very exciting accomplishment. So this is the challenge we face, and I am just delighted that you are all here and you will be part (hopefully) of an ongoing dialogue with us as we wrestle with these issues.

Dr. Richters: Thanks Peter. Kimberly Hoagwood

Dr. Hoagwood: Good morning. I want to make a couple of comments to put this meeting in the context of some larger social issues, and describe why I think this meeting is particularly important. Here at NIMH, we are in the business of both administering science, as well as doing it. That gives us a somewhat different vantage point on the whole scientific endeavor, particularly on the boundaries of science, and the management of uncertainty. Much of the time what we need to do is determine when we know enough about "X" (whether "X" is a particular human behavior, a diagnosis, a disorder, or a treatment strategy) to say with some degree of certainty that we recommend a change in public policy. We have a responsibility to do this efficiently and to do it accurately. Ways in which the scientific field makes judgments about mental illness, human behavior, or treatments matters a great deal. Part of what we are trying to do is look at the manner in which these judgments are made.

Are the modes of our reasoning as good as they can be? If not, can they be improved? What are the ways in which we are thinking about the discovery process, about theories, and theory generation, and how are we justifying and evaluating the generation of theories? Are there ways that we can compare our methods to strengthen our modes of reasoning? If our scientific judgments are flawed, this can lead to great harm. It is not just an epistemological problem. Errors in reasoning or unacknowledged limitations in judgment can lead to public policy that may be harmful. The responsibility for assuring the highest standards of scientific judgment, of course, has been around for a long time. This is not new. What is different now is that there is a certain intellectual climate that calls into question the notion of what's rational and what's isn't.

Let me give two examples of this. Neal Postman wrote a book a few years ago, called *Technopoly*. What he pointed out in this book is that extreme "statements of fact" are now believed. For example, I was in the xerox room the other day and said to a colleague of mine, "Did you see today's *New York Times*?" Well, the person had not. I said, "It turns out according to the *Times* that if you eat a pound of ice cream every day, it actually increases

your good cholesterol, lowers your bad cholesterol and you lose weight." This person believed this extreme "statement of fact." The point is that the notions of credibility and rationality are getting almost peculiar; people will believe just about anything, given particular contexts in which information is presented. The second example is this: Alan Sokal, a professor of physics, published last year an article in *Social Text*, a very prestigious journal of post-structural criticism. It was called *Transgressing the Boundaries: Towards the Transformative Hermeneutics of Quantum Gravity*. He argued, among other things, that dogma imposed by Western-dominated world views suggested that there exists an external world that has properties independent of human perception, and that these properties were encoded in physical laws. He argued that physical reality was really a social and linguistic construct, not our notions of physical reality or theories, but physical reality itself. He used quantum physics theory and mathematical sets theory in order to make his point. Then, a couple of months later in May/June of 1996, in *Lingua Franca*, Sokal published an essay disclosing that his previous essay was all a hoax. It was a joke! He did not do it just to make fun of people, but because he was very concerned about the way in which certain academic fashions had taken over the grounds of rationality. He was motivated by the fact that sloppy thinking was becoming acceptable.

These are not isolated examples, unfortunately, but they implicate our current intellectual climate. I think this is why this particular workshop is so important now. It raises the question about what gives rise to our habits of thought--our prevailing notions, traditions of science, and modes of justification. Why are those modes of thought sustainable for awhile? And what is really influencing these habits of thought? Should we take a moment to think about those influences that constrain the ways in which we justify our evidence?

Let me return to Neal Postman for a moment. He suggests that technologies themselves create habits of thought. For example, he points out that our numerical conception of reality was created by a technology. Prior to 1792, student papers were not graded; when a particular work of writing was handed in, it was not given a grade. In 1792 at Cambridge University, William Farrish first came up with the idea of providing a numerical rating to written work, so he quantified language at that point, and this was the beginning of a mathematical conception of reality that we now take as inevitable.

So part of the question is, what other habits of thought that we take as inevitable now are not inevitable? What are the technologies or other influences that lead us into certain ways of thinking about the relationship between theories and data construction? And should we be putting a pause button on certain technologies? Our technologies, of course, include things like statistical analysis packages, neural imaging, graphic displays, etc. Is it time for us to pause and look at the ways in which we are discovering and testing theories, or the ways in which we are creating new hypotheses about the data that we generate? Should we look at justification in terms of technology or has technology become our predominant system of social inquiry? Is technology itself actually directing the ways in which we are thinking about data? So part of what we want to do today is to think about the discovery process, the justification process, and the tacking back and forth between those two. What are the obstacles to progress in conceptualization, particularly, of antisocial behavior? That is the theme of this particular meeting, and with that I will turn it over to John.

Roundtable Agenda and Discussion Framework

Dr. Richters: Thanks very much Kimberly. It warrants underscoring at the outset that our objective today is not to propose specific solutions to deficiencies of the existing paradigm, but to formulate strategies for encouraging, facilitating, and funding reform efforts in the field. Although many of the paradigm's discovery and justification deficiencies are fundamental to the broad array of social and behavioral sciences, it is clear that their relevance, manifestations, and implications differ considerably across substantive domains. Thus, some of our time will be devoted to general cross-cutting problems, issues, existing obstacles to reform, and broad strategies for overcoming those obstacles-- realizing in advance that many of our ideas will be used as scaffolding for highly focused follow-up meetings in different substantive domains of research.

An overindulgence in generalities, however, would limit the constructive potential of our dialogue. So, to keep our conversation clear and practical, we'll also anchor our general discussion to some of the concrete realities of applying our ideas to a specific research domain. The study of antisocial behavior and conduct disorder in childhood and adolescence will serve us well in this regard; not only is it a research domain with obvious direct relevance to the Institute's public health mission but one in which the Institute's heavy investment over the past

several decades has yielded steadily diminishing returns. For both reasons, it is likely to be ground-zero for our initial reform efforts. Having said as much, I hasten to add that there is nothing unique about the paradigmatic deficiencies of antisocial behavior research, their undermining effects on progress, or the challenges they pose for bringing about constructive reforms.

The presenting problem in antisocial behavior research is a familiar one in many areas. More than fifty years of intensive research has produced a wealth of descriptive data concerning the correlates, predictors, sequelae, and life-course patterns of antisocial behavior. These findings have taught us much about the characteristics of children, their families, and environments that are associated with risk for developing antisocial behavior problems, the relatively more persistent and troubling nature of antisocial syndromes that emerge in early childhood, and the clinical and public health importance of focusing prevention efforts on this group. To date, however, these findings have yielded meager theoretical progress in our understanding of how the accumulated facts should be interpreted --- about *why* antisocial trajectories develop, *why* they broaden and deepen with development in some children yet taper off in others, and *why* antisocial tendencies tend to be so difficult to influence once stabilized. Although theories continue to proliferate, none has ever been more than weakly supported by the extant data, translated into cumulative theoretical progress, or given rise to powerful treatment or prevention interventions.

In the [Hubble hypothesis](#) paper I have elaborated in detail on how this failure is an inevitable consequence of deficiencies in the discovery and justification paradigms on which the lion's share of antisocial behavior research is based. For example, every stage of the discovery process-- study design, sample selection, subject recruitment, assessment and standardization procedures, data analysis-- is predicated on a faulty assumption that children and adolescents are interchangeable members of a single class in who differ quantitatively but *not* qualitatively with respect to the mental and psychological structures underlying their personality functioning and overt behavior. This assumption predefines the entire discovery process as a search for *the* right combination of putative causal factors, followed by efforts to model complex interactions among variables in the service of characterizing *the* unitary causal structure responsible for individual differences in antisocial behavior. The underlying assumption, of course, is invalidated by common sense and virtually everything we know about the unique plasticity and openness of the human brain and its extensive capacity for use-dependent and experience-based structural modification and elaboration within each individual. Nonetheless, the inadequacies of the discovery process resulting from reliance on this false assumption are obscured by the forgiving standards of a justification paradigm based on faulty logic, ritualistic hypothesis testing, and widely discredited misuses of statistical inference. Worse yet, it actually undermines the discovery process by imposing formidable statistical requirements for hypothesis testing, population generalizability, minimizing so-called type-1 errors, and maintaining low variables-to-subjects ratios.

Even in the growing discipline of developmental psychopathology, which is defined by its embracement of open system concepts and an explicit acknowledgment of the structural heterogeneity challenge, the undermining influences of the justification paradigm are devastating. 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 the closed system research strategies engendered and reinforced by dominant paradigm. The predictable consequence often is that researchers are pressured in subtle yet powerful ways to make often-unrecognized choices between adhering to developmental principles and perspectives *or* adopting research strategies based on the structural homogeneity assumption. A review of contemporary developmental psychopathology research reveals that sound developmental principles are embraced in the Introduction sections of empirical manuscripts, routinely ignored and/or violated in Methods and Results sections, only to resurface in Discussion sections. The end product typically bears little resemblance or fidelity to the very phenomena that made the study compelling in the first place.

The implications of these paradigm weaknesses for a meaningful dialogue about scientific re-tooling are extraordinarily challenging. In the context of **discovery**, a rejection of the structural homogeneity assumption perforce requires a rejection of the of the research strategies, methods, and assessment procedures that depend for their coherence and justification on its validity. The reason, simply, is that the unit of analysis in individual differences research is not an individual or group of individuals, but the sample itself. Individual data points are used merely to construct an average hypothetical individual who is characterized by the unitary causal structure presumed to be common to all members of the sample. An acknowledgment that phenotypically similar syndromes may stem from different causal structures in different individuals--- does not merely add to the complexity of studying causal processes and structures. It radically redefines the challenges of discovery in ways that can be engaged meaningfully only through significant retooling, an emphasis on exploratory research, reformulated research questions, and the development of new research strategies-- what Sigmund Koch called

'indigenous methodologies' -- with fidelity to the richness and complexity of human psychological and social phenomena. This, in turn, will require an appreciation for the virtues and necessity of embracing a guiding principle of 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, quasi-experimental, historical and ethnographic approaches. 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 discovery methodologies can be evaluated only with specific reference to the phenomena and research questions under study.

It is equally clear that meaningful reform will require far more sophisticated and defensible approaches to scientific **justification** that dispense with misuses of statistical inference, move beyond the currently narrow emphasis on 'variance explained', and incorporate widely accepted scientific criteria such as explanatory coherence, plausibility, explanatory elegance, and so forth. The justification challenge, then, will lie in translating principles such as these into workable self-correcting scientific engine that encourages and facilitates exploration and creativity while at the same time serving as a self-correcting scientific engine for guiding progress. I would urge us all in thinking about problems and formulating strategies for solutions today to grapple with the inherent interconnectedness of discovery and justification issues. Let me turn now to the business of the day and begin by giving everyone a chance to summarize their own priorities, concerns, anxieties, and expectations for the day.

Opening Comments by Invited Participants

Dr. Hinshaw Thanks very much for starting with me, John. Kimberly noted some anxiety about today's intellectual climate. I have a different anxiety about today's climate, with regard to the entire social and behavioral science enterprise. I think there is a realistic set of questions from the public, legislators, and many others, with a focus on the value and viability of the entire enterprise. Does it make any contribution? Is it worth funding, either at a basic or an applied level? I think that unless social and behavioral sciences can both promise and deliver in the ways that biomedical science has, I am not positive that over the long haul the public will tolerate it or fund it. I am not just talking about products. A "cure" for antisocial behavior will necessarily look quite different from a cure for cancer, given the highly contextualized nature and lack of disease-like nature of the former. Yet our basic and applied research can (and should) receive close scrutiny.

At a more fundamental level, since the time that I was an undergraduate, I have wondered about whether our topic of interest, which transcends antisocial behavior per se to include the broader family, neighborhood, and societal context within which such behaviors occur, were and are amenable to traditional science, to the paradigms that we use. Or, must we fundamentally retool and reconceptualize? It would take me a lot of time and probably futile effort for me to try to elaborate on the wonderful conceptual outline Bill McGuire E-mailed us prior to the meeting, regarding processes, causal structures, the need for more heterogeneous participant populations and ways of parsing them, variable scaling and measurement, verification phases versus both the creation and reformulation phases of research, the need for more complex designs, and the need to think about strategies for multi-experiment programs of research rather than just tactics about one particular study (I think that I have recalled the main points of his outline.) We could probably talk about any one of these points for the rest of the day, but I found this kind of distillation and organization crucial.

I would like to note some fairly concrete ways in which we can begin, I believe, to make some changes happen. One of those is through altering journal policy. One specific suggestion that I made prior to the meeting was that if journal editors are in some kind of consortium, there may be an explicit policy to promote, as lead articles, those that might be called "paradigm busting"--that is, different ways of reconceptualizing the phenomenon of antisocial behavior. These could be featured at the front of journals, and regular paradigmatic science could take the place of "brief reports" towards the back. Also, an issue that I have raised with John a number of times relates to the common consensus that there is not one set of unified processes that underlies the behavior pattern under consideration. However, what is the right number of subgroups? Two (early vs. late onset), as Teri Moffitt has said? Or is the number 200 or 2,000? Or is every individual so unique that categorization is not possible? In other words, what is the correct, finite number of functionally homogeneous subgroups--at a level of affect, at a level of biological process, at a level of gene- environment interaction (or any other salient variable)?

One of the fears I have raised to John is that if you take the [Hubble hypothesis](#) to one of its logical conclusions, we will have devolved back into case studies as the end point of our science, with no room for important underlying mechanisms that cut across phenotypic dissimilarity.

At a somewhat practical level, it seems to me that one, maybe not too distal outcome of a meeting like this would be to have NIMH come up with some initiatives, putting dollars where their mouth is. One of these needs to be in the language that is now in all NIH proposals about adequate gender and minority representation. As a grant reviewer and writer, I realize that in the vast majority of cases this section can become a rather empty exercise. Without enough females, or ethnic minority individuals, or representatives of low SES groups, there is not sufficient statistical power to perform either separate analyses by subgroup or "moderator" analyses, which would examine the interaction of these variables with the predictors of interest. The final point I would put forth for us is as follows: Are we simply going to fine tune from a meeting like this--or from the series of meetings that emanate from this one--or are we going to be more daring and make some fundamental breaks of set? I would challenge us to think about the latter, rather than former.

Dr. Richters: Great, thanks Steve. Ed?

Dr. Trickett: Thanks for holding this meeting, John. It is a nice exercise in response risk taking. I come to this meeting from a background in community psychology. From that perspective, I have been interested in the paradigm issues with respect to the degree to which psychological research decontextualizes the human experience. In going over the different E-mail correspondence preceding the meeting and some of the meetings, there are a half dozen I think, at least sources of tension, sort of a yin yang kind of dialogue that I think are relevant to furthering that kind of agenda.

The first, and these are not in order of importance necessarily, but the first is the difference discussed between prediction and understanding. We can look at all of these sources of tension with respect to how I have sort of been thinking about discussing their differences in forums and understanding of the human context of behavior and how that is related to what we think about -- what we know. So the first I think is the distinction between prediction and understanding. The second, which is related to that, are the discussions between qualitative and quantitative ways of knowing. Whether or not they are fundamentally the same or different, I think is an interesting question. I think the multiple methods issue is a very important one around measuring and not measuring. Third is the tension that Mike Maltz mentioned in his paper about the issue of the homogeneity, heterogeneity of constructs, whether or not, for example, if we talk about children of divorce or single mothers, whether or not we are talking about them as homogeneous groups, and whether or not trying to unpack those concepts helps us understand the social context within which those different groups live as part of the phenomenon of interest. The fourth I think is Bill McGuire's interesting hypothesis that all hypotheses are true, itself a hypothesis which I guess would have to be true, given Bill's premise. For me, from the perspective I bring to that, it pushes me to think about the ecological constraints on findings and trying to figure out where in the world we can find examples of where the hypotheses that are true here are not true there. The fifth tension I think is John Richters' [Mark Twain Meets DSM](#) thesis about the tension between a sort of medical paradigm for understanding behavior and a coping and an adaptation paradigm that results in quite different attributions of why people do things and where the locus of influence is.

A final is just sort of reemphasizing what Steve Hinshaw said about the way our general practices are constrained and decontextualized in the research enterprise. As editor of *The American Journal of Community Psychology* I have had numerous examples of people who are trying to tell rich, complicated, integrated, qualitative, quantitative stories with numbers. It always is a frustration to sort of figure out how to help them tell the story they want to tell while they are at the same time fearful of not having a journal be responsive to alternative ways of thinking about it. So those are the half dozen things I think about.

Dr. Richters: Thanks. Dan?

Dr. Robinson: Well, John, well done and thanks for having me. I come to this with a kind of schizophrenic background. My doctorate is in neuropsychology, which means I am perhaps one of the few people in the room who knows that the Isle of Reil is not in the Caribbean, but most of the books I have written over the years have to do with the philosophy and history of psychology, and more generally science, and with a particular interest in the classical world, and more particularly Aristotle himself.

So if I regale you for a few minutes with things that are old, I also have some things to say about things that are not so old. This issue goes back to the *Posterior Analytics* where Aristotle raises the question, how do we best explain anger? And he says, "well, if you are a natural scientist, if you are a 'phusikost', you explain it in terms of changes in the temperature of the blood. And if you are the 'dialectikos', then of course you explain it in terms of one thinking he has been wrongly slighted. The explanation you offer really depends upon the kind of understanding you are trying to achieve." I just want to say this so that we do not think that we have to reinvent the wheel. Let me jump ahead a couple of months to a question that was addressed to Helmholtz. He was asked by a friend to account for the radical divorce between philosophy and science, why it is that people who had been engaged communally in this enterprise just years earlier now did not even talk to each other. Helmholtz, who was not given to rash statements, said that he thinks it came about when the scientific community had reached the settled position, that the Hegelians were crazy. Now what got the scientific community to think that the Hegelians were crazy! Well, Hegel, in the patrimony of Goethe and the whole Romantic movement, reached the firm conclusion that science was a one-sided, narrow sort of thing. Goethe had written this long treatise against the Newtonian theory of optics, insisting that this theory explains everything except what we actually see. And so there was a quite dramatic departure, and it was rendered official, and even dogmatic, by Ernst Mach, who in the *Analysis of Sensations* argued that you know you are doing science to the extent that you are *not* doing metaphysics. The entire project of science is to eliminate everything that is metaphysical. He argues this in *The Analysis of Sensations*, where he insists that all we mean by a law of science is a systematic re-description of experience.

When the Vienna Circle formed and finally started distributing papers for public consumption, the first meeting held was dubbed Verein Ernst Marc, so we quite understand the pedigree of this set of ideas and precepts. Now there was a maxim that was ancient even as Socrates lived, and in the Greek the maxim was *polis andra didaska*, "Man is taught by the polis", we are shaped by the political context in which we find ourselves. As the public expression of rationality is the law itself, the kinds of entities that you get, the kind of civic beings you get will depend heavily, if not entirely on the overall political framework.

Well, I became interested just a couple of years ago in what the social and behavioral sciences were prepared to tell us about "civic development". I had not spent time with the developmental literature, and I thumbed through as much as the university library had, and could not find anything on civic development. It seemed odd that people could tell me the age at which children burp, and walk, and make funny sounds, and so forth, but on the core question of what it is a society or a family must do in order to create *civic* entities capable of sustaining a particular kind of culture, a particular moral context, there just was not anything there. Well, I obtained a small grant to host a series of lectures and seminars at Georgetown on this topic. I did not know people working in the field. David Crystal, my young and trusted colleague, did know some of them. We had six or seven come in, two of them from psychology.

One, Professor Torney-Purta from University of Maryland, who directs a project on multinational comparisons; and Connie Flannigan from Penn State. They both did estimable jobs, and I learned much from them. But there really was not anything on civic development proper and I am not quite sure that the best approach to that subject would be by way of the *Isle of Reil*! So, what happens when you develop an anti-metaphysical bias and understand that you are doing science only to the extent that you are not dealing with troublesome and vexatious variables, is that you are likely to leave outside the laboratory door precisely the set of issues that got you interested in the first instance. End of lecture.

Dr. Richters: Thank you Dan. Bob?

Dr. Cairns: There was a quote in a recent volume on developmental science that reads, "Recognition of the complexity of development is the first step towards understanding its essential coherence and simplicity." So I do not think we need to be afraid of looking at the multifaceted problems that are here. In the preceding comments I did not see much about developmental novelty. More generally, there is a tradition in psychology and sociology to get rid of developmental change at the first step of research. It is either done in the design by controlling for age, or eliminating age variation, or it is done afterward in the formulation of transformations (e.g., transformations or latent variables). The essential function of these transformations and latent variables is to stabilize an unstable system. That is fine, because that permits us to get along with our traditional methodologies, structural methodologies, but it does create havoc when you look at the essential nature of evolutionary and developmental change. It is on that score I think we have gotten ourselves into an awful lot of trouble by postulating an antisocial

trait that can be identified in infancy and tracked throughout development. When, in fact, there are new things emerging, new forms of aggression, new forms of violence, just as in normative case menarche reorders a whole range of variables. There are true novelties that occur ontogenetically and by disease process. Most of our models do not take into account or even permit the introduction of novel dimensions in a regression equation.

The issue that John raised about homogeneity of subsets is absolutely critical, or, put in another way, the lack of homogeneity in the total sample. Several of you have pointed out that even if we succeed in identifying homogeneous subsets or clusters of individuals -- as opposed to clusters of variables -- we remain with the problem of translating between those subsets and the coherence of individual functioning. Individual integrations may occur in distinctive ways that are not classifiable across the subset. It is simply a matter of working out the translation mechanisms between individual trajectories, sub-clusters, and the total sample. Now how are you going to do this methodologically? Bill mentioned the use of cross-national samples. That is a nice way to do it, or subsets within the country. Or we can think evolutionarily, looking at comparative models and animal behavior models. Here we do indeed have the capability of manipulating directly and formulating homogeneous subsets, genetic and otherwise, and intervening developmentally. Which, by the way, I think is going to be the key to using multiple sources of information to arrive at reasonable conclusions about the nature of negative phenomenon.

Our solutions are not necessarily going to be monolithic. I think we may be in a lot of trouble to try to force one model to all phenomenon. There are some phenomena that are inherently multi-determined, such as lives over time. In these cases, any sane or even insane hypothesis should be able to reject the null hypothesis, simply because we have multiple contributions to the phenomenon. Under these conditions we do need to disaggregate into clusters and subsets of subjects to look for coherence, and to get into the stepwise process from individuals to groups. Other domains, such as self-reports, or perceptions, or perhaps attitude may not require that kind of research strategy.

Dr. Richters: Perfect, thanks Bob. Geoff?

Dr. Loftus: Let me make a number of points that are only loosely connected to one another. I will try to put them more or less in an order of most concrete to most abstract. First, I have been obsessing for at least the past couple of days about what we can actually get accomplished in a one day meeting. I used the analogy when I was talking with John this morning that our task is sort of akin to that of a minnow trying to change the course of an aircraft carrier. The question is how do you do that.

The first thing is to figure out the points of leverage. Returning to my analogy, the first point of leverage is, if you are the minnow, you try to recruit the tugboats. Translating the analogy to our task means trying to figure out how we are going to communicate with people like textbook writers, journal editors, teachers --- in other words, the people who really functionally cause things to happen and determine the course of research in our science.

The second point of leverage is to provide suggestions for what we should encourage in grant proposals. That's something that NIMH obviously has some degree of control over. In my E-mail correspondence I had a couple of suggestions. First, force people-- well, maybe not force people, but at least strongly encourage people-- to write grant proposals that have very specific predictions for the experiments they propose; perhaps even suggest using something like an Excel spreadsheet to simulate their entire experiments. A simulation forces a grant writer to do many things-- to come up with an explicit model of error, be very explicit about what the hypothesis is, and maybe in the process see that the hypothesis is stupid and should not have been entertained it to begin with.

My second point has to do with inferential statistics. It will probably surprise nobody who has read anything I've written to discover that I think that there should be an explicit de-emphasis on analyzing data using standard hypothesis testing. Perhaps such de-emphasis could be encouraged in grant proposals as well. My third point is that we should use this meeting at least in part to try to figure out a structure for additional meetings that would be for purposes of education, or doing what we are doing now, except at a more detailed level. My fourth point is that, if we are going to try to have a broad effect within psychology and the social sciences, we should try to determine the degree to which the problems that we articulate apply to various subareas within psychology. Personally, I know very little about the explicit domain of focus here, developmental psychopathology. I, in my own research, straddle two areas -- psychophysics and memory. In my view, the problems that have been raised do not apply quite so much to the area of psychophysics. They apply much more to the field of memory, so in general we ought to perhaps be a little explicit about which subareas within psychology are sorely in need of help and which are not.

My last point, which is perhaps the most abstract, harks back to Steven Hinshaw's comment that the final logical conclusion of the Hubble hypothesis is that we are down to a n -of-one situation, which does not seem to me to be good. We should still be seeking ways of trying to do what actual classical statistics does, namely, ferreting out what contribution to some dependent variable is universal, and what part is determined by error --- let's say, for the sake of argument, error inherent in an individual subject. At present, doing that within the domain of classical statistics is very naive: you assume an additive structure of the "real" component, plus some "random error". I think, however, that there might be ways of accomplishing the general goal of figuring out what is real and what is attributable to individual subjects in ways that are much more sophisticated. If I had another 10 minutes or 15 minutes, I would give you a specific example of what I mean by that.

Dr. Richters: Thanks, Geoff, and you *will* have more. We are going to open it up very shortly. Ken?

Dr. Rubin: Thanks, John. Thanks for inviting me. Late yesterday in the afternoon I finally put together an E-mail to you, and I do not know if you received it. I figured you did not, so I will share some thoughts with all of you. It will not be an attempt to be abstract at all. These are really simple sorts of notions. One thing I would like to touch on is multiple meanings of behavior. One of the questions that we deal with in our research is: *Does a given behavior carry with it the same meanings within and across venues?* Children can behave in a given manner, X manner, in the same context for many different reasons. And, of course, children can behave in a given manner in different venues. In some venues, the behavior is seen as normal, and in others it is seen as not so normal.

Let me give you an example, one that Nathan Fox and I deal with on a regular basis. We can observe children playing alone in a particular setting. In some cases, alone behavior, solitude or isolation may be caused by social fearfulness or anxiety; in others a personal preference for solitude; and yet in others by notions of being rejected by those around you. Another example, in some venues solitude is called for and expected (e.g., school work or working independently to solve a problem). In other venues, it is unexpected and viewed as abnormal, as in, for example, when instructed to play a game with two other children, a focal child hides in the corner of the room, or chooses to do a crossword puzzle instead. The point is that any given behavior can have different meanings within and across settings.

Another significant factor, and I think Bob Cairns has alluded to this, is the notion of the age-appropriateness of the given variables that we study. Researchers have begun to define psychopathologies in infants and toddlers. I am going to say this rather harshly; they have done so with little understanding of developmental norms or normative expectations. Oft times their work is spurred on by parental and cultural beliefs, or perhaps inappropriately -- a scientist's inappropriate beliefs about the meanings of given behaviors. I will give you one example here. The frequency with which toddlers initiate disputes over objects with age mates during free play happens to be associated with parental ratings on the CBCL of externalizing disorders. Yet, at the same time, we have to ask what one or two-year-olds are all about-- what are they trying to accomplish in their lives?

One of the things they are trying to do is define themselves and their territories and their possessions. One of the first words toddlers learn is *mine*. John, you have a 2 year-old daughter, Katharine, and I assume that *mine* is one of the words that she uses with regularity. If other words do not exist in the repertoire, one can hardly expect a two-year-old to meander over to a peer and request, "Hey, may I have that please?" So disputes over objects occur. It turns out actually that the frequency with which toddlers initiate disputes over objects with age mates during their free play exercises is associated with behavioral indices of sociability and social competence at age four. It does not predict parents' CBCL ratings of externalizing problems at age four. So I do not know how to interpret the meaning of conflict initiation and maintenance at age two, even though it does correlate contemporaneously with parental ratings of CBCL externalizing problems.

Third, we need to attend more carefully to cultural meanings of behavior. This is something that my colleagues and I do a lot of work on. A number of a years ago, I was given a tour of a new children's hospital in Beijing. The guides included the wife of a then vice premier and a number of senior developmental psychologists. As we approached the library, the guides told us they were very proud that they received all of the most prestigious of Western journals; we could actually see these journals on the shelves. Indeed, the only journals that I saw were Western journals, and the only textbooks I saw were of the Western ilk, all English publications.

I thought to myself, but did not dare say otherwise, that here was a culture with a large percentage of the world's population believing perhaps that the findings of researchers in the West could bring with them their

meanings to another culture. I want you to think about the following: An eight-year-old girl walks into class in a North American community wearing slippers. In class she focuses on her desk and the books planted thereon. She neither looks around the room, nor engages classmates in conversation. At some moment of classroom discussion she belches and fails to excuse herself. She especially enjoys discussing mathematics with her teacher after school. How many of these descriptive units would be viewed as typical in an American third grade classroom for a girl or a boy?

Dr. Robinson: Sounds like Oxford.

Dr. Rubin: Perhaps, so. But I want you to think about a newly arrived immigrant from the PRC. Would the behaviors be anymore understandable, and to whom? Certainly not to peers, and probably not to teachers either. Researchers are now beginning to understand that given behaviors take on very different meanings across cultures, as if given behaviors lead to different meaningful inferences and attributions across cultures. The correlates and concomitants of given behaviors differ across culture. Yet, we assume that which is deemed as abnormal in the West must be similarly understood in other cultures. I will wrap up. Let me just finish with one last bit. According to Jerry Kagan, who unfortunately was unable to be here today, approximately 10 to 15 percent of toddlers may be characterized as behaviorally inhibited. They present as fearful in the face of novelty. Our research indicates, however, that if Western norms were used to identify inhibited children in the East, over one-third of PRC toddlers would be so characterized. Further, from about the age of seven, social fearfulness in the West is associated with such phenomena as peer rejection, negative self-esteem, victimization by peers, loneliness. In China, the exact same behaviors are associated with excellent scholarship, peer popularity, and leadership.

On a recent visit to Sicily, so now I will move South, rather than East, I arrived in a lab where my collaborators were assessing behavioral inhibition among two-year-olds, following procedures used by Kagan and by our group at the University of Maryland. The child I observed was the 30th they had seen in a period of six weeks. Not a single toddler of the previous 29 could be characterized or "diagnosed" as inhibited by our standards, not one. On the day of my visit, the 30th toddler began screaming as soon as the first adult stranger entered the room. She was afraid of toys, she was afraid of people, pretty much afraid of everything. The researchers ran into the observation room where I was looking through a one-way mirror and they moaned that they were so embarrassed by this child's behavior. They had never seen such a child before, and they asked what should be done with the child. Should this session be ended? Does the family need counseling? My immediate thought, which was rather perverse at the time, was that the correlation between the frequency of toddler initiation of object conflict would be negatively associated with any index of behavioral or emotional dysfunction in Sicily.

Cultural values, my own in this case, perceptions and beliefs are simply not well understood. The significance of all of the above that I have said is that in the USA-- I am a Canadian and I just arrived here not long ago so I can say this-- the American federal granting agencies do not fund cross-cultural research, at least not that I am aware of. Yet, every year hundreds of thousands of immigrants arrive in this country and we have virtually no understanding of that which is normal, and that which is abnormal in other cultures.

Dr. Richters: Thank you Ken. Mike?

Dr. Maltz: I would like to make a few comments at this point. First, as Kimberly mentioned, we hear a lot about the anti-science attitudes of the general public. But I don't think that this is a monolithic attitude. If we take a look at one of the most popular programs on television, one that is shown and viewed daily by most Americans, we find that it is filled with data analysis--in fact, very complicated multivariate analyses. It's called "the weather report," and displays data showing at least the following variables: latitude, longitude, time, precipitation, wind speed, barometric pressure, and in some cases altitude. Although the overall goal is to *predict* the weather, what is more important from the standpoint of the viewers is to get an *understanding* of the weather from their own standpoint. For example, a prediction that Chicago will get on average three inches of snow means less to me than the dynamic radar picture I see showing the precipitation heavier on the south side. It helps people get an understanding of the interrelationships among the variables that comprise that very complicated phenomenon called weather.

I think that there's a lot that can be done in this area--visualizing data. Mark Appelbaum and I talked about visualization in terms of discovery, and more can be discovered in data from its visualization than from algorithmic analyses.

Another comment I'd like to make has to do with a possible new program that NIMH might fund to support new analytic approaches. What about a program in secondary analysis? I think that we are getting better and better in collecting better and better data. The problem is that there seems to be a tendency to use more sophisticated tools to analyze the data, and that seems to get us further and further away from the reality of the data. The new methods are more complicated and more complex--to tell the truth, I don't understand them that well. We have a number of journal editors around this table, and they probably are in the same fix that I'm in--I can't make head or tail of the stuff that comes into my journal half the time, because they couch it all in LISREL or HLM or whatever the newest analytical approach happens to be.

So I think that we have a real paradox here--we have better and better data, we have (supposedly) sharper and sharper tools--but they are unable to give us any greater insight. One of my gurus, Richard Hamming, who headed the Bell Labs computer center, wrote a book, *Numerical Methods for Scientists and Engineers*. The book's epigraph was, "The purpose of computing is insight, not numbers." Of course, in the second edition I believe he changed the epigraph to "The purpose of computing is not in sight." I think that we lose sight of the fact that insight is the important thing--not prediction, but understanding. I would rather have an understanding of the weather than a perfect prediction of it.

Dr. Appelbaum: Do you live in Chicago?

Dr. Maltz: Right. You are just jealous because you look out of the window and see the same damn thing all of the time.

Dr. Appelbaum: And it *is* great!

Dr. Maltz: Another point I want to make concerns the null hypothesis significance test. It is like the Energizer Bunny: we keep on trying to kill it, but it keeps going--and going--and going. Why? Because we haven't found how to break into the cycle; it's like trying to break the cycle of poverty. We have journals; journals have reviewers trained for the most part in hypothesis testing; articles go out to these reviewers; we feel obligated to publish something that has all good reviews; these articles find their way into textbooks. How can we break into that kind of cycle? I don't think that we can break it that easily. I think that it was Kuhn who said that the new paradigm doesn't take over until the promoters of the old paradigm die off.

Dr. Richters: Right -- not until there are successful exemplars out there to move to.

Dr. Robinson: Carl Hempel died two weeks ago, by the way--I mean, if anybody thinks we are now free to change the paradigm.

Dr. Maltz: Well, again this speaks to secondary analysis of the data. Because it's been my latest research kick for the past ten years, I also think of visualization of the data.

Let me see, there was another point I wanted to make. A colleague of mine sent me some papers that he felt were relevant to our meeting. This is a quote from one of them, written by Andrew Abbott of the University of Chicago Sociology Department: "One of the central reasons for sociology's disappearance from the public mind has been the steady deprecation of description in sociology. The public--and here I mean not only the reading public but also the commercial sector--really wants to know how to describe society. It wants to know, to put it most simply, what is going on. But such descriptive knowledge has been steadily despised in mainstream sociology for at least twenty years. Our narrow focus on causality has long meant that an article of pure description, even if quantitatively sophisticated and substantively important, effectively cannot be published in our journals." So this problem is not just found in psychology, but it is in all of the social sciences.

Dr. Richters: Thanks Mike. Bill?

Dr. McGuire: I have three methodological hobby horses on which I am riding into this meeting - - three complaints about the way we live now, each correctable although not easily so. My first complaint is that methodologists (at least in the behavioral sciences) think about and teach the research process as if it involves

only the critical, hypothesis-testing, verification, aspects of inquiry; while they largely neglect the creative, hypothesis-generating, productive aspects. Our methodology books and courses seem to forget that to cook a rabbit stew we must first catch our rabbit. Before we can test a hypothesis we must first catch one to test.

All researchers will probably grant that there is a vitally important creative, hypothesis-generating aspect of the research process. Its neglect in our methodology discussions may reflect the feeling that these creative aspects cannot be taught or even described. Not to worry. I contend that the creativity of scientists can be taught and fostered by interventions on both the macro and micro levels. On the macro, institutional level, we know something about how scientific labs (from small University groups to large NIH groups) can be more productively organized (as to size, homogeneity, centralization, etc.). Frank Andrew's (1979) *Scientific Productivity* UNESCO study of 1,200 labs is an example of an actuarial, quantitative approach on the macro level; and for an example of the anecdotal/case-history approach on the macro level see Latour & Woolgar's (1979) *Laboratory Life* account of a research group at La Jolla's Salk Institute. On the micro level, individual researchers can be taught techniques that enhance their creative hypothesis generating prowess, such as the 49 creative heuristics published in my [1997 Annual Review of Psychology chapter](#)

The second hobby horse on which I'm riding into today's discussion deals with a second imbalance in methodology, that our discussions of the research process are confined almost entirely to tactical issues in research, to the neglect of strategic issues. The methodological issues we cover are all on the tactical level, within individual experiments (such as manipulating, measuring, or controlling of variables, data analysis, etc.), while we say practically nothing about broad strategic issues that arise in designing multi-experiment programs of research (such as how to select problems, where to begin, how to proceed, the chunk size of individual experiments, etc.).

The third hobby horse I am riding into today's meeting is to urge use of a wider range of criteria in judging the acceptability of scientific theory. A single criterion, namely, that the theory's derivations survive empirical jeopardy, has dominated our methodology writing (if less our thinking) since Popper and the Vienna Circle logical empiricists, or even since Comte's positivism. There are other useful criteria, some intrinsic to the theory (internal consistency, novelty, completeness, open-endedness, elegance, prodigality, etc.) and others extrinsic to it (derivability, Establishment acceptance, épater le bourgeois, subjective feeling, pragmatic value, heuristic provocativeness, etc.) as I have described in my [1989 chapter on perspectivism](#). It need not embarrass us, indeed it can be an added attraction, that some of these desiderata sound slightly outrageous or mutually contrary.

Just before this morning's meeting was called to order, Kimberly Hoagwood and I were discussing the likelihood that, if scientists actually did select among theories solely or mainly on the criterion of empirical corroboration, then the Copernican heliocentric theory would hardly have prevailed over Ptolemaic geocentrism. Even a century or two after Copernicus rushed into print, Newton still had to use fudge factors in successive editions of his *Principia* to establish that lunar motion was more in accord with helio- than geo-centric theory. Either helio-centrism or geo-centrism, or placing the center of our system at any other point in space, could account for celestial observations. Were we to clone Ptolemy today and give him all the eclipses he needs, he might still save the appearances better than the heliocentrics. Scientists' preferring heliocentric theory to geocentric used criteria other than which better fits empirical observation. We need to give more thought, descriptively and prescriptively, to these alternative criteria.

So this morning I'm pushing for at least these three methodological rebalancings of behavioral science: (1) we should pay more attention to the creative hypothesis-generating processes in research, rather than just the critical hypothesis-testing processes; (2) we should wrestle more with strategic programmatic issues in research rather than just tactical within-experiment issues; and (3) we should recognize and exploit many additional criteria for preferring one scientific theory over others, and not pretend that the survival of empirical jeopardy is or should be our only criterion.

Dr. Richters: Thanks Bill. Mark?

Dr. Appelbaum: I am not quite sure why I was asked to be here. I suppose it is because I interact with many of the fields represented here at some level or another. And probably the fact that I am member of the [American Psychological Association's Task Force on Null Hypothesis Testing](#) (APA Task Force) as well as Editor of *Psychological Methods* might have something to do with it. Before I offer some comments I do want to make it clear that I am musing as an individual -- not as a spokesperson for the APA Task Force nor conveying any policies with regard to *Psychological Methods* or APA in general. If I could use my five minutes really effectively, I

would have Mike give his talk again. I think the issues of examining data and understanding what nature is trying to tell us is critically important. All the rest of these issues are in service of our trying to figure out nature. If we forget that understanding nature is our goal, I fear that these other pseudo-methodological issues (e.g. null hypothesis testing) take on a life of their own and divert our attention from the really important issues.

But since Mike has already said this.. and I have said it again.. what else? Down here I have many folders containing the papers that were sent in support of this conference. There are several folders like this one. I would guess that this amounts to about three years worth of output, that is pages, for a journal like *Psychological Methods*. All of this has been said, and all of you have said it¹ and it has been said over and over again. Twenty years ago when I was a graduate student, we were reading papers on the problems with null hypothesis testing, the impossibility of prediction of individual behavior, and whatever. If all has been said before, why hasn't anything happened? Why do I receive submissions which say the same things that have been said since the 1930's? First of all, I think that our training in methodological and quantitative issues has virtually collapsed in the field of psychology. I am really appalled by what is happening. I, for instance, have approximately 20 weeks, six hours a week, in which I am supposed to convey the totality of not only statistical knowledge, but the methodological knowledge of the field. We have moved into a training system where the number of required graduate courses in psychology has been cut in half in less than 25 years. We have a model now where what we do is to bring graduate students in, give them minimum possible course work, and put them into labs, mainly in an apprenticeship model. What happens there is that the next generation of psychologists is seeing people who are making the errors that you all have written about in all of these papers now training another generation of students in exactly that way - there is no progress being made.

When most of us were grad students we were getting opposing views because we were doing things like pro seminars. We were being trained by a broadly based cross-section of the field. We were often expected to work in several different laboratories. We heard many voices and not all speaking in the same voice. I do not think that what we are doing is irrational, it is just unfortunate and costly. That is because what we are doing is responding to a reinforcement system which I think is basically warped, but I do not think you can do much about it. Why do we get students into labs so quickly? Because we know damn well that if they do not have six papers published somewhere by the time they get out, there are not going to stand a very good chance of getting the kind of job that we think students of ours should have. And so we pull them into the lab very early. We have them doing research, analyzing data, writing papers before they really have a clue as to what they are doing. Then we go a step further. We create new journals so that there will be a mechanism for publishing the increased volume of work. God only knows how many journals there are now, so that there is room to publish all of these papers. And having more papers published increases the number of papers that have to be published in order to get the job we want our students to have. And so the spiral escalates. And then we are surprised when the system gets funny. We have created a system we cannot change.

This pattern is not just one of graduate training. We have done this same thing in grant getting side of the profession. I sat on a study section for eight years, so I cannot say I am pure and good - you get into the system and maintain the system. I was recently reading a review that a colleague of mine got. Among the things that were said that particularly caught my attention was a sentence from one the reviewers [Under the new system you see the actual words written by the reviewers - not a sanitized version created by an Scientific Review Administrator]. What was said, as a negative comment, was they are not using "cutting edge" methods. Now notice they did not say they are not using the right methods. They did not say they are using inappropriate methods. They said they are not using "cutting edge" methods. And so what are we supposed to do? Any rational person is then going to write their proposal in the most God-awful LISREL terms, or whatever cutting edge is at the moment, with little regard to what is actually the appropriate thing to do. Then these analyses get done and they get published and we are surprised and worry what the field is coming to. Can we really expect anything other than what we are getting?

Let me go on to my third thing and then I will shut up and talk later about the APA Task Force. I think there is a fundamental misunderstanding about what quantitative methods do versus what we want to have done. In hypothesis testing the only thing we can do is to talk about the probability of the data we obtained having been obtained if a certain theory is true. That is we talk about the data conditional upon the theory.. We get data that looks like this. We talk about the probability of getting this data if A were the case versus if B were the case. We compare the two and if one probability dominates the other we conclude that the one theory is better than the other theory. We can say that A is better than B. We can never do very much about saying whether A is very good - just that it is better than B. What we really want and what no one knows how to do in terms of statistical and quantitative approaches, is the probability that the theory is right given the data. This is what we want, what we get is different. I will argue that assessing the theory given the data is what we as creative thinkers do. We look at the

data and try to generate a theory that we think optimizes that data. If we do not look at data, if we do not do what Mike Maltz has talked about, then it seems to me that we end up not doing what we as scientists really want to do. We need to look at data, we need to do secondary analyses, we need to look at historic data as well as generating new data. Through this process we may generate worthwhile hypotheses and theories - ones that we can then subject the theory to the inferential science of replication, of saying how well does this explain data over a wide range of situations, over a wide range of sub-populations, etc. I will now shut up.

Dr. Richters: Thanks, Mark. Not for shutting, but --

Dr. Appelbaum: You'll be doing it before we are done.

Dr. Richters: Nathan?

Dr. Fox: I am a social developmental psychologist and I have written -- I guess what is most relevant for today's meeting, I have written quite a bit about the number one driving model, that is in our field for social development, and that is the model of attachment. And so I would like to raise five issues which I think are relevant for the discussion today. The first, I think, was from Bill McGuire's email-- that there are no rules of disconfirmation. We have a theory in social development called attachment. It does not matter how many studies one reviews which disconfirm this theory or parts of the theory. It does not die. It is like a vampire. You can put a stake through its heart. You hang garlic in your windows, but it keeps coming back raising its head, no matter how many times you think you may have stamped it out.

As a journal editor, I am constantly involved in debate with individuals who send papers to me which are disconfirming of the attachment model. Yet, the entire introduction and discussion sections of their papers laud and praise the theory of attachment and say how important attachment is for understanding social development. Along with this problem is the issue of the file drawer problem. There are numerous individuals who cannot publish their papers which show disconfirmation.

The second issue is the one which deals with homogeneity versus heterogeneity. Attachment researchers classify individuals into four different types. These four types of individuals are viewed within-category as homogeneous types, and there is little recognition of the possibility of within-category heterogeneity. With no data to confirm within-category homogeneity and the likelihood that between-category differences on external measures are not great, the classification system fails. A third problem with the the attachment model is notion of determinism. The attachment model argues that the same underlying motivations that elicit infant crying for its mother will motivate an adolescent to bond with a friend, or motivate an individual to marry a particular person, and will motivate an adult in the way that they will parent their child. The lifespan notion is inherent in many models of development. Some developmental psychologists believe that there is something about that initial bond between mother and which determines social and emotional development subsequent to that. That is a very powerful notion in developmental psychology. However, there is little if any evidence to support this position.

As a journal editor, I can tell you that the first paragraph of every paper examining social development cites Arnold Sameroff and Michael Chandler, and use the words *transactional model*. These papers advocate the transactional model. They seem to mean that there is biology, there is environment, and somehow they interact. Now it is clear what the transactional model is and what it means. However, if we are not able to parse behavior into its components we will be forced to? everything is an interaction, and that ultimately does not help us to explain behavior.

Finally, there is a positive bias to traditional attachment theory, which has been lost in recent more analytic work. That bias is the notion of evolutionary functions in terms of trying to understand ultimate causes in behavior. When John Bowlby wrote his three volume treatise on attachment theory, he creatively tried to use evolutionary theory to understand early motivational patterns of behavior in infants. This approach has been misused by our field. There are those who would use an evolutionary adaptation explanation for diverse behaviors such as homicide or teenage pregnancy. It is impossible to have any independent corroboration of the truth of these causes.

I would like to finish with one problem, which I have and which I -- it is a comment on Stephen Hinshaw's first point. Steve said that he questioned whether or not the public will ever -- will continue to tolerate social and behavioral sciences research, and he said that he doubted whether that is true. I take exact opposite tact. I think

that they will forever love us. My reason is also our problem. When I was a graduate student I heard a lecture by Stanley Schachter. Schachter told the story of when he was a graduate student and he came to his grandmother and his grandmother said, "Well, Sammy, what did you learn today?". And he says, "Well, grandma, I learned that when people live near each other they are more likely to know each other and be like each other than when they live farther apart." And his grandmother's reply was, "For that you have to go to graduate school, Stanley?" And it goes on and on. The point is we are engaged in a discipline in which we create and use terms which the public use and love. As long as there is intuitive ours will be a different science, because we cannot rise above those particular terms and terminology which we have created for ourselves. So, in some sense, we have created our own monster. I personally do not know how to kill it, but I would be interested to hear how others would. I am thinking maybe it is a good thing. We should adjourn, or at least recognize individual differences in bladders.

Dr. Richters: Thanks Nathan. This is probably a good time to break, so let's reconvene in 10 minutes sharp. *Whereupon a brief recess was taken)*

Roundtable Discussion

Co-Chairs: John Richters & Kimberly Hoagwood

Dr. Richters: All right. We will reconvene.

Dr. Hoagwood: Part of what we want to do next is hone in on the problem of antisocial behavior. In particular, we want to be looking at problems in the discovery strategies that are used, theory development, conceivability itself, and problems in the justification of theories-- the ways in which we organize our evidence, the rules or standards of evidence that we use. So those are the two domains. We will pull together recommendations towards the end of the day. Some of what we have heard about I think is very pertinent to both discovery and justification. In the discovery end, Mark, you were asking about where do our theories come from, how do we develop those theories or those models to begin with? Do we do that on the basis of observations, and how can we do that appropriately?

What Bill was talking about was selecting the problems. How do we go about selecting the problems, rather than getting too caught up in the strategies in advance? Can we build programs by thinking about the models and the theories and do that in a systematic way? And your point, Mike, was that insight should be clearly differentiated from prediction. What we want to do is try to nail down some of the major problems in either discovery or in justification, and what you see as the big obstacles. Why are we at this impasse? Are there solutions that you can see for trying to get us out of the morass?

Dr. Appelbaum: I'd like to comment for a moment or two on what happened in the first meeting of the APA Task Force on Null Hypothesis testing? I think it is relevant to what we are discussing now and also perhaps will encourage us not to distinguish all that much between the discovery and confirmation phases. As you know, there has been a long, long history on *does null hypothesis testing do us any good? Does it cause harm?* The Board of Scientific Advisors of APA (upon hearing some rather radical proposals such as mandatory banning on the reporting of hypothesis testing in APA journals) convened a Task Force to advise APA on this issue. It is a rather impressive group of people in that it has a wide range of perspectives and domains represented, ranging from mathematical statisticians to those dealing more with philosophy of science. What was the amazing thing was in the very first meeting it took us all of five minutes to say that the null hypothesis testing question is really not the important question, and to essentially dismiss it by recognizing that when it is used properly it serves a very important but limited role. And like many other tools when it is misused very unhappy things result.

The question before the Task Force got into why do people do so many horrible things --- the kinds of things that Geoff Loftus and others in this group have pointed out. That discussion led to what I think is going to be the most important recommendation that will come out of the Task Force. It is really a recommendation to journal editors, a recommendation to funding agencies, to recognize the fundamental value of exploratory studies --- of studies that are used to *generate* hypotheses. There are 15 or so of us on the Task Force and I think that virtually every one of us believes that the most horrible violations of the fundamental principles of statistics occur because researchers generate theory out of whole cloth. We have developed a culture where hypothesis testing, theory

confirmation and/or disconfirmation, is the end all and be all of scientific research. This culture forces us to prematurely formulate theory, to set up models which we have no *a priori* belief in, and then to utilize theory disconfirmation methods to disconfirm what we don't believe in the first place! It seems little wonder that we continue to have "twisted" logic in the use of null hypothesis statistical testing. It just isn't the right tool for some of what we do.

Dr. Richters: This is an excellent point, Mark, and the reason we decided not to separate artificially between the two. Go ahead Kimberly.

Dr. Hoagwood: Part of what we are trying to do is identify some of the obstacles in discovery or justification. If that is not a good distinction, or if that dichotomy itself is creating an obstacle, I think that we should think about this as well. In order to help us think about ways to organize the obstacles, so that later in the day we can start coming around to strategies, approaches, solutions, etc., Peter Jensen has come up with some categories for organizing the discussions. So do you want to present those?

Dr. Jensen: Well, there is nothing sacred about this, of course. But it might help guide it. Oh, thanks. This might help guide our thinking so we do not get too off into narrow hobby horses. We want to keep this at the conceptual level. In thinking about all of the email dialogue preceding this meeting, it seemed to us that we do not want to say "Well, golly, we just need to include another variable or methodological refinement", such as better sampling of ethnic or minority groups. This is all fine and good, but this is not the conceptual level at which we want us all to struggle with the issues. Or, one might say "We need new approaches and new statistics", but we hope to go beyond these ideas as well. And so the issue of using new methods or multimodal, multi-informant, or graphical approaches may all be useful, but if we stop at this level our feeling at NIMH is that we are not going to get to where we need to be.

Our concerns have to do with this: How do we construct theory, develop theory? How does theory guide lots of these things, or sometimes the discontinuities between our theories and our actions. Do we do violence to our theories when we collect data in a certain way, with the notion that we have to justify any findings with a certain statistical approach at the end? And so we want as much as possible to wrestle with all of the conceptual levels. So when you offer comments, let me ask you to be clear about the conceptual level at which you are working. For us, where we have been most troubled is at this fundamental and epistemological level. Now it may well be that tactical and strategic problems of psychology training are the "big issues". Maybe that is really what is going on because everybody, as Mark said, everybody knows --- or at least leaders in the field know --- that there is a big problem, and it is the just the way we are training, and so we have to do new training. But we suggest that part of the problem is that we are having difficulties, logically and epistemologically, in knowing when we know. What are our rules of logic, our rules of inference, our rules for scientific evidence? It boils down to induction versus deduction.

Allow me to provide a case in point that shows that the nested problems. I submitted a paper recently to a journal in which I came to the an outrageous claim at the end that we do not get valid data from children in certain diagnostic areas, such as oppositional defiant disorder, and attention deficit, versus what we get from parents. And so I laid out the logic for this in the paper, and I showed that when children reported their symptoms they were not validated by clinicians. I showed that child-reported symptoms are not related to impairment. They did not seem to have all of the other external correlates, whereas parent- reported ODD symptoms have face, construct, and concurrent validity. Obviously, you could say, well, this gets behaving badly. What does it mean to say that a child says he is behaving badly and no one else sees it? Well, that is interesting, but I cannot make a credible argument why that should be meaningful in terms of diagnosing ODD. And yet we as a field say we have to have multi-informant data, something from everybody. In fact, one of the leaders in the field has even suggested that when we examine risk factors, all correlations between symptoms and risk factors should be informant-specific. In other words, we this person argues that we should just take the symptom perceptions of the child and, taking them at face value, always treating them separately, because they are so different and so independent from parent's reports. I thought that any sensible person --- like my grandmother --- might ask "wait a second, you mean you just *assume* that what anyone says is true, and you are just going to throw it into your correlations?; aren't there *some* other external criteria under which you could say well, this is good or bad data?"

Well, in that paper I had to bring in multiple lines of converging evidence so that I could say, "Look, this is a very plausible working hypothesis and I put it to the field to test it empirically". Well, that took a lot more journal space, and I had to tap into other lines of evidence including what we had done in the same study. So I had to go

back and tap into other lines of evidence. But the reviewers were at another level, saying "Well, you know we all know we need multiple informants and multiple approaches." The integrity and validity of the information was less important than doing it the right way! And so in some of your comments, you have alluded to the fact that we do not have good ways of teaching our grad students and our scientists a programmatic conceptual approach to research that is driven by solid logic. We sometimes stop thinking once we have seen the p -value, and we do not say, "Well, under what conditions does it or does it not hold true? And how can I go about testing that hypothesis, and laying those questions out in a theoretical program of research that has that sense of vision and perspective?"

So as we talk about the problems of discovery and justification, this might be a useful frame for us to ask "At what level are we talking? Are we talking at a relatively superficial level?" We may need to, because many of the obstacles may have to do with training, but maybe we need to consider carefully these conceptual levels. Maybe we are wrestling with problems about e we are really at other levels. There is no controversy here. Maybe we will come to that conclusion, but maybe not. Maybe we are wrestling with problems in the way our mainstream thinking has been working, having to do with these linked problems of justification and discovery.

Dr. Hoagwood: Thanks. Dan, you had a comment?

Dr. Robinson: Yes, following your instructions, Kimberly, I do want to say a few things about models. I am reminded of Norbert Weiner's comment that the best model of a cat is a cat, preferably, the same cat. Physicians in the room are probably quite familiar with the name Einthoven, who discovered the EKG, the electrocardiograph. I doubt he would be funded under the usual arrangements today, because what Einthoven argued was this: That the best representation of the heart --- which is in the chest, is covered by a rib cage, blood flowing through it, etc., messy, fatty, etc. -- is essentially an infinite spherical volume filled with an electro conductive medium, with a dipole in the center of it! I now record from two different points and recover a waveform that would constitute a worthy model of the heart's dynamics.

It turns out that the difference between Einthoven's model and the electrocardiogram is something like five or six percent. It is not a big difference. The reason I mention that is to say that, at least in my understanding of the history of science, models inevitably are simpler than what it is they set out to model. Now if we take the social science or psychological approach to development or criminality and the like, *the sample* is very often used as the model of the individual; and the sample of course is horrifically more complex than is the individual. So there is something backwards about the approach to begin with.

Now if I seem to be lapsing into an $n=1$, it is explicable partly in terms of a misguided youth spent in psychophysics, where the only reason you throw in a second subject is just to make sure your first one was not dead; and then you need a scale factor to separate the data you get from the two subjects. If I seem to be lapsing into an $n=1$, I do so for this reason. Nobody studies 100,000 people in order to determine whether the Krebs' cycle is the right model for carbohydrate digestion and the like. When the epidemiologist goes into a village and finds that 99.9 percent of the sample is running a temperature of 105° , and there is this outlier at 98.6° , he does not start bringing treatment to bear on the 98.6° "outlier". This is because there is already in place a model of what constitutes a flourishing, and worthy, and adapted physiological system. It is through the examination of the individual case that one establishes normal functioning. We can speak of "psychopathology" only to the extent that there is already in place a model of what constitutes socially acceptable behavior, etc., that is, there has got to be an essentially *civic* framework with the usual freighting of ethical precepts and the like. And *then* you can look at the given case and say, "Ah-hah, well, this isn't quite working."

Freud, for example, never thought that in intensively studying the individual case he was departing from the traditional approach of medical science. His theoretical excesses would disturb the medical community. After all, operating in the shadow of Mach's positive science, Freud --- a well-trained member of the medical establishment --- understood that it is in the exhaustive study of the individual case, that you see the general precepts, the general laws that must be operating everywhere. Otherwise, you would have no opportunity to develop a science at all. If the individual is not reflecting the general principles, characteristic of the species, then, in fact, you do not have a characteristic species. So I do not think we should dismiss $n=1$ on the grounds that it is not assimilable into a scientific framework. $n=1$ may be nonscientific, but not because it is $n=1$. The point being that the individual is the best model of himself.

Dr. Richters: Thanks Dan. The email and phone dialogue leading to today's meeting reflected a common core concern about the structural heterogeneity challenge and two different--- but compatible--- approaches to resolving it. Many of us emphasized the need for new strategies and methods at the beginning of the discovery process to generate the necessary data. Others focused on ways of making better use of data already in hand by employing data analytic strategies and graphical displays capable of modeling and identifying structural differences among subjects within samples. We seem to be in general agreement about the existing problem, though: The lion's share of sample-based research concerning the causes of antisocial behavior in childhood and adolescence-- and individual differences research more generally-- is logically predicated on the assumption that there is a single, complex, multivariate causal model underlying phenotypically similar syndromes. What everybody believes, though-- including most researchers, police officers, probation officers, teachers, parents, and judges-- is that there are qualitatively different subtypes of antisocial children and adolescents in terms of what gives rise to and maintains their antisocial behavior.

Dr. McGuire: Well, models, case histories, and structural-homogeneity assumptions are typically yielding flawed knowledge. I will be the Devil's advocate and admit that *all* knowledge is limited. The thing known is inadequately represented in our knowledge of it; and our knowledge of it is inadequately represented in our communication of this knowledge to ourselves and to others. It is the tragedy of knowledge that our necessary representations of the known are necessarily triply misrepresentations of it - - under-representations, over-representations, and mal-representations - - as I've discussed in my [1989 perspectivism chapter](#). The only thing worse than knowledge is not knowing.

Dr. Appelbaum: But isn't that fundamentally an empirical question?

Dr. McGuire: Yes.

Dr. Appelbaum: As opposed to a question that you start out *a priori* with?

Dr. McGuire: Absolutely, absolutely!

Dr. Appelbaum: So that following the model of what I believe, as you do, normative science to be, you go out and you make some observations, careful observations. I mean if you think about the people that have really influenced developmental psychology, there is a Piaget with two cases, rather than three cases...

Dr. Robinson: Darwin's observations on his infant son.

Dr. Appelbaum: Yes, that is where the ideas would generally come from.

Dr. Robinson: Well, let me join Bill McGuire. There may not be a unifying principle across all cases, but you could have a subset. It probably would be a tiny one. Let's say a subset of youngsters that have temporal lobe focal epilepsies. The best explanation of antisocial behavior then would be in terms of a neuropathy. That may be a vanishingly small set of the total set of -- now, one might say, ah-ha, yes. Many, many factors are involved. Let's say 12, and there are exactly 12 identifiable classes of antisocial behavior, each responsive to this causal nexus. So I think, again, theories are underdetermined by data.

Dr. Fox: Can I just ask a question of that? You would not say that every child that has temporal lobe epilepsy --

Dr. Robinson: No, no, no, no, no, no!

Dr. Fox: So the converse is not --

Dr. Robinson: No, that is right.

Dr. Fox: When does it --- to know that you have the subset of children who have temporal lobe epilepsy?

Dr. Robinson: Well, only that in the individual case you may want to do an MRI or something to see whether this is a member of the subset where the best explanation of the behavior is in terms of a neurography.

Dr. Fox: You might want to do that.

Dr. Robinson: No, no, no! -- I offer this example of a prediction. The social sciences will never be able to claim it themselves experimentally. The *a priori* probability, if you are mugged in Harlem between 2 a.m. and 5 a.m., that your assailant is Afro-American, is .999999. All right. Well, it is 3 a.m. and you are on 138th Street and St. Nicholas Avenue and you are mugged. Your assailant gets away and the police cordon off the area and they line up 20 suspects. Nineteen of them are Norwegian sailors, whose tour bus took a wrong turn off Broadway; the 20th is a young black chap in sneakers breathing heavily. Once you come to grips with the fact that you are still going to have to have a trial, you begin to see the difference between prediction in the social sciences and explanation in what is sometimes called the juridical sciences. So, no, the datum itself cannot be dispositive. Quite so, again, *the underdetermination of theory by data*.

Dr. Richters: But then suppose we were to take seriously the proposition that the temporal lobe deficit is causally relevant to a particular subgroup of antisocial children. How would you go about establishing empirically that this subtype, or any other for that matter, exists and both can and should be discriminated from other subtypes?

Dr. Robinson: Well, no. Well, what Wilder Penfield would say is that following the extirpation of that calcified glioma, if there is a total cessation of antisocial behavior of the sort that made this an interesting case in the first instance, you certainly have a leg up on the explanation. Now these things are always fraught with interpretive difficulties. But, yes, the post-operative picture is surely the one that neurosurgeons would use to determine whether that is an apt surgical procedure in cases of this kind.

But you know if we take that specific example and we take a look at the category, antisocial behavior, and we find out that what the kid did was strong arm somebody else for his lunch money versus suddenly erupt and shoot people in the -- well, shooting is the wrong example, because that implies premeditation, but do something spontaneous. The fact that the kid has temporal lobe epilepsy may mean something entirely different. That is number one. Number two, temporal lobe epilepsy would effect different people differently I assume, depending upon where in the temporal lobe there is a problem.

Dr. Appelbaum: So are you interested in temporal lobe epilepsy, or are you interested in antisocial behavior? Depending on which one you are interested in, you may look at a tremendously different host of things in either case.

Dr. Hoagwood: Just one comment. It seems to me that the concern about description, which is very important, we ought to get back to, applies not necessarily only to subtypes of kids, but to the description of the labels themselves. I mean the very labeling act itself has its own set of differentiations that need to be considered I think, without regard to the specifics and what the behavior in the kid might be. I mean that is the Mark Twain example, as I understand.

Dr. Richters: Well, we all understand how it is done now. The question we face is how to do it differently-- how to move forward in a responsible way, framing the heterogeneity issue as an empirical question? The rationale and justification would be that the coherence and the potential of the current model absolutely depends on an assumption for which there is no empirical, theoretical, or conceptual support and plenty of *counterevidence*, and therefore the heterogeneity hypothesis warrants serious consideration. How would we move methodologically from where we are to where we need to be given that-- as I emphasized earlier-- it will require a rather profound reorientation to discovery processes?

Dr. Robinson: John, a very simple question. What *is* antisocial behavior?

Dr. Richters: Excellent question Dan, and, as I argued in the Mark Twain paper, it depends.

Dr. Robinson: Yes, but you see if that is the case, then sending people out with thermometers into the community and getting this tremendous range of temperatures leaves you with nothing but a set of descriptions that cannot be hung on anything. It is only in virtue of knowing something about the normal functional anatomy and physiology of the system that the measurement means anything. Now as you move from young children, and Sicily children, and Beijing (I thought Ken Rubin's account was wonderful), this does not relativize everything. It does not say that anything goes depending on where you are. What it does say, however, is that there is an inextricable connection between the vocabulary of your axiology or system of values -- the terminology you actually use, whether it is negatively or positively freighted --- and very broad cultural precepts that have been in place for the longest time.

I should say this particular division of NIMH has every reason in the world quite bravely to say, look, we are not doing blood counts here. We are trying to match up our programs with bona fide social problems that are at the very core of the cultural and civic life of this country. That means then that we have got to take a certain position on what constitutes acceptable behavior. Clearly and abstractly, there are imaginable worlds in which what we are judging negatively would be judged positively. We are not functioning in an imaginary world. We are functioning in this one, and we can tell you the one we have. We have a constitutional order respectful of individual liberty. That is the framework. We did not put it there. We are the beneficiaries of it. That is the framework. Behavior that is unacceptable is behavior that it is incapable of sustaining a civic world like that, and *that* then is the "axiology". There is no need to argue it further, unless in fact the political and civic realities of the supporting community become so transformed that it now becomes inapt, to say all this. If you do not have that framework, then the thermometer readings are irrelevant. You are running around trying to build better thermometers; wrong task. The level of accuracy here is not the issue. The question of *what constitutes health?* is the issue. Failing to settle that, the thermometer is a useless instrument.

Dr. Appelbaum: Why not follow Mike's suggestion? Use data that you already have, imperfect as it is, and set up some mental model of what it would take for you to be able to say comfortably, "Wow, there is enough similarity here that there might be something worthwhile looking at further". If you are then comfortable with the idea that the variation is just the variation of individual differences in a more or less homogeneous population, you stop. If, however, you find "clumps" --- groups of individuals who are extremely similar to each other (low variability) but are very different from other individuals or other clumps then you might want to take a different approach. The point is that we can use extant data to inform us about this. We probably don't want a blanket approach.

Dr. Hoagwood: But the problem that I am hearing Dan talk about, and it relates also to what you are talking about, Ed, is a pre-data problem, having to do with our classifications and the assumption that there are natural divisions of the phenomenon (i.e., antisocial behavior) under investigation. We assume that these divisions arise out of nature. The divisions and the classifications we create for describing antisocial behavior arise in relation to the specification of a particular goal. It would seem as though what we need to do is articulate or at least acknowledge what the goals are--what is the intent to classify. This connects to a point that you had made in one of your E-mails, Dan, about "adjustment to what end." I think if I were to select and frame one problem or obstacle in developing a program of antisocial behavior, this problem would be the classification systems we use, the way in which we divide up the phenomena we study, and the ways in which we do so without acknowledging the goals or the teleology by which these phenomena are classified.

Dr. Appelbaum: I do not understand what that means.

Dr. Hoagwood: Let me give you an example from the area of functional impairment and its measurement. In the functional impairment area, the measures that are commonly used reflect different kinds of theoretical models. A well-known measure was originally developed by Vineland for mentally retarded, institutionalized adults. The definition of functional impairment used by Vineland earlier in this century was "social usefulness,"-- the extent to which an individual could economically contribute to, rather than burden, society. That had real meaning for institutionalized populations in the 1930's. Now when we use the Vineland with children, are we still interested in social usefulness as the underlying construct for what we are measuring or not? The measures we use divide up behavior according to underlying assumptions. There is always an underlying assumption behind any measure, and it is also behind the data. My point is that we ought to recognize and acknowledge assumptions and their influence on the data.

Dr. Appelbaum: Is it that or do you just have multiple ways of indexing what is seen as a kind of a

complicated but, more or less, homogeneous --

Dr. Robinson: No, Mark, the teleological issue is this: Do you understand the data in such a way that you fear you are not going to get very many "good Nazis" this way, so you better work harder, or do you step back and say, "We are not trying to make good Nazis. We are trying to make something else. Therefore, this data set should be understood as portentous, as opposed to Göebels saying "this data set looks very good to me." There has to be a framework of what constitutes health for the measures to mean anything. Otherwise, you are simply playing with data.

Dr. Appelbaum: No, no, I understand that. What I am trying to figure out is this: Is this just developing two different indices of what you might think of as a kind of a common underlying trait in each of us? Think of the measurement of mental abilities and all of the various ways that we have chosen over the course of the science to index those things. They are still referring back to a similar underlying process of how you input information, do something with it, and make decisions. I do not quite get it yet, I am afraid.

Dr. Richters: Let me try a different tact -- and this may be only part of what Kimberly was referring to or totally my own. How much can we reasonably expect to accomplish with existing data sets? Given that the discovery strategies used to generate most existing data sets were predicated on the assumption that subjects differed only quantitatively with respect to the theory-relevant variables-- that they were structurally homogenous-- the necessary data for identifying subtypes after the fact may not be reflected in the resulting data set. So when you stop at some point and say "Maybe there's more than one causal structure operating in this sample" the data may be rich enough to go back to and look for the necessary patterns, but in many cases the necessary data may have been systematically excluded at the point of data collection.

Dr. Appelbaum: I wouldn't disagree with that, but I would think that at least a starting point would be to look within what we already have. The people who developed these measures were not totally idiots at the time they developed them. They were based upon, I assume, some reasonable sort of thought processes. Now if these differences are so major and so profound, then perhaps on what we already have we would expect to see some sorts of differences. If we don't see these differences on those, then you might want to say, "Wait. I think that it may be that we have not looked at the right attributes of the individual." Then go ahead and develop new measures, which will be seen 20 years from now as naive as the measures were before. Then go ahead and do that, but ask yourself before the fact, "If I were to find different subtypes on these measures, what would that do for me?" So I am saying you can take multiple levels.

Dr. Richters: It seems to me that Terrie Moffitt's review article on early-onset versus adolescent-limited antisocial behavior is a powerful statement about the status of the field. The distinction clearly captures something important in the sense that early-onset antisocial children are more likely to be characterized by all sorts of negative personal, environmental, and family factors and are more likely than the others to suffer a variety of negative short- and long-term outcomes. On the other hand, it is totally ambiguous for theoretical and practice purposes how these group-level differences should be interpreted. I have little doubt that one could find similar support for a distinction between early-onset and late-onset, time-limited coughing in a heterogeneous sample of patients whose coughing was due variously to such distinctly different causes as lung cancer, asthma, cystic fibrosis, allergy, industrial pollution, working in coal mines, etc.

Dr. McGuire: I think that most of us agree on the problem but disagree on the solutions. We agree with John Richters' [Hubble hypothesis](#) that the usual behavioral science assumption of structural homogeneity is erroneous and is distorting our priorities. Most of us probably agree that antisocial behavior manifests equifinality and multifinality. A given psychopathology can arise via different paths and can eventuate in radically different outcomes. To investigate this some of us use razzle-dazzle new analyses (logistic regression, structural equation modeling and the like) to tease out different ways of getting from here to there. But others of us argue that such state-of-the-art analyses often obscure the phenomenon being described, because they offer no one-to-one correspondence between parts of the model and of the phenomenon. If so, we may end up with improved predictions at the cost of less understanding. Are we willing in our theorizing to pay for an improved product simulation at the cost of poorer process simulation?

Dr. McGuire: But, as Reuven Dar pointed out, some advances may sacrifice understanding for prediction, process simulation for product simulation. For intervention and prevention, understanding may be more useful than prediction. I am not convinced that the two are antithetical; perhaps we can develop theories and analyses

that constitute the best of both worlds. Or if, tragically, efficient prediction and provocative understanding turn out to be antithetical criteria, then perhaps we can strategically plot out a program of research with a mix of methods and weigh criteria differently to fit a variety of purposes.

Dr. Richters: Peter?

Dr. Jensen: A question for Dr. McGuire. Could you state a little bit more about why it is an either/or? Why you would have to give up one to get the other? Are you saying that you either go for the sometimes, you go for the process within it sometimes? Is that kind of what you are boiling it down to?

Dr. McGuire: I think it is either/or, but not in an important sense is it either/or. It is either/or, in that some workers will have to be working toward this super computerized product simulation, and others will probably use more traditional laboratory or field manipulations where they are looking at one thing at a time, and study the process within each. I mean that science is a social interpersonal process. Some researchers will use one approach and some will use others. Just the same, each study might be better designed if one knew what one's objective was in this study and give up temporarily on the other objectives.

Dr. Hoagwood: But its objective was to compare those two approaches. That could also be built in to a design.

Dr. McGuire: Yes, you could try to merge them, but I would regard that as a more distant aspiration.

Dr. Appelbaum: Bob, does this sound very familiar if you just substitute learning disabilities for social whatever problems?

Dr. Cairns: Yes, well, possibly. I think that one of the problems in arriving at subtypes, or subgroups, as John pointed out, is that it is a theory-driven process. This shows us both in the selection of measures (i.e., what you throw into the pot to get your subtypes), and when the subtypes are computed. For example, look at the trajectories, as Mark Appelbaum and Robert McCall did in their 1972 SRCD monograph. They classified trajectories and came up with about five or six. But you also classify outcomes, and then work backwards to get to the process of how the outcomes were reached, the processes of development. Another issue is that the subtypes could be identified on theoretical grounds. It does not have to be with empirical procedures. Consider *a priori* attachment subtypes, or multiple subtypes, the B1s, B2s, you can cluster on the neurobiological status of organisms. I think your point is well taken: that to get the process it looks like you are going to have to combine procedures.

One way to do it would be to classify in terms of antecedent conditions and cluster at the outset of the study. This permits the researcher to pay as much attention to negative cases as to the positive cases. Such antecedent clustering permits a kind of a yoked control, a matched subject design, that could then be used to give a more powerful analysis of developmental events that occur differentially in positive and negative cases. You can then fit all kinds of hierarchical models to explicate the differential process. Again, I think it is the integration of strategies that is called for: thinking carefully and logically about the steps at each phase of analysis rather than counterposing one strategy to another.

Dr. Robinson: Take one of the founders of developmental psychology in the modern period, James Sully, whose *Studies of Childhood* was foundational. He begins with expressed debts to Darwin and the evolutionary framework. He concludes with a long appendix, which is the autobiography of George Sand. What does he think the autobiography is doing? He thinks the autobiography is giving in the individual case the instantiations of all of the influences on development that haven't been developed in the body of the text. He sees no tension between the two approaches. He sees a complementarity. I should mention a book of mine that came out just last year. It's a only history of the legal concept of insanity from the time of Homer to the present. It is a book titled, *Wild Beasts and Idle Humors*. You can see the problems we are wrestling with especially in this. As the epidemiologists of old are telling us whether rural life is more conducive to insanity than urban life, you can see that what is taken for granted are standards and criteria of insanity which we would find insane! So you have this excruciatingly careful level of description, based on a theory that within 10 or 15 years is going to be thrown out the window. When the physiologists gain hegemony in this domain quite early in the 19th Century we come to learn that moral insanity is the result of-- are you ready for this? I want you to take notes on this --- a lesion of the will!

Dr. Appelbaum: And I have an imaging technique that --

Dr. Robinson: Yes, that did pick it up. It is probably somewhere in the spleen I should think. So, again, the question that plagues me is whether the characteristics of a sample, no matter how accurately and statistically treated, constitute a good model of the individual. That to me is the core question, because the only reason we are interested in the sampling statistics is presumably to be able to understand the individual case and make predictions about how that life is going to unfold.

Dr. Appelbaum: But, fortunately, that would be good index for that.

Dr. Robinson: May I just say that if it were the case, that the powerful statistics brought to bear on the data were doing that job, this division of NIH could issue an actuarial table or something comparable to all families. You could pick it up at the supermarket, and parents would know that if Billy does this, this, and this, lock him up now. It turns out that we don't do a good job with that, and that, in my example, it was not the Norwegian sailors, and it also was not the young black chap. They actual assailant got away, and no degree of sophisticated statistics on the *a priori* probability of assault by race is going to help here.

I don't believe that our data bases constitute good models, nor do I believe that you find out anything about an individual by studying individual differences. If you know that in some sense Ed and I differ by six, you don't know anything about Ed and you don't know anything about me. You just know that we differ by something. The idea that an individual differences paradigm is the right model for understanding individuals, if I may say this with respect and admiration, could only be achieved within what Williams James called, "... this nasty little subject." I do not know anybody who would have dreamed this up before becoming a psychologist.

Dr. Richters: Well, given all that we have talked about, if our charge were to forge a new scientific approach to grapple with the structural heterogeneity challenge, how would we go about it? How would we alter the discovery process? What role would theory generation, exploratory research, and hypothesis testing play? What standards would we adopt for evaluating our ideas, critically examining our theoretical notions, deciding whether we are making forward progress?

Dr. Appelbaum: What are you trying to do? I mean do you want to do your problem of predict an individual?

Dr. Robinson: That is not my problem. My problem is this. Unlike the natural world, which is given, the human social world is brought about. And, thus, one can do a thought experiment and say, "What kind of world are we striving to achieve, or what kind of community, or what kind of neighborhood, and the like? That then becomes the Newtonian framework. We just stipulate it. And then the question for research would be, based on what we know and what we can find out, what would you have to do to bring this about, and what sorts of things militate against bringing this about? If you have not taken step-1 -- I am sorry to be so repetitive. If you have not determined *a priori* what constitutes the desiderata, then perfecting the thermometer achieves absolutely nothing. So, I would say, specify a number of possible social worlds. These might include everything we know about the existing one. Try to render this in the most coherent way you can, and then take a look, by way of the data, at the study of individuals, and the study of groups, and the factors and influences that seem to be capable of supporting or seem to be capable of subverting each of those imaginable (and several of them actual) worlds. They all sound so very simple. From my reading of philosophy and history of science, it is very much the way these things are best done.

Last point on this, I promise. When Newton approached Kepler's laws, Kepler had Tyco Brahe's data. This was a data base to beat all data bases. What did Newton do? He moved away from this entirely and said, "Let's imagine an object, a unitary object going around a force field." What could he do with that? He derived Kepler's laws! Simplify, simplify, simplify. Not *multivariate* but *univariate*. Specify the world that is coherent to you, and then determine what variables are necessary and sufficient to bring it about. Don't start off with this bloody world and say, "Let me see what is working here." This is finding out what a radio is by smashing it on the floor and counting the pieces.

Dr. Hoagwood: Dan, give an example of specifying the social world.

Dr. Robinson: Well, this is done very often in quite limited social context. For example, you want to bring about a team that is going to win the Super Bowl. Most people could specify what you would have to do to pull that off. You are not going to do this perfectly, but you certainly know what a bunch of non-starters are. If the defensive line has an average weight of 60 pounds, chances are you are not going to do very well, etc.. People who try to solve problems within a corporate context, or within a departmental context, generally can list the sort of things likely to promote or retard what the end game seeks to achieve. Now, at the level of society this looks hopelessly complex, but it may not be. There is a wonderful book I would recommend to everyone interested in American history. It is a book by Barry Shain, which was published by Princeton a few years ago, called the *Myth of American Individualism*. What Shain points out in this work is that this country, before it became an independent (1735 to 1765), is very much operating in a reformed Protestant ethos. It is *Communitarian*. Anybody standing up and asserting his "individuality" would have been regarded as a lunatic. As you move into the 1790s and closer to the age of Jackson, with westward expansion, you get a radically different conception of what a meaningful and proper social life is. It is very much an insulated corpuscular kind of life. You are out there as a frontiersman doing the best you can.

Now if you wanted to realize the American picture circa 1765, you would be trying to encourage things quite different from what you would be encouraging in 1820. Well, we have a political debate going on that features things like "family values". Some of this is absolutely horrific, and the people who make these speeches recall to mind Shakespeare's "Shame, where is thy color?" But there is this sense that the social world we live in is unraveling, and that the civic dimension of life has atrophy to the point where it might be beyond recovery. Well, that must mean we *do* have some conception of how things might be better. All right. Start putting together what the schoolyard might look like, what the classroom might look like, not in Sicily, not in Beijing, but in *American* communities. And now start tallying up the behaviors, and tendencies, and family practices that either militate against or nurture just this sort of thing. There is research that has this teleological dimension in focus. It is not NIH dictating what society should be. It is NIH saying, "If this is what you are striving for, forget it. You are not going to get there this way." That sort of thing. Sorry. I have taken much too much time.

Dr. Hoagwood: Nathan, and then Geoffrey.

Dr. Fox: I just wanted to make one brief comment to that, which is, not being an economist nor a Marxist economist, I would simply say that your vision of the just society, and particularly in this country, is tainted by the economic conditions and the economic system under which the country operates.

Dr. Robinson: There is a variable you have identified.

Dr. Fox: So that, ultimately -- well, ultimately, we, as psychologists, call it social class or socioeconomic status. But, ultimately, unless one is going to study it and how it works on the individual, which we do not want to do because it is forever too complicated. So we will simply "measure it in some indirect way." We are not going to be able to be able to answer the question that you pose, because we live within a particular economic system and a society which will not give everyone the particular civic outcomes. It does not want to give everyone the particular civic outcomes that you postulate as being good for the society as a whole.

Dr. Robinson: But the history of civic unrest does not track the history of socioeconomic stratification. So there must be something else.

Dr. Fox: Well, it does not do it perfectly, but it does it rather imperfectly better than most other things do.

Dr. Hoagwood: Geoffrey, you wanted to say something.

Dr. Loftus: I would like to take some of Dan's comments to perhaps a less eloquent level, and talk about some of the things that John Richters actually brought up in his [Hubble](#) paper involving the issue: how do evaluate the quality of your science?

Part of the problem, as John said very clearly in the Hubble paper, is that psychology operates as pretty much sort of as a closed system. So we have this discovery process, and then we have this crummy justification

process, and the crummy justification process validates the discovery process, and around we go without any external indication that things are not going very well, except for sort of a feeling of angst that some of us get from time to time.

Now there seems to me that there is a very obvious set of external criteria for the quality of a science, and that is simply the degree to which it translates into practical applications-- to policy, to inventions, and things like that. If you asked most people *why* chemistry is good, they'll say it's because we have plastics. If you ask why physics is good, they'll say it's because we have transistor radios. A question that should be before us is: Psychology is good because we have what? I would like to talk a little bit about --

Dr. Richters: Jobs.

Dr. Loftus: So why do we need psychology if we have Oprah, for God's sake. Anyway, I would like to provide an example of this, partly in answer to Kimberly Hoagwood's question. I am going to break the rules a little bit here and not do it within the domain of developmental psychopathology, because I really do not know much about that area. I would like to talk a little bit about my own domain, which is memory and perception. I spent a lot of time testifying before juries as an expert witness. In a typical case within which I will testify, perhaps somebody has been mugged, and there is some suspect, and who is now a defendant sitting in the defendant's chair, and the witness gets up and says with great confidence, "*That is the guy who mugged me,*" which is a very compelling statement to a jury. Meanwhile, in this hypothetical case, there is a certain degree of physical evidence, an alibi, let's say, or something else, mismatching fingerprints, that this defendant is not actually guilty. Now the reason I am there is to try to explain what the scientific study of memory can tell us that might help reconcile the discrepancy. We have this witness giving a confident assertion that this guy is the mugger. We also have this contrary evidence. Under what circumstances do people tend to be confident but wrong in their memories? Well, much of the data that I testify about is admittedly incomplete. Numerous people, including, for example, Mike McCloskey at Johns Hopkins, have written about what the problems are with people like me getting up and testifying in front of juries about how memory works. That is sort of the bad news. There are problems in perception and memory research, in terms of being able to generalize what we found to the exact problem at hand.

At the same time, I think the good news is that this experience, for me anyway, sort of provides a path to discovery, in a sense that the term is used in the Hubble paper. Having this practical problem that I need to solve constitutes the desiderata, as Dan Robinson said. That is, it provides hints about what we should be -- what kind of research we should be doing, what kind of answers we should be seeking, and what kinds of theories we should have that would allow us to solve these problems and other similar problems.

Part of the reason that I use the example of memory to illustrate this issue is that memory and perception are fields that are mostly done because of their sort of basic intrinsic, scientific worth. We want to find out how memory works. We want to find out how the visual system works or whatever. >From the perspective of most psychologists who study memory, it is only incidental that all of this stuff could be used for practical issues like informing juries in ways that they can better evaluate eyewitness testimony.

In the field of development psychopathology, it seems to me that -- correct me if I am wrong, but it seems to me that practical applications are pretty much the whole thing. Well, maybe not the whole thing, but most of why the field exists to begin with. You have these kids doing bad things and you want to know what to do with them. You say "*All right. We have these problems. Here are these kids out doing all of these bad things. We have a particular kid; what should we do with him?*" It seems to me that this is the fundamental question, which at least starts to provide the vision that underlies the discovery process.

Dr. Trickett: Your comment about the American individuals and reminds you of the Jean Baker Miller's question, "How many women does it take to keep one man independent?" Let me clarify what I think is an important difference. That is the difference between talking about antisocial behavior and talking about antisocial children or kids. It seems to me if the question is on antisocial behavior it would allow such questions as what day of the week do more antisocial acts occur? How come school-A kicks out 14 percent more kids than school-B does for certain things? Questions like *What kind of neighborhoods promote antisocial behavior?*, in contrast, allow a kind of multilevel look at the phenomena, rather than locating it on a particular individual level. I am just wondering if that is a useful kind of distinction. I was thinking about research --- acculturation research --- where there is something that people talked about in adolescence and in families in terms of the acculturation gap, where

children acculturate more quickly than their parents do. It is a cause of conflict, et cetera, et cetera. It turns out that kind of phenomenon is really much more prevalent in Latino immigrants to this country than immigrants from the Far East or Russia. If you look at the phenomena of this acculturation gap across populations you get a different set of understandings about what it means. I think it is important to distinguish between antisocial behavior, which can be looked at a variety of different levels of analysis, and kids as sources of antisocial behavior.

Dr. Richters: There are a variety of ways, Ed just named a couple of them, that are equally valuable for answering different kinds of questions. I really resonate what Dan said about the civic framework within which we're working and the powerful influence of social and cultural factors in shaping our views about what's important, what's good, what's bad. That does not in any way conflict with a recognition that in certain cultural contexts antisocial behavior patterns over a long period may be "normal" and normative, even though they are morally and legally despicable. So I think those two are not necessarily at odds.

One of the big problems we face is an inadequate monolithic paradigm. Depending on the particulars of a research question and the state of our knowledge at a given time, meaningful research efforts may require fundamentally different approaches to how we frame the questions, how we address them, how we try to answer them, what we accept as reasonable evidence. And, again, I need to draw us back to the question of, given that, what can we do to break set in responsible and rigorous ways where it necessary? This does not presuppose that everything we are doing now is horrible.

Dr. Jensen: Well, I was just going to re-raise a question going back to Mark's earlier comments. The statistical approach is only one of several tools, and yet there ought to be other tools out there. As a research agency we ought to be exploring, exploiting, and understanding what takes place at the level of the individual. We might be able to embark on a process where we find what the types are, and then following them to identify the processes within these types. Yet our mechanisms do not easily allow that kind of research. As a funding agency, we are basically employing one kind of approach, but it is just *one* line of evidence on the justification side.

Dr. Loftus: I am sorry. Why do you mean by just one approach?

Dr. Jensen: Well, basically, the predominant problems are seen in quasi-experimental group-based studies in the area of developmental psychopathology, not so much in experimental research. I am often impressed, if I look at some of the basic behavioral research where someone comes in with a grant application saying "I have this very interesting theory to test. I have three or four nested experiments that I am going to run. If my theory is correct, then **X** should happen, and under other conditions **X** should not happen." And so they build a very elegant argument that rests on a series of tests, an overall set of inferences, that should arrive at, with these conclusions based on the overall pattern of the evidence. Now we do not do this very much or very well in longitudinal, non-experimental research, where we begin with our diagnostic categories on a false set of assumptions. We have ignored basically the desiderata. We have taken all of that for granted. We basically have either large scale epidemiologic (or sometimes smaller clinical) samples, and we follow these sometimes tweaking the variables here and there. Then it rests on the multivariate analyses of trying to kind of figure out what is going on. But there are other ways to approach that data, or to design and conduct a study so that one could come to the end of the study and say, "Golly, I have four lines of evidence at the end of this study that all seem to converge on this conclusion. Instead, there is only one line of evidence, perhaps all built into some path analysis or structural equation modeling, etc., but it has got a "loosy goosey" feel, described as one that has "captured" 15% - 20% of the variance, and then someone else comes along and simply does another analysis, adding one more variable to the equation. And then somehow another variation on the same study gets carried out all over again. Now one might ask how, if we had a different way to approach that kind of research, we might build in multiple lines of evidence to test a given hypothesis. How do we follow up and test those hypotheses as rigorously as we can without experimental paradigms? So that is what I mean by following John Richters' and Mark Appelbaum's sense that multiple strategies are available but we are only using mostly one.

Dr. Mitnick: You know, Peter, I think one has to recognize that even within the community of research study there are different cultures, and that the kind of application that comes to a review committee reduced by cognitive social psychologists will get bombed in an AIDS review committee, because of the step-by-step little kinds of studies that you do. Whereas, the AIDS review committee has focus on the past of what is the end result? Where are the interventions? How is it going to change? So I think we have to recognize the heterogeneity of the communities that we are talking about, and not assume that we are always studying.

Dr. Robinson: I should think it is not heterogeneous enough. I am not a potential grant applicant, I assure you. And to say one more thing on the subject. One thing I would say is that within this NIH grant setting I should think it would be very hard for someone like myself to win a grant. I maybe wrong and I hope I am wrong, but I do not think I am. It will be very difficult for one to win a grant on the grounds that a *book* would be written; one that attempted to integrate research areas and findings, and come up with richer or more various perspectives on what it all means.

Now may I say, based on my misspent youth, in which I published somewhat regularly in *Science* and other estimable places, that it isn't easy writing books. It is not what someone turns to, because of an impoverishment or lack of cleverness! At the same time, one discovers many, many things in critically reviewing the findings of others. One would not be likely to learn as much when actively engaged in just that field of research.

Where do we begin, the question is? John Richters raises it and Kimberly Hoagwood raises it. I should think one place to begin, if you think any value comes out of something like this, this is much more a narratological kind of approach --- this is what books do, as we all know; just what the journal article shouldn't do, doesn't do, can't do. I would like to see a division like this, given its unique mission, to show some receptivity, and maybe even invite applications from people who would get a semester or a summer off, to work on book-length monographs. I do think that this should be encouraged, and graduate students should begin to learn that the scientific community is not hostile to undertakings of this kind, that the 2 X 2 factorial design cannot be finally the only grounds for --- I was going to say the only grounds for tenure --- is not heterogenous enough.

Dr. Cairns: I would like to come back to Peter's comment about longitudinal study and the realization that *NIH* does change. Twelve years ago longitudinal studies were not particularly favored by *NIMH* and *NICHHD*. You could not get projects funded for longer than a three year period. So we had to write a series of three year grants to get our longitudinal study done. But such work paid off. One of the big changes that has occurred has been that longitudinal findings have changed our perspectives on antisocial behavior. We did not know 15 years ago that antisocial behavior patterns and their sequella were reasonably predictable. Although I think that the level of stability has been overplayed, it's a lot more continuous than the profession believed 15 years ago. This is an example where there has been a real payoff of behavioral research that has accumulated and been replicated. It has yielded us new information about the nature of the human condition and patterns of antisocial behavior development. I think that those longitudinal results are actually forcing a re-evaluation of standard methodologies, because those methodologies simply have proved to be inappropriate for the tasks. We are losing much more information than we should be losing about individuals over time by using the techniques that eliminate time and developmental dynamics.

Dr. Fox: Can you give a specific example? Of that, I would be really interested, exactly what you are alluding to.

Dr. Robinson: What we know?

Dr. Fox: Yes, both in terms of what we know and how that has eliminated certain methodology.

Dr. Cairns: Well, let's see now. I wrote a book in '79 on social development discussing the sorts of issues of prediction, and developmental dynamics, and the rest. When I got to the section of negative and antisocial aggressive behavior, I spent about half of the chapter showing that it was considerably less predictable and stable, at least we did not have information relevant to it. That has totally changed. The information is new and novel, and I think it has made a big effect. As far as the cumulative impact of the longitudinal studies, as far as the difference between regression analysis, say, in clustering procedure that John referred to, think of isolation and rejection, social relationships. They tend to show up in regression equation as predictive of some problem behaviors. When you do a clustering algorithm and you identify those kids who only have those problems of isolation or rejection, but no other problems associated with it, in terms of poor academic performance, and irritability, and stuff like that.

It is not predictive at all. These kids look -- actually, in some studies they look even better. You know they stay home and study. I think it helps illuminates the issue that Rubin pointed out earlier. There are multiple reasons that you can get these kinds of social adaptations. This gets covered up on a general model, and you have a lack of specificity of relationships and variables to outcomes. In any case, I think that we do not want to underplay the advances that have been made, both within the institution, and how these advances have provoked

this very discussion we are having today. They lay open the problems of treating individual coherence as they are variance, as I think as Dan has been pointing out. Sometimes we forget that we are moving ahead, that institutions do change, and that this can be another opportunity.

Dr. Hinshaw Well, there are different definitions of what that standard would be. That is, is there a discrepancy between reading and IQ, or is there an absolute level of reading achievement that is low? Knowledge of the underlying process, specifically phonological awareness, has helped to clarify a lot of disparate findings in the learning disabilities field, even at the level of genetic relationships. Without such knowledge, there can be despair regarding the multivariate problem of clustering disparate but potentially related phenotypes into more or less differentiated subtypes--and then seeking external validators of those clusters. But if we learn something from the learning disabilities field, it is that with better resolution of the phenotype, by looking for an underlying process like phonological awareness, a lot of issues get clearer.

Dr. Appelbaum: As having been part of, but where we were involved in subtyping of learning disabilities. Conceptually, subtyping sounds like it is relatively easy. In fact, it is hell! I mean it is very difficult. Part of the problem is you have to know exactly what variables are to be included in the clustering solution. This is partly what Steve was talking about. The problem is that every variable that you allow into the system can be very influential on the formulation of clusters or sub-types. Even if there is a "true" natural clustering you can easily lose the "true" cluster structure if you have irrelevant variables in the system. Despite the problems with clustering and related sub-type identification systems, one can ask whether it is a worthwhile approach. I think in our work on learning disability subtyping we found the answer to be yes. What it did for us was to separate major domains of learning disability. For instance, you would not use phonological awareness to look at math dyslexia. Now maybe, like Stan Schachter's grandmother, you should have known this without the benefit of "cute" quantitative approaches -- perhaps we should have known that there would be different processes in terms of problems with math versus problems with reading, but at the time the field was going for a generic theory of learning disability.

Dr. Robinson: But Mark, you could only have had those two groups, if they were right answers, and the gifted kids got more right answers right than the ones with learning disabilities.

Dr. Appelbaum: Oh, no! No, that was not the way the diagnostic system worked. The point is --- and I think Steve Hinshaw is saying this -- you have to take a many way approach. If you say, let's now stop all this and go into subtyping, because we have to solve the subtype problem first, I do not think it is an efficient way to move through a complicated field. I think you really need to be doing multiple things, and I think you need to be using techniques like Michael is talking about, in order to be able to inform yourself. You need to be doing some theory generation and theory kicking out. We need to ask do they have enough here that we can sort of stop these other activities?

Dr. Maltz: I think that rich description is really relevant there, that we need more case studies of kids with different types of pathologies, just to see what their life-course trajectories have been.

Dr. McGuire: In the last hour we seem to be more positive than we were in the weeks of exchanges that John Richters organized before this meeting. Then we seemed to be listing what is wrong with the way we live now. Now at this morning's meeting we seem more positive in proposing what is to be done. I have myself long been uncomfortable with the way the null hypothesis is used but so many of us clobbered it in our pre-meeting exchanges that I began to feel sorry for it, sympathizing with the underdog. Some of my best friends are null hypotheses.

Dr. McGuire: In fact we may end up identifying approaches that deserve more NIMH support, say, cluster analysis, case histories, longitudinal designs, graphic representation of relations, and so forth. In terms of the categories that Peter Jensen just presented, most of the suggestions we have been making this morning fall into Peter's first three categories and very few into the big programmatic epistemological-foundations categories. I am a bit sorry about this because the three hobby horses on which I rode into this morning's meeting involve some epistemology basics. But in retrospect our focusing on Peter's first three categories seems inevitable and even desirable. I suspect NIMH can exert more leverage on the more concrete issues than on the epistemological foundation issues.

Dr. Hinshaw The point I was making, now that Mark has left the room for a minute, is in agreement with

him. Clustering and subtyping is one strategy, but searching for some basic unitary underlying mechanisms can proceed in parallel. I am not talking about mutually exclusive ways to the truth.

Dr. McGuire: Product simulations and process simulations can be carried on simultaneously by different people submitting different proposals.

Dr. Cairns: I think an epistemological issue has been raised. It is really the issue of holism and what is the best level of integration of information. Is it at the individual level where the interactions among variables are taken together rather than being teased apart? Does one arrive at reasonable generalizations at by studying individuals or groups? That has driven I think a lot of the comments and discussions made here, and how it is you can draw reasonable generalizations at the individual levels that pertain, not only to first, to subgroups, and potentially to entire samples. I think implicit are some of the epistemological questions concerning, whether by the holistic study of individuals or the statistical analyses of groups.

Dr. Jensen: Because I think that is often why it takes a while to get to really the application, because we don't how to draw inferences well from a lot of disparate, maybe less than careful programmatic research, and so much of a research is non-programmatic, I would argue. When we have a center come together or a program project, then often you can build in and nest studies, and come to some conclusions again, but so much of it is disconnected R01s.

Dr. Maltz: But we do know how to draw inferences. Dan's example of being mugged in Harlem, an inference can be drawn. It may be incorrect, but we can draw inferences. It does not have to just be based on the statistical process of drawing inferences. I think that is where we go wrong, where we have this mechanical way of doing things, and that is the only way that seems to be scientifically justifiable. It has to do with fear of making judgments, rather than a lack of means of making inferences.

Dr. Hoagwood: Ellen?

Dr. Stover: I thought I would throw something out right before -- I assume there will be a lunch break for you to think about, and kind of the time when there is room to exchange with one another. It comes from an arena that I am from, which is the AIDS arena. I think what you are talking about -- at least what I have been hearing for the last three hours, has to do with how do you construct research strategy for a country that will answer the kinds of questions and get the kinds of data that will allow people to, as it was much more eloquently put than I, deal with these kinds of civic, or public health, or social, or whatever you want to call it, kinds of issues. I was very struck by the two comments made that were completely the opposite of one another, that the behavioral and social sciences do not provide anything, and the behavioral and social sciences basically provide everything.

I think one of the things that happened in the field that I was so wedded to for so many years is, there was no question that we had to answer and had to get people to stop engaging in certain behavior. There was no question, otherwise there would be no funding; funding would be cut off. Advocacy groups stepped in. I mean we have no advocacy groups at the table, and it is something that has struck me in almost every meeting I attend in the new structure.

It seems to me that there needs to be a way to establish from the academic environment and from the community environment, a way of framing the questions, and having the research bear directly on those questions, so that the answers are directly useful to our society. That is not easy. But, basically, what I hear you talking about is a multilevel approach of research focused on individuals, on families, on communities, some research on the society, and probably more globally on systems. Those are very different kinds of research approaches that was just pointed out. You would not expect to come in all as R1s or in a center. They may require long-term strategies. One of the things that now I look back on and say it was terrific that we did, but when we were doing it, it was very painful, was literally sit down and write out a strategic plan. These things are not fun -- and then target.

And, again, that is what I hear you kind of -- you are not using those words, but that is what you are talking about. Target where you believe the directions ought to go. The other thing that emerged, and it is kind of a radical comment whenever I make this, but then it serves as just part of the norm. We don't have an evaluative system like the drug industry, or the pharmacologic system, or the FDA, to judge when we have done a hypothesis testing

or a discovery kind of study.

Are we ready to go to the next step? We do not have anything like that in the behavioral and social sciences. The decisions are not made about how much data do we have, and then we should move to the next step. There is no body, there is no entity, that looks at that. Again, in the AIDS world we ended up with a consensus conference. It is the closest kind of way of getting the information out to the field. I used to say, how many studies do you need, 14, 18, 25, 63? There is no standard. I thought I would kind of throw out those concepts for you to think about, because they may begin to help frame the direction that you go, not only for after lunch, but for many decades. Yes?

Dr. Jensen: I think that is a great comment, and I have to say --

Dr. Stover: Well, you and I had this discussion.

Dr. Jensen: Yes, because one of the things I think that is happening to the National Institute of Mental Health in the last year and a half, since Steve Hyman has taken over the Directorship, is a little bit of a sea change --- a "show me, I'm from Missouri" attitude. "Show me that this is going to make a difference; how do you know what you know" That is how these issues touch back on epistemology. When do we have enough data? When are there sufficient lines of evidence? After we have manipulated a variable experimentally and have shown its effects longitudinally and across different populations? When should we convene a consensus development conference on well established risk factors? For example, assume that we had shown that 10% of antisocial behavior in this country is due to some critical variable, and this variable is manipulable. Also, for argument's sake, if you say that antisocial behavior affects maybe five percent of our youth and has very significant consequences (i.e., much more than MS, childhood leukemia, sickle cell disease, and cancer), why aren't we moving ahead strategically with some precision, systematically trying out those lines of evidence? And when will we have enough data? Four studies? Five?, Six? Other areas of research seem able to arrive at firm conclusions. For example, in the area of passive smoke, we had seven studies; three or four case-control and several longitudinal studies. But finally we said "This is enough evidence. We are now going to regulate smoke in the work place". SEVEN studies!.

Dr. Appelbaum: Peter, for most statisticians there is a distinction between data to know and data to show. These two are not the same things and to talk about a single standard probably is not realistic. We tend to think at the "data to show" level. We are not a product regulated area and have no tradition for that. I am not saying that this is good, but I am doubtful that we are at the point where many would accept the duality. The drug industry goes out there and it produces a product, but then before it can market it they have to participate in a very formal system, I do not know of any case, and there may be cases, but I do not know of any cases where, say a therapeutic system, or a clinical therapy type approach, or even an educational approach has actually gone through any sort of structured clinical trial validation before it is released and can be "marketed". A school system picks up a method because someone went out and gave a neat talk about the system and perhaps had data from one or maybe even two studies- probably not conducted by a disinterested party

Dr. Hoagwood: I think that happens a lot, Mark. I know certainly in the services research area that is the whole point, is to go in the real live settings --

Dr. Stover: Absolutely.

Dr. Hoagwood: -- do the studies, and then see what the dissemination uptake is.

Dr. Appelbaum: Right, but that is a rather different thing. What you are trying to do, is you are trying to get the product established, as opposed to saying you cannot use it until a demonstration is made of its safety or efficacy. It is just the opposite, I think, of the FDA approach. I mean people would have loved to have taken what was the apricot pit, laetrile, or what have you.

Dr. Stover: I threw it out using that term to generate discussion, but certainly not to suggest it.

Dr. Robinson: A profoundly important point though; the distinction between the two is a profound distinction. In the FDA situation you are certifying something that will treat a condition which voluntarily the subject population

seeks as a therapy. Suppose you found out within this division of NIMH that there is never an instance of antisocial behavior ever in the history of the world performed by someone who has been brought up as a Scottish Presbyterian. You certainly then could not precede to mandate that children be brought up Presbyterian.

Dr. Appelbaum: Let alone Scottish.

Dr. Robinson: Let alone Scottish. If there are any other avid gardeners in the room, then with me and my wife, they know that the definition of a weed is an *unwanted flower in the garden*. Of course, the definition of a gang is an unwanted "team" in a community, and the definition of a sect is an unwanted religious orientation, vis-a-vis the majority's view. The social psychologists know this better than I do. Didn't Ash do this camp study where he put jerseys on children that had been factionally divided, and whether or not they got together or remained apart, depended on whether they thought of themselves as being on the same team? That is, you could throw out 20 T-shirts, and radically transform the character of social interaction. This is a profound and powerful display. NIMH cannot then say to the public --- "We now command that to maintain your insurance coverage, your children must wear these jerseys". You must be extremely frustrated in this. No, I say this. You must be extremely frustrated to know, that even and maybe especially as you make great progress, the likelihood that you can have the social assimilation of what you have uncovered broad enough and wide enough to vindicate the entire operation and the expense, is not very great at all. Your activity is much more like an academic one than it is like a public health one. You can say these things have deleterious effects. Stop doing it. That is all you can do. Whereas, the FDA can do more. They can say "No, no, no; if you mark it 'laterite' and say it will cure cancer you go to jail.

Dr. Jensen: Well, we can do more. If we have evidence we can say, now we need to establish standards for day care. As they say indeed, the issue of day care and its effects on children is a very good example of how the field, pooled as it was, was able to mount a concerted effort, and organize a study of which ultimately use the confirmatory measures that exist to demonstrate ...

Dr. Fox: ... to demonstrate to their own satisfaction that indeed day care did not have deleterious effects that it had originally postulated that it did. Given that, it is frustrating to convince public policymakers that indeed the next step is the provision of day care or the establishment of standards for care on a regional or a national basis. That, obviously, is the frustration that psychologists currently encounter. But there is a history of that success which can be emulated.

Dr. Jensen: There are multiple venues. One might think about educating teachers about certain classroom techniques, more effective management, classroom size, and parent training under certain conditions, where we could intervene at a policy level. Now, changing policy is a complicated issue. But it requires coming up with lines of evidence that converge, so that one knows when one knows. But how do we pull the science together when we have a lot of disparate unconnected studies? It seems to me that, given the monolithic approach to so much of our research, it is going to be a very slow way of getting there. I am intrigued with one of Jerry Kagan's comments. He said, "Would we begin our research differently if we started from the supposition there are always interactions, there are always individual differences and processes?" Rather than studying the universals, as in mainstream psychology, if we assume from the start that similar processes lead to very different outcomes and different processes lead to similar outcomes, how would we design those studies? What would we put into those studies that are different from what we are doing now?

Dr. Richters: My concern is that if we do not harness the morning's discussion in the service of some specific ideas about how to bring about constructive change, the sea change that Peter talked about metaphorically may turn out to be an ebb tide. Let's break now for a working lunch and reconvene officially as a group in 45 minutes, 1:15 pm on the nose. (*Working Lunch*)

Continuation of Roundtable Discussion
Co-Chairs: John Richters & Peter Jensen

Dr. Jensen: Okay. Well, we have to move to where the rubber hits the road here, so to speak --- a phrase

that comes from my military training. I think there is a general consensus that we have converged around perhaps in three dimensional space, but not too far apart, about the kinds of problems that are out there, how they are interdependent, and difficult to resolve up to this point. What we would like to do this afternoon is to let you think as freely and creatively as possible about the kinds of things that we might do that will help kind of move the field forward. So be as daring as you want, in terms of --- it might be changes in the things we are doing specifically at NIMH, in how we have reviewed proposals, or certain funding mechanisms, or how studies ought to be designed, what we at NIMH should ask for, or new training initiatives, etc. We want to be able to take your brainstorm ideas and begin a process of iteration with you and with others in the field to see what we practically can do. Since there is general agreement that there are important problems that we might be able to work on, we want to find ways to move a little more rapidly forward. So let's get started..

Dr. Richters: Again, let me emphasize that our charge is *not* to propose specific solutions to existing paradigm problems but to propose Institute initiatives for encouraging, facilitating, and funding retooling and reform efforts in the field.

Dr. McGuire: What mechanisms, regardless of the content?

Dr. Richters: Absolutely, yes, yes! As I mentioned earlier today and even earlier in my email messages, our options are completely open --- workshops, institutes, RFA's and Program Announcements to stimulate specific kinds of research, analyses of existing data sets, working collaboratively with other organizations, convening follow-up meetings in specific research domains. That entire landscape of options is open to us.

Dr. Jensen: Geoff, and then, Mark. A little bit of early Alzheimer's.

Dr. Loftus: I have three suggestions which are embarrassingly simple and concrete. I have alluded to some of them before. They all have to do with what might be suggested for what goes into grant proposals. The first is the most simple, which is that any graphs you have in your grant proposals should not get counted against page limits. The second is that, right now, when you write a grant proposal in memory or perception you are supposed to list what practical applications your work might have. So everybody makes something up. For instance, if you are doing research in visual perception, you routinely say that it could help with reading disorders. It is something that you think about for about 10 minutes before you have to send in the proposal, and everybody makes up the same thing, and that is that. People don't really have to take this part seriously. What would be more challenging and I think more useful would be to have a serious section on what contributions related work or foundational work cited in your proposal *have* made in the real world. That would force you to think seriously about whether what you are proposing has any potential benefit at all, and you are forced to do it the way that is real rather than making up some random pie in the sky justification.

The third thing is what I mentioned in my E-mail, which is that, I personally find it very useful to do simulations of one sort or another of experiments that I am considering. This is something that John Richters and I talked about this morning. Let's suppose that in a proposal you were required to take what you consider to be the most important experiment that you are proposing, and pretend that you have done it, and produce data as an appendix literally as if it were coming from the experiment. You could do it by a simulation, for example. That forces you to do a whole variety of things. As I said before, it forces you to specify your hypothesis, thereby providing some insight about whether your hypothesis is a stupid one or a reasonable one. It forces you to come up with an explicit model of error that you anticipate in your experiments. It has other attributes too numerous to mention in this meeting. So those are three modest suggestions.

Dr. Appelbaum: I was thinking back to the things over my career that have been unusual, different, and made some difference. One of them was an NIMH activity, and I am afraid it has gotten lost. It was Vickie Levin. Before she moved to Review she was the program officer for the Prevention Branch. What she would do, is once a year she would bring in her grantees. The people that were actually out there now doing the research would bring them in, bring in some people to help out. At one point I came in to do some statistical work with the group. It was set up in such a way that it was noncompetitive. These were already funded people. It was confidential, and what they were doing is they were talking about the problems that they were having as they were doing the research, and worked as a team trying to solve some of those problems. She had several different clusters, depending on what the content of the problem was. It was truly one of the most interesting and exciting things that I think I ever did, because you had a feeling that there was a level of honesty that you rarely saw in the research

process, because we are competitive and what have you; plus they were dealing with those problems immediately, and they had all sorts of really very good insights for one another.

I thought that was a really novel thing. I have never seen it done any place, except Vicky doing that with NIMH. The other thing is, this goes back to my tirade on training. Another major problem is that senior research scientists are rarely able to keep up with methodological advance. Our training essentially freezes the day we pass our doctoral writtens. We move ahead in our own area and have little time to keep abreast of other domains such as quantitative methods. I do not know about the rest of you, but I do not have time to read much beyond my own narrow area. I am a journal editor, so I spend a lot of time reading papers, 80 percent of which will never get published. It is very difficult. I keep vowing that I am going to learn a new technique, once a month or some such thing, and I really mean it when I promise myself. But I do not do it; I just do not have the time. So if there were a way of identifying some of those advances and getting them into the common domain you might make some progress--- the kind of thing that Mike Maltz is doing. He has a --- you all should get a chance to see this --- a beautiful graphical display on murder rates that immediately suggests a whole series of hypotheses that you would never have gotten by doing any kind of standard analyses. I was showing a couple of people here a product, *JMP*, which is a commercially available data visualization program. You can tell people about it, but unless you enable them to actually use some of these things, they might not become part of the tool kit. I do not think that, in terms of the technology, I was talking just in terms of graphics and statistics. It is the same thing in measurement, in coding systems, in the whole research enterprise, but we essentially "freeze up" in everything but our areas. Bringing researchers "up to date" in research methodologies would certainly be a worthwhile investment on the part of the Institute.

Dr. Jensen: Just to tap in a bit on Geoff's comment. Would it be useful to follow further to say, not just does not count for graphs, but to encourage the use in review criterion --

Dr. Loftus: That would be part of my suggestion --

Dr. Jensen: -- say visualization, et cetera, et cetera?

Dr. Loftus: -- for stimulating experiments. I mean, presumably, the result of that would be the display of the data, presumably, in as insightful a form as possible.

Dr. Appelbaum: This is one of the big concerns of the *APA Task Force*, because we have strongly recommended the use of graphical approaches. This recommendation will have some implications for journal space. We have met with the *APA* journal editors to talk about the trade-off between the number of articles you are going to be able to publish and whether or not you have those kind of graphical presentation --- they take space. At least at the verbal level they would be willing to publish two or three less papers a year, and to dedicate that page space to more graphical presentation. *APA* has made a commitment to allow color and other graphical devices in the journals. We will start with mine. So some of the barriers are coming down. We now have to make sure that the people that are going to use it, know how to use it, so it is advantageous.

Dr. Jensen: Dr. Robinson.

Dr. Robinson: Well, as I am not going to submit a grant, I am liberated from being constrained by my own recommendations! I certainly think that if I were reviewing proposals in an area like this, minimally I would want these considerations included. I would want the proposal to make clear, and distinguishable, the implications of the research for policy, the implications of the research for theory, and the implications of the research for interdisciplinary influences in both directions. What does this research match up with, if anything else? I certainly would want the author of the proposal to instruct me in the assumptions made, what I would call the scaling and mensurational assumptions, such that I understand the rationale for those particular measurements as opposed to some other set of measurements. That is, I can grant that a ruler is a very accurate way of measuring things. If the problem has nothing to do with length or height, the defense of the ruler would be nugatory; it would be just the wrong thing for the task. I would want to know what it is about that particular set of measures that match up in the right way, intelligibly, with that set of phenomena. I would want the author to be the source of my education as to what the most significant limitations are in this research design. I would want the author to serve as a critic of the research design. The author is probably as well positioned in this regard as a study section is going to be.

Dr. Appelbaum: He gives every reviewer the easiest way to do his review.

Dr. Robinson: Yes, well that is right, but better that it comes from the -- then the author understands the grant on which this was --

Dr. Appelbaum: But this is really counterproductive from the point of view of the individual applicant.

Dr. Robinson: Yes, well, I do not think so. Some lessons are learned on the stool of repentance, as you very well know.

Dr. Appelbaum: I would agree.

Dr. Richters: Plus, its always a bad strategy to leave a weakness undiscussed and risk having a reviewer discuss it for you!

Dr. Appelbaum: Oh, no, no! The far worst one is to right there in the proposal yourself, point out the critical flaw. I have done this for enough years to know that is the cheap, easy way to get that review done, and move on to the more difficult ones. Just a reality test.

Dr. Robinson: I would also want the author to disclose, and fairly carefully, the axiological or evaluative assumptions, not only in the choice of terms in the proposal (as in *antisocial*), but in specifying policy implications. That is, there must be some core, a sort of social ethics --- axiology is the right word. It is a pretentious word, but it is the right word. There is an implicit axiology in research of this kind, and I should want that made explicit. Then, finally, I would want at least this much by way of a converging operation. I think unless an argument to the contrary is successfully made, the experiment, along with its sample and the like, really is intended as a model for the individual case. So I would like to see, where possible, a research design that brings the findings from this sample to bear on an individual case, that being the relevant test of the theoretical implications spelled out under implications. If all these findings and the statistical operations performed on them do not really give me much of a handle in the individual case, then I have some question about the extent to which the model is faithful to what it is modeling.

Dr. Appelbaum: Do you have a problem, for instance, that demographers who are studying the population "problem" recently found that the average number of children was 2.5 in this particular time period, and then dropped to 2.3, even though we know that no one has ever had 2.5 or 2.3 children?

Dr. Robinson: No, it is actually not that sort of problem. I will tell you what I do have a problem with. Someone does a study and discovers that, let's say in 75% of the cases, diets that produce an increase in the titre for sugar are associated with antisocial behavior. I do not know if that is the case, but let's say it is. The conclusion reached is that there are significant dietary sources of antisocial behavior. This then should allow me, knowing the dietary habits of Billy, to make certain statements about how Billy is likely to behave. Now it turns out that what passes for prediction is my ability to predict a data set, not my ability to tell what Billy is going to do. Then, in fact, what the model is, is the model of the data set, and not the model of the individual about whom we are trying to make statements. So, no, it has nothing to do with whether somebody can have 2.3 children. Let's put it this way. The Newtonian model tells me something I want to know about the retrograde motion of Mercury. It has to. If it did not tell me that, it would not have last for 300 years. The Einsteinian model tells me that light when it passes the sun, is undergoing gravitational effects.

Dr. Appelbaum: On the other hand, the pressure/temperature volume laws, which are very useful under some circumstances, tell you nothing about any individual molecule. They tell you about an aggregate of what happens when you have lots and lots and lots of molecules. Now I would argue it does not tell me anything about --- I can make no prediction about --- any individual molecule.

Dr. Robinson: May I just say this? If that were going to be your last word on the subject; that, in fact, the gas laws allow me to provide a statistical description of an ensemble of molecules in the aggregate, but will tell me nothing about the behavior of the individual molecule. You then go to Congress for appropriations for a division like this. I think the Congressman that passes something like this probably should be subject to recall. The question

we want to know is how Jack and Jill ought to be brought up in a society in such a way as to live full and flourishing lives. Not that, in the aggregate, you are always going to lose some, but by and large things are going to be okay. I do not think that would match up with your mission. Am I right?

Dr. Jensen: I think we do not want to predict the behavior of gas volumes. We want to predict the behavior of societies. We want to predict and understand children that come into our real world settings and do something. I would like to be able to say --

Dr. Robinson: Do you have to say it with regard to the individual --

Dr. Jensen: Well, hang on. I would like to be able to say -- yes. As a clinician, I would like to say with 95 percent degree of certainty, this is the right treatment for this child.

Dr. Richters: Dan, when you speak of Johnny, you mean *Johnny-like* people, don't you?

Dr. Robinson: Well, there I would say that is an empirical question. Yes, there are Johnny-like people in some intelligible sense. Yes, I am talking about Johnny-like.

Dr. Jensen: Right. So it would be the laws that predict Johnny. They may be the same laws that are offered about many other Johnnies in jails. But I would like to know enough about those laws so I can say, now I understand Johnny.

Dr. Trickett: But you do not have to. You can be concerned about population outcomes, on the basis where population then would be reasonable model.

Dr. Appelbaum: For instance, consider picking a reading model -- assume you are working in a state system where they are going to either be teaching phonics or whole word. We know damn well that if you pick an individual at random it is going to be very difficult to tell whether or not that person will do better under a whole language or a phonics. I may have to make a decision that we are going to either teach phonics or whole word, even though I cannot predict the impact on any given individual with any great certainty, beyond knowing that 75 percent of the time or whatever, a kid will do better under this than that. I can see at points where, yes, we would want to. Our test would be the degree to which individual behavior can be understood and modeled from that, but I do not think that is a criterion. It may be a criterion for some applications.

Dr. Robinson: Well, they are not. There is something about Ryle-type category mistake here in assuming that this sample is something other than a constellation of individual cases being worked on. The reason generally for looking at the ensemble is that life is tough, variables are numerous, and what we are going to have to do is some kind of smoothing operation, and make a best guess by getting significance out of a large enough sample. The point of doing it is because you want to know something about what makes people tick. There then has to be at some point a converging operation that says, okay, here are the conclusions reached by this research. Now let's see if it tells what makes *Jack* tick, you see. At some point, it is going to have to do that, but I do not think the analogy or metaphor, or however you intended the gas laws of Charles and Boyle, really do match up with developmental psychopathology, at all.

Dr. Richters: Well, in one sense we would like to know what makes *this* Johnny tick in the sense of --

Dr. Appelbaum: Doing what is called a primary prevention, like the fluoride in drinking water. So we decide to try to bring in the ad counsel to mount a campaign because we believe that understanding of this will probabilistically lower the rate, will not take care of Steve, who is going to go out there and do his antisocial thing, but probabilistically.

Dr. Jensen: Geoff, and then Mike, and then Steve.

Dr. Loftus: I just want to go back for a second to Mark Appelbaum's example of whole word versus phonics --- there are lots of equivalent examples of course. What you want to understand depends on what you are going

to have available to you. So if you are going to have to make a decision, as in Mark's example, about whether you are going to teach a whole community via one method or the other, then it suffices to know what is best for the whole community. You can understand things at the statistical "gas law" level. Suppose on the other hand that you have a computer instruction system, such that any given kid can be assigned either one method or the other. Well, now it becomes a lot more important to understand what is going on at the level of the individual kids.

Dr. Appelbaum: And later we want to know what is this process --

Dr. Loftus: I mean it goes back to Dan's original point, that what determines -- what you want to know from the research, at least to some degree, is the desiderata of the society.

Dr. Jensen: Mike and then Steve.

Dr. Maltz: Something that Dan mentioned made me remember something from my classical physics course, the ergodic hypothesis, where the trajectory of, say, one molecule in space, may mimic an ensemble of trajectories. In other words, is there a relationship between cross-sectional and longitudinal data, and I think that in this field there is not. Again, this is why I am pushing for more longitudinal analysis, perhaps case studies. To get off that for a second, and to get on to something that Geoff and Dan mentioned about, what about these criteria for -- these additional criteria for submissions?

Well, I tell you, the more criteria you put down, the less likely I am going to be to submit. I look at those forms and I say, "Dammit, I do not have two and a half to three months to figure them out, to put together a submission." So I think that you should be concerned about that. One other point Mark was talking about. Well, would it matter if there were 2.5 children before in the average family versus 2.3 now? I say those numbers are suspect, that those are averages, and we should not be looking at averages. What we should be doing is looking at distributions, because those two numbers are meaningless. Again, because nobody has 2.5 or 2.3 children.

Dr. Appelbaum: Well, it depends. If I have to then forecast what we need in terms of school resources for the next five years, the average number of kids per household is very useful.

Dr. Maltz: It depends upon --

Dr. Appelbaum: It depends again on the use and the level of question.

Dr. Maltz: -- on whether those who have seven children are becoming less frequent in number, or those who have one child are becoming less frequent in number. So I think you do have to look at the distributions, rather than just -

Dr. Appelbaum: I am not going to disagree.

Dr. Maltz: Pardon?

Dr. Appelbaum: I am not going to disagree. I mean, we are talking about different --

Dr. Loftus: Situation of distribution assumes --

Dr. Maltz: Well, but the point is that all of the experiments that are done usually look at the mean as the only

--

Dr. Jensen: Steve.

Dr. Hinshaw: Just picking up on Mike's last point. You were reminded of an undergraduate physics course. I am reminded of the first study I did some years back of stimulant medication effects on kids with attentional

problems. It was published in a prestigious journal (Journal of Consulting and Clinical Psychology). I found that Ritalin showed a linear dose response curve for reducing the aggressive behavior of such children. What I thought was the showpiece of the work was that when you examine the individual dose response curves of each participant, precisely 2 of 25 kids showed that same trend. This is the old "patient uniformity myth" that Kiesler discussed over 30 years ago. Yet the reviewers did not like discussion of such individual differences, stating that the field does not have statistics for such examination. "How can you talk about these individuals? Keep it to the group level." So even in the relatively basic clinical trial method, if you are at the group aggregate level, you may be misrepresenting nearly every individual. Fortunately, more modern approaches to clinical trials are starting to take notice of this key point. For example, random regression models take into account modeling the trajectory of change or growth in each individual case.

Dr. Maltz: Only 23 out of 25.

Dr. Hinshaw: Well, that is right, it was a lot.

Dr. Maltz: That is the converging operation here. That is why I am passing around those transparencies, showing what a difference you see when you look at individual level data versus -- imagine putting that data set into the more of SPSS, and having it crunch, and you would get a very high correlation with absolute statistical significance to 15 decimal places. It missed something. You miss about four or five different really important features.

Dr. Appelbaum: By the way, you see that is why individual growth curve modeling is so much better.

Dr. Jensen: Yes, right. I am wondering if we could pick up on Dan Robinson's last point about converging operations. Frankly, this is the issue with which we have been wrestling, but we are having a hard time wrestling with it. What might such a study look like? It is one thing to tweak review criteria. And, as you point out, it is already bad enough to submit a grant. But what might studies look like that take advantage of converging lines of evidence? A great example last night came from Mark Appelbaum. He said, you know when they published on cold fusion, a bunch of people all around the country said, that cannot be true, and they rushed to their labs, and within two weeks they said it was nonsense. Is there a way that we can design studies that bring enough converging lines of evidence so that someone could say at the end of a non-experimental study (by relying on converging operations) -- "Wow, that really seems to make sense?"

Dr. Fox: I do not know that I would advocate a study, but I think Mike mentioned earlier this morning secondary data analysis. I think if you wanted to replicate that parable or story that you told in cold fusion, then there should be some way, if someone has a multivariate longitudinal data set, and they come out and they say, oh, look what we found, for that data to be available for other investigators, for a panel of investigators, to look at again, and to see whether or not they can replicate that.

Dr. Jensen: Okay. But I will say, that is pretty easily done. But I want you to wrestle at a deeper level.

Dr. Fox: If it is easily done I would like to know where it is done.

Dr. Jensen: People do have data sets and are making available them available as such through Cornell and other places. Such data sets and archives have been established. So whether it is easily done or whether they were paying for it is one question. But I would like to push us further down into theory and perhaps even into epistemology.

Dr. Robinson: Well, that is a good example then, if that is the way you would want to do it. You must be tolerant as I review Psychophysics 101. Let's take a look at the difference between a classical psychophysical experiment and something done under signal detection terms. In the first case what we do is use catch trials to eliminate or greatly reduce guessing, the threat being that if the subject guesses on the catch trials, the whole data set will have to be done over again. Under the second set of circumstances the subject is encouraged to guess. If you did not get it right during that interval, what interval do you think the signal might have been there? Then you look at the *a priori* probability of guessing right under these different arrangements. Now when you go from the classical paradigm to the signal detection paradigm, you discover the threshold is much, much lower and you

might even raise the question under those circumstances, of whether there actually *is* a threshold.

Similarly, in psychophysics, which has wrestled with scaling problems, even since the time of Weber, there are very strong anchoring effects that can be measured in a psychophysical context. The rating scale behavior, with respect to physical stimuli, will be radically altered, depending upon whether the ratings themselves are linearly arranged, or whether you have outliers that the subject has to deal with. Now suppose you went into an area like developmental psychopathology, about which I think I can say I know nothing. I assume some of this research assigns values to the children's behavior. This is how you end up with some statistics to do in the first instance. If you were dealing with a genuine natural phenomenon, then the measurement of that phenomenon should be relatively resistant to anchoring effects. That is, if you gave the raters nonlinear scales to deal with, where there were several extreme values at one end, and not many at the other, you could not pull the ratings all the way in that direction merely by way of anchoring. It would not happen if they were measuring core temperature, or digestive physiology, and the like. The more and more judgments are subject to anchoring effects, the more those judgments in fact are inextricably bound up with factors, which for want of a better term, we would call subjective and contextual.

One thing I would be interested in, in a rating scale type approach to antisocial behavior, is the extent to which the scale itself has been properly scaled. Can you manipulate the scale values and still preserve the phenomenon? If not, you may be looking much more at how people use scales. In psychophysics we know that not all numerals are numbers! We know that when you go from the street address, 300 East Main Street to 150 East, you have not cut anything in half actually. So I think there are probably considerable scaling problems. To say no more, there are probably *mammoth* scaling problems here, and one way of doing a study that involves converging operations is to see whether in manipulating the scales you can still get the same effect. You asked for a concrete case.

Dr. Jensen: That is a very interesting idea. Bob?

Dr. Cairns: Yes, a couple of comments. One is, I do not think it is that easy to bring together longitudinal studies. I think it is easy to do at the variable level, but not at the concrete operations level. NIMH has invested in an awful lot of funds over the last two years in longitudinal studies in this area. I felt there is just a crying need to bring together people like Mark, actual investigators that really know their data down to the concrete level, as opposed to the foggy variable level, and then to examine for commonalities across studies. That is a big task and it takes a lot of honesty and exposure by every investigator to let other people snoop around, and see what they actually did. But that would be a great advance if we --

Dr. Fox: Right, but that would begin to address the problem which is being raised here, aside from the fact that it is not a new research. It would begin to address this now.

Dr. Cairns: Ken, I will yield, but that is not really what I wanted to say. It was just kind of a thought. But I want to come back to the issue of convergent, because I think that is one area of convergency.

Dr. Rubin: Well, I was just going to follow up on what you were saying. We convened a conference on the development and treatment of aggression in 1989, I believe it was, and then we published a book in 1991. Almost everybody who attended the meeting was funded. They came to this meeting; it was in Toronto. At the formal presentations everybody gave their best show, and said " *here are our predictors of violent and aggressive behavior.*" Ken Dodge would talk about social cognitive deficits; Gerry Patterson would give his spiel and so forth. Then for every sort of data piece, we had people working with those same data sets, but who were doing intervention sorts of studies. It was really interesting.

So those with family models of the development of aggression, were presenting with people who were doing family therapy interventions. Then we had dinner with these folks. In casual conversation, when the microphone is off, you start asking *what is the take here, and how effective are your treatments?* For every positive, every significant finding, every publication that gets accepted, there are 9 or 10 that do not work, using the same procedures, and you suddenly start wondering what it is gets published! I was reminded at that instant of work that was going on in the area of peer relationships, where there were specific characteristics of rejected children that were pervading the literature. Other characteristics that were not deemed associated or related with rejection by

the mainstream researchers, that seemed to others of us, just intuitively, *ought* to be associated with rejection. It turns out that the minority opinion was valid and accurate. They just could not get published. Once "reality" is published, if you fail to replicate the one model study you were considered unable to replicate "reality", the acceptable! I think that this a dilemma for the field, in that, that which is published is published in stone, and 9 out of 10 investigators may not come up with the same findings, using almost identical procedures, and it will not get published because the original study's results are considered "real" and carved in stone. The other part of it was this admission, with microphone held like this (unplugged), that most of the aggression intervention procedures did *not* work!

Dr. Hinshaw: Bill, this reminds me of what you said in your '83 meetings, about what we are training residents do is to be entrepreneurs, to present finally after 15 failed tries, the one.

Dr. Jensen: In a way, Dr. Suppe -- but it reminds me, your comment, the notion of converging operations as you have described it, and has Mark, basically add converging scientists to it, in bringing multiple manipulations of data and exploratory in a very interactional context.

Dr. Rubin: Actually, the most productive part of the meeting was not the book, but the conversations around the table.

Dr. Jensen: All right. Dr. Suppe, and then -- boy, I am having a hard time with names today, Ed. You might introduce yourself to the group, Fred.

Dr. Suppe: Yes, I am Fred Suppe. I am chair of the History and Philosophy of Science Program, University of Maryland, and I am Visiting Distinguished Professor of Nursing at Columbia. The question on convergence from an epistemological approach I think comes down to this. When you look at a population and the standard population statistics, both the mean and standard deviation are telling you about what is happening in the two central -- the two middle quartiles. When you are applying two individuals, you are primarily concerned with what is happening in the upper and the lower quartile, and if you want a reporting of data that will bring the two together, what you need is a measure of central tendency, and it should not generally be the mean, because if your sample is n is below 400 the mean is more bias than the media, so you want median. You want a real spread on the upper and the lower quartile, and you want an examination of the high and low points of data. If you put those together, you know what is really going on in the population. Most of the information is really in details, a reporting is mostly of what the central tendencies that leaves out most of the useful information applying it in concrete circumstances. If you have a reporting that brought all those pieces together, you would essentially get your population central tendencies, but you would get the useful population data, apply it to the ends. By the way, this is not an original meeting. This essentially John Touke's first pass, exploratory data analysis by elements of reporting. I would suggest that the heart of the epistemologic problem is precisely what quartile you are looking at for what classic problems.

Dr. Jensen: Okay. Ed.

Dr. Trickett: I was thinking about the idea of convergence without necessarily the goal. I was trying to link it to some of the concerns about the degree of which findings are tied to methods. I was thinking about antisocial research methods, and certainly risk factor research, which does not express any interest in the meaning of antisocial acts to those doing them. I was wondering about whether or not parallel stories of those kinds of phenomena would be enriching. I am not sure they would converge. Mike Agar published a nice book on heroin addicts, for example, that talked about the reasons for heroin addiction in New York -- how addicts made sense of their own behavior, which is absolutely unrelated to the way social science makes sense of the heroin addict's behavior. I am just wondering, in terms of those multiple methods, if that would provide certainly an illumination for theory, as opposed to convergence being the goal.

Dr. Jensen: Could the multiple methods converge, or --

Dr. Trickett: Sure.

Dr. Jensen: -- I mean they may or may not, right?

Dr. Trickett: Sure.

Dr. Jensen: Wouldn't that be one way to see across different scalings, if you will, if the phenomena hold up under different --

Dr. Trickett: Sure. It gets back to the use of discovery. I mean, if mothers rate kids in a certain way, and kids rate kids in a certain way, and teachers rate kids in a different way, within the discovery perspective the idea is to understand what it is about those different raters that yield different outcomes, as opposed to whether or not they can converge.

Dr. Appelbaum: Or whether really kids behave differently when observed by those three sets of observers.

Dr. Trickett: Right, sure, exactly.

Dr. Jensen: Dr. McGuire?

Dr. McGuire: Is it okay to get off the convergence topic?

Dr. Jensen: You want to divert.

Dr. McGuire: This afternoon we are supposed, I think, to come up with what is to be done and how it is to be done within NIMH prerogatives. We might carry this out by constructing a matrix whose row headings would be the content innovations that we feel are needed in behavioral science research and whose column headings are mechanisms that NIMH may have at its disposal for encouraging innovations. In each cell of the matrix we can enter how the column mechanism can be used to promote the row innovation, leaving cells blank where the intersection seems weak. For example, as regards the column headings of the matrix (the mechanisms that NIMH has at its disposal) a preliminary list (that need reorganization and extension) might include setting up a methodology (or "psychology of science") review committee to handle applications that propose to use dramatic innovations that are to be promoted (e.g., analyses that exploit longitudinal designs). Or NIMH might at least designate an ombudsperson to spot and track proposals using desirable innovations through the ordinary review committees. Also the NIMH might sponsor summer workshops to bring together scattered researchers who use approaches like content analysis of open-ended response. Also, NIMH might set up email addresses to facilitate contacts among members of invisible colleges, or support curriculum development such as developing syllabi or textbooks on mathematics for the behavioral sciences. Another NIMH mechanism could be awarding training grants to centers with strength on underused approaches such as employing statistics as discovery processes rather than just as tests of a priori theories. Or pre- and post-docs (junior and senior) could be awarded to individuals to work in these areas. NIMH could provide subsidies to appropriate journals for the added costs in implementing recommended approaches (e.g., graphical representations of relations, even color graphics).

As regards the row headings of such a matrix (the topics to be encouraged by the NIMH mechanisms) most of us have been mentioning a variety of these all morning. I just mentioned 5 or 10 of them, like longitudinal methods and graphical representations of relations. Others to be considered which I recall hearing this morning include secondary analysis of archival data, meta-analysis, case histories, heuristics for creative theory generating, macro level analyses such as using nations rather than individuals as units of observations, as in psychopathology studies like Dane Archer & Gartner's (1984) 110-nation study, Violence and crime in cross-national perspective. But this is just a sample of innovative-topic nominees, likely to be extended this afternoon and in follow-up meetings.

Dr. Cairns: Can I add a couple?

Dr. McGuire: Yes, yes, please.

Dr. Cairns: Bill and I were talking about this in between one of the sessions. I am thinking of the P50, P01 type mechanisms that do research at different convergent levels under one umbrella (e.g., interventions,

longitudinal, short-term experimental). These integrated funding mechanisms do not seem to be in favor nowadays, but that seems to be precisely what is called for if we are going to bring together these different worlds of methodology. Another way would be to cut across branches. Many of the branches of NIMH are problem-oriented or methodology-oriented. Perhaps the branches could get together to form common frameworks. Moreover, there could be a variety of integrative training conferences. Not just summer institutes, but even shorter term workshops. When we talk about methodology and measurement, or theory of measurement, or logic of measurement, it is not immediately obvious that self report should converge with external reports on all domains. You may have ipsative theories by the self that are quite different from the normative theories by others, but address some of these issues at an intellectual level for the field.

Dr. Maltz: What is a P50 or a P01?

Dr. Cairns: P01s are integrative studies, different investigators, even from different institutions. But I do think there are mechanisms available that could and should be exploited to address some of these advances.

Dr. McGuire: Let me ask a question about how NIMH works. My observation of its workings has been mostly limited to review committees. Is it feasible to ask a wide range of review committees to be more positive on proposals that, say, use statistical analyses as a discovery process? Or can or should such weighting be introduced at the subsequent Council review? Should proposals include a checklist on which the PI can indicate employment of designated underused approaches?

Dr. Richters: There are several answers. One possibility is to use set aside money to attract and fund the necessary research using the *Request for Application* (RFA) mechanism. Grants submitted in response to RFAs are not reviewed by the standing IRGs.

Dr. Appelbaum: It is very difficult to influence a study section to do anything other than what that study section is going to do.

Dr. Richters: The job of an IRG, in fact, is to serve as a standard bearer for the paradigm.

Dr. McGuire: All right. We may be trying to encourage, say, using statistical analyses as a discovery process. Applications proposing this use might come in areas scattered from childhood psychopathology, to critical flicker fusion frequency and everything in between. Can review committees be expected to have expertise in such techniques as well as in the committee's subject matter area?

Dr. Jensen: Go ahead, Ellen.

Dr. Stover: One of the things that I alluded to in the very brief opening comments this morning, is we are truly in the midst of redesigning study sections as we speak. So these kinds of issues are very critical to get, not only on the table, but into our process. So you do have a unique opportunity right now over the next couple of months, either as a group, or individually, to E-mail us or give us those kind of thoughts. That is off your question. I think John was still in the middle of answering it. But the other distinction then, is that kind of effort, if it is mission-specific, would be reviewed within the institute with a special kind of review group. The effort I am talking about would go over to the centralized DRG. It has a new name now.

Dr. Richters: I would be interested in any thoughts about how we might stimulate this kind of dialogue at a more detailed level in substantive research domains, deciding first what those domains should be, who the people should be, whether or not we should be developing pockets of activity devoted to these issues, such as meeting with journal editors--- three or four of whom we have here today and with whom the Institute has been meeting in the child adolescent area for several years. That is, steps we might take to help the field --- to lift existing barriers and signal our commitment to reform and re-tooling efforts where necessary.

Dr. Robinson: Well, Geoff Loftus engaged in what sometimes is referred to as a futile gesture, which is to say an absolutely heroic one, doomed to fail. I must say, preparing for these meetings I read your editorial, *Counsel to Contributors*, Geoff. I had not read it when it came out, but I said, "Fantastic." Then I asked Geoff,

what effect did this actually have on what was submitted. I think you said not much. Is that right? Not too much?

Dr. Loftus: Yes. One of my colleagues, Tony Greenwald, had the discourtesy to point out that every empirical article in my journal had used hypothesis testing. I was delighted to see at Psychonomics a week or two ago that virtually every data set I saw had different error bars. I, of course, attributed it to myself, but in reality it was probably because the software packages these days lend themselves a little bit more to computing error bars than they used to.

Dr. Robinson: Well, you could not have had a paper in psychophysics published before 1940, where you did not have confidence levels included. And, in fact, before 1935, I think you probably had to include all the raw data.

Dr. Jensen: Let me pick up on Bob Cairns' point about these broader mechanisms, and raise the question, where does an experimental study say that basically hetero-science is where you put two or three little critical tests of a guiding hypothesis together, and that be part of a converging operation, at least inductively in the sense. Do you see a role or a possibility of how that might be brought to bear in a similar way in nonexperimental studies, and how might such a thing look?

Dr. Cairns: We have a prevention branch, basic research branch, and other branches, that in some respects regress to the same issues at a logical level and explanatory level. I think this is the way it is conceptualized at NIMH. Those prevention studies are now viewed as naturalistic, or non-naturalistic interventions, that are field and experiments. To break down these distinctions between what we call longitudinal studies and prevention studies, and to lend that also to short-term experimental studies under control conditions. This is the way they do in animal behavior. They move from the lab to the field, back to the lab to test a specific hypotheses. We tend to not do that in human behavior, even though the logical issues are exactly the same. If we can think of models to at least mimic some of the successful procedures in other sister areas of science, I think we will be making some headway. The R1 mechanism does not support that sort of thing. I mean, you are in big trouble.

Dr. Jensen: Not that it could not. I mean, you could argue the R1 mechanisms better support that in the basic behavioral sciences where --

Dr. Cairns: They do it animal behavior.

Dr. Jensen: You have 3, or 4, or 5 tests that are built into this same R1.

Dr. Cairns: Yes, and in behavioral pharmacology you have the same sort of thing. Where problems arise is when review groups are sent reasonably sophisticated behavioral analyses, combined with reasonably sophisticated neurobiological analyses. Bu panels rarely have expertise in both domains, so that the total package is not considered or evaluated. You do not see the integration evaluated, only the separate components that are represented by the primary and secondary reviewers.

Dr. Fox: Let me just second that in saying that, I think someone mentioned the multi-method, multi-measured approach. In general, those kinds of approaches are frowned upon because they necessitate at the level of theory or model some sort of integration across measurement or approach, which may not be the current state of the research. They are hypothesis-generating and not hypothesis-testing. They are frowned upon by the study sections because you have your neurophysiologists on the one hand, who will criticize the use of the behavior, and your behavioral researchers who will criticize the level of the neurophysiology. So to the extent that the Institute could support that kind of effort, I think it would be a step in the direction of the multi-measure, multi-method.

Dr. Jensen: It gets to your point about the scaling problem.

Dr. Robinson: Well, sometimes the role of this is absolutely pivotal. Some years ago, 20 years ago, I could not make sense of electroconvulsive therapy (ECT) literature. ECT was quite effective in treating depression, in treating classes of schizophrenics in England, but not so here. I was giving a lecture at the Royal Free Hospital, and the Director of Psychiatry there at the time was Alick Elithorn, I said, "How do you account for that," and he smirked. He said, "Oh, well; it is a hell of a lot harder to be classified as schizophrenic in the UK than it is in America". He simply took it for granted that the effect had nothing to do with the relationship between

schizophrenia and electroconvulsive therapy. It had to do with how you *scale* schizophrenia and the effects of electroconvulsive therapy. Most of this is just folk wisdom, but the question is how you get that incorporated in to the research protocol, so that you really know what *Smith* means when *Smith* says "antisocial behavior". Does he mean what Jones means? I rarely see, when I do consult literature of this sort, attention to such matters. I do not know the last time I even saw reference to an attempt to determine the validity of the scaling operations themselves. They are more or less taken for granted. Nouns are used as if they were quantities. People are given rating scales. You are not going to get anywhere until these terms of art -- all the statistics in the world will give you mere mush, if at the lexical level you are dealing with mush.

Dr. McGuire: But some progress is being made in these areas. For example, Luce and Tukey's conjoint measurement and Norman Anderson's functional measurement approaches allow researchers to scale the variables and quantify the relation between them within the same experimental design.

Dr. Robinson: Is that common now though, Bill?

Dr. McGuire: No, not as common as it could be, but there is some use of such procedures.

Dr. Jensen: How does he do that? Could you just say something about that?

Dr. McGuire: Well, without getting too technical, one can start off with a model, say, a parallel or a fan model depending on whether additive or multiplicative relations are theorized. Or one can use and exploit novel response tasks, beyond pick-one and paired comparisons, such as those Clyde Coombs (1964) proposed in his *Theory of Data* book. Or one can use the kind of cross-modality scaling that Volney Steffire used years ago (although his work was sponsored by whiskey companies more than NIMH).

Dr. Robinson: Which one is the *holbein*? You remember that? That was Fechner founding "experimental aesthetics": which one is the real holbein?

Dr. McGuire: These things are not done as much as they should be. But I mean, at least we can point to some of them-- this is an example, not a great one, but a good one.

Dr. Appelbaum: If we can go back to study section behavior. I cannot remember in all the proposals in eight years on a study section, and we got measurement proposals, and they failed, they failed, they failed. Study sections would not recommend them out-- *This was work that should be done before you...*

Dr. Fox: Proposals that were strictly designed for investigative measurement.

Dr. Appelbaum: Measurement, measurement issues. How do you measure? Instrument development is not something that a review committee will generally look favorably upon.

Dr. Hinshaw: Yes, I just wanted to take up a different point about the grant world. I think there is a tension. If people who study antisocial behavior do what a lot of us are suggesting, which is to cluster, to design confirmatory measurement studies within larger longitudinal work, these endeavors will be time-consuming and costly. It does not take too many of those studies to put a squeeze on an agency's grant portfolio, I would think..How, at the same time, are we to encourage younger investigators, or people doing ethnographies, or people performing secondary analyses, or looking more qualitatively, to get their own research programs going and to interact with the people in charge of the big longitudinal data sets? There has to be a pluralism, but there also has to be some integration among the elements of such studies.

Dr. Jensen: Geoff.

Dr. Loftus: I just wanted to second Dan's opinion, that scaling is really at the root of a lot of problems in my opinion; certainly that is the case in my area. They think of Clyde Coombs and Duncan Luce and they think: "*ugh, mathematics*". Mathematics is one thing I do that is very straightforward and it helps a lot in terms of at least starting to deal with scaling issues, is to insist that people --- either writers for journals or grant proposals ---

specify explicitly what they believe to be the relation between the dependent variable that they measure, and whatever internal construct the dependent variable is meant to reflect. Even if they just indicate that it is assumed to be a monotonic relation, that at least sort of gets you started in terms of understanding or, as a reviewer, being able to understand, for example, whether you believe that different measures are measuring the same thing, or whether they are measuring different things.

Dr. Mitnick: Part of this is -- and Mark is absolutely right in the study sections that he has been on, and I am sure you fought very hard to have all these methodological studies --

Dr. Appelbaum: Some of them, not all of them.

Dr. Mitnick: Kelly and Pugil had it right. We have met the enemy and it is us. So it is the people sitting around this table who you are talking about, who would not support those kinds of studies.

Dr. Appelbaum: Well, it is because competition for funds, studies introducing new data were inherently better than studies that were infrastructured.

Dr. Loftus: That is not the purpose of the review committee. The purpose is to look at each individual application.

Dr. Jensen: Well, to further pursue the scaling point, there are some relevant studies in our portfolio. My sense is that they might be increasing, such as when someone samples within an epidemiologic framework, obtaining information from multiple informants. But they also go in under blinded conditions, perhaps doing some high/low sampling on some risk factors. Then they send ethnographers in. Basically, this relies on a scripted approach, to say we want you to find out **X**, **Y**, and **Z** of the following. While you get one thing from your standard survey measures, someone else is going in blinded to what you may have gathered in this other "objectively gathered" data set. And they say, "Now let's really figure what is going on, constrained according to certain development models and approaches." And so it requires the nesting of two different approaches, and they look for the convergence of concurrent and the prospective information. We are seeing a little bit of this type of research, but it is certainly not yet mainstream.

Dr. McGuire: Just a brief point to build on something Steve mentioned a few minutes ago, that after we nominate individual techniques that deserve more use, there remains the problem of putting them together in some organized way. But perhaps at this meeting we can settle for just listing isolated needs. We could leave for a later meeting, after we have had time to incubate, our proposals for organizing these isolated recommendations more systematically.

Dr. Jensen: Yes, Dr. Suppe.

Dr. Suppe: A couple of scattered thoughts relating to the issue of scaling, relating to the issue of how do you do the kinds of development of instruments, the kind of methodological developments, without it becoming a huge high-funded, career-long process, how does the young guy or woman get into this game? A couple of scattered thoughts on that. First, one thing journals could do would be to insist on the text size as a part of the reporting and each piece of data, which would then allow meta-analysis as a way of getting some of the benefits of large studies by just simply looking at the population of studies done. If you have text sizes, you can do it, if you do not, you basically cannot do the meta-analysis. The second thought is, even if we succeed in getting people to do more adventuresome and more thoughtful research using something other than just hypothesis testing, where are they going to publish it?

Dr. Loftus: *Memory and Cognition.*

Dr. Suppe: The basic point is that a lot of journals -- I will give you an example. Look at ACA guidelines and supporting data. We are talking about doing things that are not allowed. I mean, you should be showing not just one competence level, you ought to be looking at sub-significant levels, the ages, and so on, and seeing how robust even the levels are. You cannot report non-significant data under APA guidelines.

Dr. Appelbaum: That is absolutely not true, and it has not been true for 20 years.

Dr. Suppe: Okay, then where does it get published?

Dr. Appelbaum: In the publication manual, the segment that says "only significant data should be reported" was eliminated .

Dr. Suppe: Oh, good, glad to hear it.

Dr. Appelbaum: -- at least in the last, two editions. I am sorry. I am very sensitive to that one.

Dr. Suppe: The basic point I am trying to make is that, the journal editors have to be actively involved in having review standards that allow, encourage this kind of methodological adventuresome.

Dr. Jensen: And the funding incentives for the grants and so forth, we have to signal to the field and put our resources behind it, our willingness to be bold and rigorous in thoughtful ways.

Dr. Suppe: And NIMH could do things, like for example, have a program of subsidy to journals that are willing to open things up methodologically, something like that.

Dr. Jensen: Rather than forming another -- yet another journal that will be marginalized, taking the standard there in journals, and giving them incentives to -- interesting idea.

Dr. McGuire: But then we will get associations, researchers, journal editors, saying *you are not going to dictate what we publish!*

Dr. Jensen: Nobody wants to. Nobody wants to.

Dr. Appelbaum: Right. You are just going to subsidize us if we do what you want. Of course, you are not going to try to force us or influence us. Get real.

Dr. Suppe: We have this program for subsidizing journals to do this. If you are interested in getting money for your journal and it is something you want to do, fine, you are welcomed to.

Dr. Jensen: You are perfectly free, Mark, to be --

Dr. Appelbaum: I would not dare speak for APA, but I can just imagine the reaction there. As chairman of the APS Pub Committee, I think I know quite strongly what I would recommend to our organization.

Dr. Fox: Mark, you have to draw the distinction between bribery and blackmail. This is bribery.

Dr. Appelbaum: I see. Talk about antisocial behavior.

Dr. Jensen: But, seriously, Mark, if one of the obstacles, one of the difficulties, is that the journal editorial process is very fixed in defending this monolithic approach to science, then one of the solutions must be, at some points in some ways, a willingness of journal editors to entertain the problem, and embrace a commitment to working out solutions.

Dr. Appelbaum: I think the way to do that, however, is to engage in dialogue, not to get into either bribery or whatever you want to call it. I think also, John -- I don't know. I do not think you would find as high an agreement about monolithic models of science among journal editors as you are going to find in this group. I think that there

has been probably -- there has been a lot of discussion of these sorts of things, and I think if you look at the journals, at least the APA journals, and what is being published now compared to what kinds of studies were published, 15, 20, 30 years ago, the Board of Scientific Advisors, with regard to things like null hypothesis testing, and alternative ways, I think you are going to find at least the perception of those groups is that science is not as monolithic, that there have been considerable changes. I mean, APA journals are publishing ethnographic studies. Can you imagine that?

Dr. Jensen: So do you see the problem is basically we need more time and more money, but we are essentially heading in the right direction of science as basically sound in most areas?

Dr. McGuire: Well, I think we have to make distinctions on subsidizing journals. Subsidizing journals for the incremental costs of using color graphics where they are useful might be acceptable. But subsidizing journals to accept and print articles that are tougher on null hypothesis manuscripts might be an unacceptable intrusion.

Dr. Appelbaum: Actually, in a way you already do it. Because I think, at least the way APA journals are going now, if you want color publication you have to pay a small page price, which then can come out of grant support.

Dr. Maltz: We have a special issue, Mark. Is that just as coercive, money, a subsidy for a special issue? Would that be as coercive?

Dr. Jensen: Yes.

Dr. Maltz: Okay.

Dr. Appelbaum: And, besides, individual editors, at least to the APA journals cannot do that. The Pub Board has to approve special issues.

Dr. Jensen: We have a lot of interests here. Let's see, Ed and then Mike.

Dr. Appelbaum: One thing this group might recommend is that the journals have data archives. Publishing data in print form is not a very good form, because you are probably not going to be able to punch it in right. But the idea of data archives, which are maintained --

Dr. Cairns: Exactly. On the positive side, again, I just wanted to underscore that one of the points Bill raised, and I seconded it, was something we have been learning I think about antisocial behavior. Sometimes the best defense is a good offense to positive engagement through seminars, and institutes, and workshops, and that is a very productive way to spend money to change the field. Along that line, there is another positive feature of studying antisocial behavior. There is a positive side of dealing with essentially relevant variables. Most of the esoteric issues that come up in the laboratory on psychometric scaling are handled rather nicely when we have the concrete realities of life. There is little ambiguity when you count the number of homicides, the injuries that occur, or the fights that are experienced. These are concrete realities that can be pointed to, observed, and analyzed. Such concrete referents transcend many of the problems that have arisen in conceptualizing and operationalizing psychological constructs.

Dr. Jensen: I think Ed was next, and then Mike.

Dr. Trickett: I was just thinking, Peter, as a member of the editorial consortium over the number of years, about how one of the editorial constraints around various kinds of methodological issues involves not having the developed knowledge base in, for example, ethnographic work or qualitative work that is available in others. One of the possibilities for either promoting multi-method or deviant-method approaches will be to support methodological development per se, and it can be in a variety of areas. Again, that is one of those marginal things just like measurement in terms of how it fares here, but it seems like a relevant kind of way to support the larger enterprise.

Dr. Jensen: Great. Mike.

Dr. Maltz: Yes, we have been talking about trying to change a field and finding a good lever for doing it, and it may be the journals, but if you read the instructions to contributors -- how many people have read the instructions to contributors in the last 15 years? Really? I never look at them.

Dr. Appelbaum: But if you only send us two copies of the manuscript, we are going to send it back to you, then you have to send five.

Dr. Maltz: What I can do is-- in fact, one of my associate editors, John McCord, suggested this-- is write a different kind of letter to reviewers (and I am thinking now of sending that letter to the authors as well, saying, please look at the following issues: Were the tests used appropriately? For example, if you want to encourage replication, is it a replication that is worth considering printing? For example, is the methodology unique? Is the data set of importance? Does it open up a new area? Ask them questions of that sort, rather than only "is there statistical significance". I do not have all the ideas yet as to what is going to go into that letter, but I am thinking of using that as a mechanism whereby I can suggest to reviewers that they look at certain things.

Dr. Jensen: Nathan.

Dr. Fox: On the other hand, as a journal editor you are on the receiving end of the process, and not at the beginning of the process.

Dr. Appelbaum: The next editor that comes and says, "Where are the significance tests?"

Dr. Fox: So I would say that, if you ask an individual to search out in the field, well, what is going to affect your decision in terms of how you are going to articulate and design your next study. It is not the editorial in the journal, but it is what NIMH writes out in its RFA for the kind of research area that they see as important over the next five years.

Dr. Richters: Point taken. The common complaint in the field is that existing funding and editorial incentive structures are biased against change.

Dr. Jensen: Mark, and then Dr. Robinson, and then Dr. Suppe.

Dr. Appelbaum: We have talked about how to change things on the input side. There is also the "down the road" side of it. I participated this last year in a process that I had never heard about before. It is called a Dobbing Conference, after Professor John Dobbing in England. Of all things, this had to do with the effect of LC PUFAs, long-chain fatty polyunsaturated acids in baby formula. The question is, should fish oil supplements be used in formulas. It was a really quite amazing operation in the sense that it was set up in quite a different way to try to do something kind of like a consensus study, only it was centered around the book rather than a conference. So participants did chapters. Chapters circulated before meetings, feedback was given to the author, revisions were made to those chapters, and then a three-day meeting was held where everyone had already read everything. The conference focused in on what we knew, what we could agree upon, and what we couldn't agree upon. It led to a book that was published in less than 6 months. It was a quite different approach, but it was one that was quite amazing, because it brought in people from all the levels. From basic biological science (visual retinal developmental system that these fatty acids influenced) all the way up to sociodemographic factors. So something like that, in terms of kind of, what is it that we know, and then what are the holes. What should those RFAs that you guys put out that are targeted -- what should they look like?

Dr. Hann: Mark, who paid for that? Who sponsored that?

Dr. Appelbaum: That was actually sponsored by -- I am not allowed to say.

Dr. McGuire: It wasn't Johnson & Johnson, was it?

Dr. Appelbaum: No. It was a private organization, but it was held quite independently of the funding source.

Dr. McGuire: But you are proposing another NIMH mechanism would be sponsoring a book or a big plan?

Dr. Appelbaum: Right. But something like that instead. Every one of these consensus conferences I have participated in had certain "knee-jerk" features. You have to react to the things you hear in two days, as opposed to a very systematic, reflective process.

Dr. McGuire: Not the last one.

Dr. Robinson: I have a specific point than a rather more general one. NIMH could actually -- I assume this could be done -- could have some sort of monthly newsletter, and of the like, that would feature research judged in-house to be exemplary in introducing promising methodologies, or promising alternative research strategies. This would have once served the young don's need to be recognized for work of consequence. It would encourage submissions preliminary to proposal writing, or perhaps as part of research underway, and thereby maybe draw up some drafts, getting the reading community used to the fact that this is going on, without ordering journals to do anything of the like. Not as a journal, but as a kind of newsletter, saying, wonder if anybody thought of doing it this way.

Dr. Richters: Dan, what about using the Internet as a vehicle as well?

Dr. Robinson: Sure, sure. But when you reach my age, if the thing is not on paper, it has an unhealthy -- I mean, I look at this screen and assume that there is some rake-like being behind this.

Dr. Loftus: Print it, Dan, print it.

Dr. Robinson: My printer confuses me. It interferes with the quill that I --

Dr. Fox: He has a remotely operated system for his telephone, yet.

Dr. Robinson: You know, the worst of it is that in my late 20's I actually directed the summer workshop in computer mathematics at Columbia. We had a Burroughs 220 that was in a room twice this size. The floor had to be adjusted, and the temperature and the humidity, and its power was such that you could go to CVS pharmacy now, and for seven bucks put the thing in your pocket. So that is where I am with this. The other point I did want to make, just to keep us all awake this afternoon, was a rather seditious point. And that is, in a way I would be prepared to defend at a length you would not permit the experimental phase of this science is really supposed to be, to some extent, self-liquidating. I will tell you what I mean. We do not have large numbers of people rolling balls down incline plains. We, more or less, think that there are other things to do, but not that.

That will always be true in a science developed enough so that recourse to the laboratories for the express purpose of determining whether some general law is worth a continued investment. This is obviously not the current history of our nasty little subject. That isn't the way it goes. Thus, and with all fully earned respect for the journal editors and for APA and the like, I have never viewed it as a sign of health within a profession, that there is this profusion of experiments, journals, papers, and the like. There is a part of this story that is worrisome, that is, even if you don't want to say ---. The second thing is, is I have seen studies reported from time to time on the impact of experimental publications, on workers -- I hate that expression -- on people who are doing studies in the same field, and the impact, I gather, is not only slight, but ephemeral. Now if it really is the case that we are doing more and more of what progressively has less and less an influence on fewer, fewer people, who claim nothing less than a kind of moral or a sentimental attachment to the same set of issues, then it seems to me that the place to begin to turn things around might not be the journals at all. Bill just looked up. What shall I say about this? I would like to see the journals devoted to research of such consequence that they were published infrequently, maybe an annual research publication that is really reserved for stuff that puts a period at the end of the sentence that says, "All right. Now move on." There are lots of things you can do around this. This was pivotal, it was important, it changes the way we understand the damn thing. Nobody would have ever thought of it before. The editor actually says, note this readers. There may be years in which we do not appear. You cannot do this stuff on command. I say the seditious comment is that the increased number of journal pages may be a sign of ill health

within the scientific part.

Dr. Fox: Can I ask a question about the way science is -- if you can answer it. Is there no place for the incremental nature of work within a science?

Dr. Robinson: There is tremendous room. Physics is prospering now. It is prospering chiefly as a result of its theoretical richness. It is not prospering chiefly as a result of previous empirical work. It is the theory that generates all this stuff. We do not have theories like that in psychology. The last time -- here I speak out with bias, Hullian theory, and I give everybody an opportunity to laugh for the official 90 seconds. Anybody interested in reaction evocation potentials with the bar and the dot over it -- the oscillatory -- Hullian theory, lack although it might have been, was at least structurally the sort of theoretical thing that sent you to the laboratory, and said, if this damn thing is right, then rats in harnesses ought to do this sort of thing. I think we have almost been cowed by the tragic nature of that enterprise, but we shouldn't be cowed by it. It was an important attempt. What Hull was trying to do is develop a behavioral biophysics. He wanted to look at energy exchanges between organisms and environments, and reduce the entire affair to a kind of thermodynamics. You know, this is not going on.

That kind of theoretical risk-taking is highly productive, and it does not matter if 10 years later you decide it was a non-starter to begin with. That is an important lesson to learn too. When you get rid of that kind of risk-taking, you end up with the Journal of Tweedle Dee, published six times a year, and twelve times a year, the Journal of Tweedle Dum, because Tweedle Dum has twice the data base.

Dr. Jensen: We have a review paper in progress right now that reviews all of the "known" correlates to see how much of the research is duplicative, the same old "me too" research based on regressions, etc. So this is underway. Dr. Suppe, I think was next.

Dr. Suppe: Four points. First, with respect to the physics comparison, if you actually look at shelf space, that is shelf, the physics publications, it is mostly experimental, not theoretical.

Dr. Jensen: Oh, inevitably.

Dr. Suppe: But if you do a citation analysis on the experimental research, lifetime citations by a non-author of a paper, the expected number of citations is less than one.

Dr. Robinson: Yes, well that is the impact.

Dr. Suppe: The second comment. Going back to Nathan's comments, which I strongly support his comments about the potential and the importance of NIMH initiatives in terms of changing things, I do think you underestimate the potential the editors have, because a large number of the papers are published only after resubmission or revisions, and you had a mid-point potential there to have a very strong influence on what happens.

Dr. Fox: And you are also sending it out to so many reviewers that will also spread the word.

Dr. Suppe: Third point has to do with the question about publishing data. Several decades ago, a number of portions of social scientists attempted to provide -- they were actually -- companies started to do it commercially. They were going to be data archives, where you would get them on a fiche, and the paper would say, you can get the data for this study from such and such a place. I took in one particular area in the social sciences, several years worth of studies that had those invitations, and sent off for all the data, and I got a return of zero. Either the company was out of business, or the authors had moved to somewhere, and they were all no longer known at the home institution. The mail was generally returned, and I had a zero success rate in getting any data. If you really want to have data archived, I think it is something that is going to have to be done in one of two ways. Either NIMH is going to have to set up and fund an archiving institution that does it, and I think probably the Web or CD is the way to go.

The second thing is, I think increasingly journals are going to have to go to CDS as the form of publication,

just because it is a quarter, the cost of a hard brown copy, and it saves shelf space. Certainly, if you talk to university librarians, that is the way they see it. Now, as intermediate step would be, in each of the years, the volume of a journal you have a CD that has the data for the papers you publish in hard copy. The fourth point is, if you really want to change -- and this goes back to your comments about not being terribly computer savvy. If you really want to change things methodologically, the point where you can have the greatest influence is not with the senior researcher, but it is with the person at the dissertation stage. NIMH could fund a set of dissertation research scripts for grants that were tied to methodology.

Dr. Richters: Good point, Fred, excellent point! We also have K-award mechanisms--- the training mechanisms for career and salary support-- that can be an excellent source of leverage.

Dr. Suppe: If you here is the competition for research dissertation and research that does this, it forces the senior faculty to get involved in the methodology, because they still have to -- they have to direct the dissertation. There is always a problem of funding, and a lot of people are going to be very uncomfortable, saying well, we will not let you get this funding that you could get because I am not comfortable with the method. I think you can out a nice mind on the person through this method, and it has great potential.

Dr. Maltz: First of all, Ann Arbor and the ICPSR, Interagency Consortium for Political and Social Research, does archive an awful lot of data. In fact, the Bureau of Justice Statistics and the National Institute of Justice fund within it the National Archive of Criminal Justice Data, which is from where, by the way, somewhere I downloaded the 42 megabytes of homicide data for those figures I passed around.

Dr. Suppe: That is also widely done in the physical side. For example, earthquake data, southern California has mass store that everybody can get --

Dr. Maltz: And that point about the sorry state of most of the contributions, reminds me -- and this is just a little aside, one of Murphy's laws. "Everybody lies, but it doesn't matter, because nobody listens." I mean, who uses those studies? The third point, also with regard to Dan, with regard to experimentation, is that I think we have to remember that what we are doing is not the same thing as physics. We are shooting at a moving target, and every time we have a new finding, we have a new context in which we have to do a little bit more discovery. So perhaps the experimental side should continue.

Dr. Rubin: One of the things the Department of Education does (and they in fact fund applied research in the area of social and emotional development) is organize a meeting of the principal investigators and co-investigators once or twice each year. They are all brought together to discuss what it is that they are doing within a given theme, and there are some outsiders who are invited to talk about new machinations in the area. And then there are incentives that are applied to bring the researchers together. There are small pools of money made available, so that people who are already funded can begin to think about collaboration, and exchange of ideas, and so forth and so on. So it is a very nice model. Now, relatedly, when you bring people together it often entails the sharing of data. Therein lies a problem, because to whom do the data belong? Here is the issue of territoriality, and the reasonableness or reasons of an individual to share data with colleagues. And I do not know what the incentives are to allow the sharing of knowledge from any given lab. To whom do the data belong?

Dr. Rubin: One would hope.

Dr. Richters: Well, for all intents and purposes, the individual investigator.

Dr. Rubin: Well, I am hearing the public. I am hearing the individual. Well, you see, so then now we have the public, whomever that may be, the individual, the investigator, or the university.

Dr. Robinson: The university and the investigator, that is a technical --

Dr. Rubin: There seems to be virtually no incentive to share data, to pool data, or to fund investigators to examine, second-hand, others' data, having very reasonable and reasonable questions.

Dr. Robinson: It is a horrible state of affairs. It shouldn't even be thought of in terms of -- To whom do the data belong; who owns the fact?

Dr. Appelbaum: No, not a fact, the data.

Dr. Rubin: But think about it. I mean, there is --

Dr. Suppe: It depends on who is funding it. It depends who funded it. I mean, universities, for example, if it is patent able, will have a policy then. If university resources were involved in the development of it, it belongs to the university, not to the individual. A funding organization like NIMH could have a condition of grants that the data must be shared --

Dr. Jensen: Legally, we get across cooperative agreements, and we reel that in to say like within three, or four, or five years, we will make the data set available to the larger public, or we could do it on contract, and then we would say, "This is our data set." We buy RTI's time, but --

Dr. Richters: For the vast majority of applications, investigator initiated, we have no right legally to require that.

Dr. Fox: I would simply like to ask a question of Dan, with regard to the model of physics, as to whether or not that is the appropriate model for us, or whether, for example, biology, in all of its glory of multiple experiments, and multiplicative experiments going on at the same time, and many journals being filled with those experiments, is more analogous to the state of psychology right now. Yet they consider themselves to be a, "science," and we have discussions and meetings here to decide as to whether or not we are as well. But it seems to me that our analogy is not physics, but it is biology.

Dr. Robinson: There is an interesting history-conceptual story here. It just happens we got our kickoff from Weber and Fechner, and that mid-19th century was a time in which the proposition would have been rejected, but there is something about biology that somehow in principal removes the physics and chemistry. In fact, Helmholtz, and Ernst Brücke and Carl Ludwig, and Dubois, entered into a pact: they would accept no statement in biology unless it were reducible to physics and chemistry. Of course we have moved from that point. Now when Wundt put the discipline as --- well, he did not put it on the map --- but Wundt had *two* major projects in mind for psychology. One of them is the one we adopted. But the multivolume *Völkerpsychologie*, the anthropological psychology that dominated the later stages of his life, was predicated on the assumption that the social dimensions of life are not uncoverable in a laboratory context the way basic processes are; that you do psychophysics and that sort of thing because you are interested in processes, not persons. The analogy I sometimes use in my own teaching is that psychophysics is interested in *vision*, not *seeing*, if you get the point. Now Wundt's argument was that explanations of social phenomena have to be in terms of motives which properly understood are a species of rational purposes, et cetera. And that is to be distinguished from the essentially causal processes that underlie basic physiological events.

Well, this is a story very well known around this table. Hempel, you know wrote his book, and William Dray, and Peter Wench, and others said no, you are getting it all wrong. The Battle of Waterloo is *not* a repeatable event, it is a singular event. You take Napoleon out, it is a different story. There is no substitute. The whole Hempel-Oppenheim thing is inapplicable. This is not something that is going to be resolved here, but I would say for a division like this one, where one is concerned essentially with ineliminably, irreducibly, *social* processes, the physics just seem wrong, in every single way.

Dr. Richters: So Dan's point was *not* that our model should be physics, but unfortunately it tends to be.

Dr. Robinson: Well, what I am saying is, to the extent that the commitment is firmly to a kind of multivariate experimental approach, that is already grafted onto the physics part of the disciplines history. Well, if that was the wrong turn to begin with, then of course, everything you have grafted onto it is probably going to suffer from that defect.

Dr. Jensen: Steve and the Dr. Suppe.

Dr. Maltz: By the way, we are going down to epistemology.

Dr. Hinshaw I had three comments from the last bit of discussion. One is about whether journals and journal editors can influence things and the sense that journals are not the optimal point of intervention to effect change in the field's standards. I disagree. Journals are not applicable just at the submission phase, after the research has been completed. Rather, they set the standard for the field and therefore influence how research is conceptualized and executed (not simply how it is written up). If journals alter standards-- in terms of graphic display of data, in terms of policy about "types" of research, or in terms of pluralism regarding approaches--they can certainly shift research thinking and research practices.

Another is an NIMH issue. It is very difficult to get NIMH training grants anymore for departments of psychology, psychiatry, et cetera, and I think that this is unfortunate for at least a couple of reasons. One is that, if we want to get students who have either GRE Quantitative scores (as Geoff has suggested) or Analytic scores (which, to me, may be the more valid index) that are high, we must realize that the social sciences are competing with both business and the hard sciences. To the extent that graduate training is not an act of impoverishment, funding would help. But even more importantly, training grants might set the stage for what Bill McGuire has talked about, which is training not just in the tactics of doing a study, but in the strategies and procedures of developing a program of research. To do so optimally requires more than a single mentor with a single student; it requires a committed faculty.

Third, I wanted to reiterate a point that I made in the preliminary E-mail for this meeting, which is about NIMH funding mechanisms for experimental trials, in other words clinical trials that subserve theoretical, mechanism-linked, process-related research. John and I have had some argument about this point. I contended that experiments can elucidate mechanisms of change, if carefully designed; for example, Patterson's work has shown that a key causal mechanism regarding the development of antisocial behavior is the parental interactions with the child. John countered, however, that unlike lab experiments, clinical trials are multivariate in nature and they typically address independent variables that may be many steps removed from ultimate causal factors. I agree to the extent that even though we can perhaps determine what mechanisms are responsible for change in a clinical trial, we may not know whether these are ultimate causal processes.

Nonetheless, with forethought and with theory, clinical trials should be able to tell us a lot more than "independent variable X influences dependent variable Y and therefore has some causal status." A well-designed trial can help to elucidated mediators of treatment effect and clinical change--those processes that are key to making the treatments work (e.g., Dishion and Patterson's demonstrations of the causal nature of parent-child interactions). And, the experimental control of independent variables in a clinical trial will always yield far cleaner and more powerful tests of mediator and moderator effects than the best-executed multiple regression can ever hope to offer.

Dr. Suppe: I think we may be doing ourselves a disservice by overestimating the influence of the physical sciences, especially physics, on what in fact methodologically prevails in the areas of research we are concerned with. If you look historically at the actual statistical methods, you will find they mostly come from agriculture. The correlation/coefficient is Galten, for investigating inherentability of traits under animal breeding. The randomized experimental design is developed by Fisher from the newer studies. Galton developed student's T-tests for Guinness Brewer, for evaluating the superiority of various hopped groups, and he publishes each student because Guinness will not let him publish it because it is proprietary information, and they think it gives them an advantage in the brewing industry. Draft analysis is developed by Sewell Wright as a way of getting causal knowledge into improving the inherentability estimates over the Galten sort of approach, and he does it as senior animal ombudsman for the U.S. Department of Agriculture, and it publishes in the *Journal of Agricultural Research*. We go on to Storie Yates, and his influence on aggression analysis. Most of the techniques we area talking about here in fact are agriculture, and they are largely input/output studies.

Dr. Robinson: Wouldn't you say that is really after the fact, because all of these developments are after Fechner, after Wundt's laboratory, and after psychology had already adopted a perspective on the laboratory as *the place*. The statistical part of it is really the long footnote to what was already the accepted model.

Dr. Suppe: The statistical part, however, is the focus of the complaint, at least it is a major focus of the complaint, the methodology restricting us, being the hypothesis of testing, at least that is the rallying point that at

least -- the information I have seen for this conference, and earlier consultation on it. And you do not find the physical sciences doing this kind of statistical work. There is that heritage in the thinking of how sociology came in, but you also need to look at things like all of the demographic studies that were done on suicide, and so on, in the 19th century too; there is that other heritage. I think in terms of when we get down to the methodological issue, it is the fact that the model is agriculture, and the optimization of output is the basic model, for all the research; and the problems we are dealing with when you get down to the individual level are all concerned with suboptimal systems. Where you have multiple variables you cannot optimize all of them, and optimization research is just an inapplicable design at that level, and that I think is the real problem.

Dr. Robinson: I would just like to second what you said. I think that the broader concern is not biology, physiology, botany, agriculture, or physics, but the distinctiveness of behavioral conditions that necessarily involve dynamic exchanges, the very phenomena that we are trying to explain, which does not fit with any of the preceding and requires the special conditions, or even conceptualizing our variables.

Dr. Cairns: That is what I originally meant by saying that novelty and the dynamics of the interchanges make it impossible to reduce our explanatory variables to just intrapsyche or intraorganismic phenomenon. The very nature of the phenomena we are dealing with demands conceptualizations and analyses that can capture the dynamics of interchanges. So to the extent to which we are wedded to models that have a static basis, we in fact are in effect fighting against the developmental assumptions that seem to be guiding our explanations.

Dr. Robinson: Well, Bill, I am sorry, one more entry for the necrology. Four or five months ago I read that Peter Winch had died. Peter Winch, for those of you who are not familiar with him, was a philosopher of social science. He published a book called, *The Idea of Social Science*, which was architectonic for a number of critiques of social science, not the least of which would be William Dray's *Explanation and History*. Winch wanted to make clear in this still eminently readable book -- I strongly encourage it, if it is still in print. Winch wanted to make clear that there are certain conceptually necessary attributes, in virtue of which a phenomenon becomes social in the first instance. It is precisely what renders the phenomenon *social*, that renders it uniquely unsuitable to the kind of reductive systematic manipulations of the laboratory. Now, why is that? Because the sort of phenomenon that renders itself fit for experimental manipulation and causal modes of explanation would be a phenomenon that itself is determinatively regulated by some core set of laws. Well, to that extent, it is going to be largely liberated from the influences of personal initiative, autonomy, rational purposes, complex motives, social interactions, the individuation of the participants. It is only the Battle of Waterloo if Napoleon is on the losing side, however. Therefore, once you start parsing the sentences in virtue of which you have something called, a *social science*, you begin to realize that the right set of methods -- I hate the term -- the right methodology cannot possibly be drawn from wherever it is we use in our experimental modes of inquiry. Now, this was not a brazen or bold book; it was a clear book, and it is a point of view. It is a point of view that has very deep roots. You can recover this point of view in Aristotle. You certainly can find the point of view quite fully developed in Hegel's critique of phrenology. You can find it in R.G. Collingwood's, *The Idea of History*, on which Winch himself depended to some extent. And of course you find it in William Dray's, *Explanation in History*.

Dr. Suppe: Well, Drude was explicitly implying it in ---

Dr. Robinson: Quite. Yes, on the Wittgensteinian analysis the social domain comes about through essentially discursive practices. Everything that takes place in that domain is meaningful. Meaningfulness itself is a cultural product. And, therefore, all the rest of this scientific stuff is inapplicable. Now look, *meson atriston* --- the middle course is best. Obviously, there are studies to be done, and they are important, and we are not going to make much progress until they are done. I think we have to understand that there is a kind of authorizing metaphysics behind our enterprises, that is prepared to assert itself at key points, and say "Look, thus far may you go, and no further in the experimental end, because if you keep this up you are going to dissolve the very phenomena that constitute the social world itself."

Now, what can NIMH do about this sort of thing? Well, it has to stop embarrassing people who come out of that tradition, and want to do a precis on the findings, and raise serious stern and critical questions; questions about the method used to produce a data base. Could it *ever* possibly be applicable to a social context, on the Winch and Wittgenstein view, the social context being precisely the context that is immunized against this mode of analysis? That dialogue, which has been a very fruitful one, I hope you will agree, within philosophy itself, is a dialogue that does not take place in the behavioral and social sciences. It is viewed as a kind of -- you have to be regarded as some sort of know-nothing even to bring this up. The person that brings this up obviously thinks the

Isle of Reil *is* in the Caribbean. This is a barrier that has to be broken down. Philosophers of science have something to say to people who do this work. So do philosophers of social science and intellectual historians. There has to be some way of getting that part of the community into the debate, otherwise we are going to spend all of our time applauding our useless achievements.

Dr. Jensen: There is a very tangible example that illustrates your point. There has been some debate in the child area in the last few years about how we should assess child psychopathology. People have developed these very structured interview approaches so when an interviewer asks a question, if one admits a symptom, then the result is a branching series of follow-up questions. Well, it does not take the interviewee too long to figure out that if he/she says yes it will result in more questions and a longer interview. Yet, there is some debate that if subjects are told too much ahead of time about how this is all going to proceed so they understand the procedure, and even him/her decide on the order in which questions will be asked, that you will lose the psychometric rigor. They would rather have an invalid but reliable procedure than communicate as effectively as possible with subjects by bringing them into the information sharing and discovery process!.

Dr. Trickett: Bringing up the idea of the philosophy of social sciences reminds me of the issue of how to increasingly open NIMH's systems to people who are not already part of it. I was thinking about the review committees and how, in part, it is the same as asking Congress to deal with campaign financing reform. I am just thinking about *how* to open the doors to people who we do not know about as relevant constituencies to inform these kind of discussions who are not already part of the organization one way or the other. This seems to me an important organizational thing to take on.

Dr. Fox: I raise this as an issue and as a question, rather than knowing the answer. It seems to me that as an outsider looking in on NIMH's response to varying political considerations, that part of what drives these questions is the fact that other scientists are quite critical of the level and scientific approach of the behavioral and social sciences.

Dr. Robinson: Who, for example?

Dr. Fox: Harald Varmus, Director of the National Institutes of Health.

Dr. Robinson: And what is his specialty?

Dr. Fox: Molecular biology.

Dr. Robinson: Oh, well.

Dr. Fox: You say "oh well", but on the other hand his comment to *us* is, "oh well". And if it is not at the level of the gene, or at least at the level of the cell, then he does not really, at least as I understand it, understand its value. Now, the question is, how does NIMH -- how do we as scientists within the social sciences -- meet that particular challenge? There have been varying responses and reactions to it since it was first brought up when he first took over. And, again, I am now viewing this as an outsider and watching it. Some have defended our mission as "scientific" in the same way that physics, biology, or chemistry are scientific. And some of them --- I think the more practical ones and the more successful ones --- are, as was mentioned earlier this morning, to show the tactical progress that we as a discipline have made, in particular, in certain areas. That response is obviously one that is going to govern then the kinds of questions that we ask, and the way that we go about asking those questions.

Dr. Robinson: Well, I will never have an opportunity to talk to Dr. Varmus about this, but I would say there would be no interest at all in the molecular biology of the gene if we did not have at the gross phenotypic level and at the level of evolutionary theory some reason to be interested in things like genes. Just as there would not be any interest in brain physiology if there were not some reliable relationship between events in the brain and things like learning and perception. We do not have a mind/spleen problem! I think what I might suggest to Dr. Varmus is this: There does not seem to be much progress made in completing the bridge from the molecular biology of the gene to contemporary evolutionary theory at the level of actual aggregates of animals in complex settings facing extraordinary selection pressures. That bridge one day presumably will be built. Therefore, he should not expect

there to be much bridge building between, let's say, the molecular biology of the gene and social unrest in troubled communities, but that if there were not social unrest in troubled communities occupied by bona fide hominids who have needs, desires, and the like, then the research that he did, that probably some agency paid for, would have been just another leaning on your shovel, WPA project.

So I submit, with all due respect, he really ought to just sit back and thank his lucky stars that there really is a relationship between genetics and the rest of this.

Dr. McGuire: I think that most of us here would hope and expect that today's meeting, concentrating on how behavioral science could be improved, would enhanced the perceived value of behavioral science. But we should recognize that there is some risk that accentuating the negative at this meeting could backfire and make the field look bad. The report on today's meeting should be sensitive to this danger.

Future Planning

Dr. Robinson: Well, but we have taken criticism as seriously, enough to constitute a committee and address these matters in a sober, serious, and disinterested way.

Dr. Cairns: Well, I am not exactly disinterested, or uninterested. The issue in terms of behaviour is self-evident. For children and adolescents the major threats of health have to do with behavioral events such as violence, car accidents, suicide, and risk-taking. This does not divorce us from those earlier comments about psychobiology, but it does presuppose we are going to do the kind of innovative studies and cross-validated studies that link behavior and biology. I guess my concern is the extent to which we continue to force these kind of issues into a methodological mold that was ill-suited for the task. That is why I made those comments about some of the distinctiveness of developmental dynamics and emerging qualities. These characteristics do not fit most regression models in terms of analysis of variance or reductionism. On that score it would seem like we are currently making significant headway with longitudinal methodologies, person-oriented strategies, and sophisticated measurement strategies. There is now a much broader consensus among leading researchers on these matters than I have seen before.

Dr. Richters: A behaviorally silent majority.

Dr. Cairns: Yes, and some of them are becoming much less silent. They are beginning to speak up, and this group can capitalize upon that sentiment. It is not just a sourness about existing procedures or reviews. More productively, it points to these alternative strategies of disaggregation, of scaling down by a prodigal analysis to individual levels, and then coming back up to group generalizations. There are some major changes that are underway, and it looks like this group is beginning to crystallize at least some of these ideas.

Dr. Jensen: I cannot imagine a better comment to -- it comes back to the notion of whether it is an ebb tide or a sea change, but I personally feel we are talking about an increasingly *qualitative* change. We thought it might be worthwhile just going around, and giving everybody 30 seconds for maybe a final parting shot, after which both Kimberly, John, and I, will give a few summary comments about where we are going to head and some of the future activities we will undertake. So we will be finished by 4:00 pm. But Steve, since we started with you last time, why don't we start with Nathan. And his final parting shot, what will happen?

Dr. Fox: Well, let me make two comments, just very brief comments, because I only have 30 seconds. The first is that I think that we need to understand the difference between the bad use of current statistics and the use of new approaches to understanding our data. And the thing that is interesting to me and most important to me is the attempt to understand data through graphical analysis and through the display of data. I think that is critically important. I think that displaying relations, rather than presenting a correlation is critical. The second thing I would say is a plug for extreme groups analysis. I think that psychology has been adverse to the publication or the presentation of data that looks at extreme groups, but I think it was Dr. Suppe who said that perhaps at the tails you are going to get some of the interesting data, and I would say that perhaps, in terms of the homogeneity/heterogeneity arguments, that looking at the extreme groups might be a profitable way of approaching some of the issues that have been raised today.

Dr. Appelbaum: The El Nino was actually a wonderful example. Being a San Diegan and watching most of the models which were coming out of the *Scripps Oceanographic Institute* what you are convinced of is that it will not be an average winter. On the other hand, we have half the models that say we are going to be in drought, and half that say we are going to flood out.

Dr. Robinson: So compute an average and nothing will happen.

Dr. Appelbaum: Right. I think it is difficult to speculate where things will go. I think there are many issues which are interesting. I am really rather impressed with the fact that collectively we seem to have the answers to almost all of them. The problem is we don't know how to get the rest of the scientific community to "play right". I am sure that another group could meet and say the same thing about this group. They are very difficult issues. It is hard, perhaps wisely, to have a system that can shift too fast, too many times. I think that models where we think of ways to allow new paradigms to be tested out on good problems is helpful. If you want to kill one of these paradigms, all you need to do is put it to the wrong problem or a problem that is just beyond that level of solution. So I would caution that if you want this process to succeed, to first cut the problem down to an intellectually reasonable size; don't try to solve all the problems of antisocial behavior, and use that goal as the way of testing a different paradigm, because you can assure the failure of the paradigm.

So somewhere I think there needs to be some paring down to a reality level, and then trying multiple approaches. If you want to take the Kuhn model seriously you always have to do the same thing, using the paradigm that is currently there, because the final test is whether or not the new paradigm out performs the old paradigm, and if you do not have the old paradigm working on that problem, then you cannot do that.

Dr. McGuire: Well, I hope and expect that the record of today's meeting will show 50 good suggestions (some specific and others general) and maybe 10 to 15 not so good suggestions for substantive and methods improvements that would move the behavioral sciences forward. Our discussion may also have mentioned 15 or 20 mechanisms that NIMH could use to promote these improvements. Later these suggestions could be arranged into a 50 x 15 matrix, visiting the cells of which could be a creative guided tour. I doubt if the 50 suggested improvements will now or ever meld into one big paradigm shift but it might coalesce into one or two big clusters and maybe four smaller clusters and a dozen or so singletons, all deserving of some attention.

Dr. Appelbaum: Do you think it might be reasonable to break the order for the guy that is standing there with his hands on the suitcase, worrying about when his flight leave?

Dr. Cairns: At first blush it sounds like we are adding more complexity to the task. But I think it is just the opposite. If you shift analysis to the level of the individual integrative organism you are not going to look only at what is inside the person, but you are going to be looking more carefully at those forces outside. This requires the field to develop and employ levels of analysis beyond the individual, the dyad, and the social network. What would be an impossible task if you use the usual paradigm of studying single variables in isolation becomes eminently reasonable if you integrate them in the individual. The role of social networks and relationships has not been a major theme of this workshop, but they seem to be essential to any coherent account of antisocial patterns. One of the happy advantages in reconceptualizing our dominant methodology is that it can become more open to social influences. We have not talked about that stuff at all, but again, I think that is one of the happy advantages of making a shift.

Dr. Maltz: First of all, I am very pleased to have met you all intellectually and personally. I have gotten a lot out of this, and I hope that I have contributed. Second, I'd like to underscore what Bob just said about complexity. One of Edward Tufte's points in one of his books on graphics is *to clarify, add detail*, which, essentially, was what I was trying to do in the figures that I displayed. If we look at the real detail, if we go down to the nitty gritty, we can see a hell of a lot of more. We can see more patterns than we can by squinting, and to me focusing on the mean is essentially squinting.

The other point I want to make has to do with training, and quite frankly, the use of computers. I look upon computers as being used in two different ways: One as an autopilot; and the other as power steering. I think that we have been teaching our students too much to sit back and leave the driving to us. You just throw the data in there, and it will come out, will spit out a couple of *p*-values that are low enough that you can do something with--

instead of you doing the driving and *you* saying, "I want to look at this. I want to see what this relationship is." We do not teach them that enough. This again is the logic of discovery, and again as Geoff and Mark will probably hit on as many times as we can, graphics, I think, is the way to go.

Dr. Richters: Plastics.

Dr. Rubin: I would like to reinforce notions about additional fellowship support for graduate students. I think that is very important. I think there are probably countless decent data sets out there that exist, to which very bad questions have been addressed. I think that we owe it to the young graduate students who are coming into the field to address better questions with these probably decent data.

Dr. Richters: These pre-docs and post-docs.

Dr. Rubin: Both, right. Canada, for example, has a wonderful system of funding graduate students at many different levels. I was telling Della that the average research grant for a Canadian social scientist is in the level of impoverishment, relative to American social scientists, but we do not have to fund our graduate students, they are funded. They (the students) apply for funding from the federal government (e.g., from the Ontario Mental Health Foundation, which would be the provincial "relative" of NIMH). With decent ideas they get to track data that may already exist, or new data really nicely funded, and I think that is very important.

The other thing I would like to say is that much of the discussion around the table has been addressed to dealing with issues pertaining to individuals. >From the writings of Robert Hinde we learn that there are several levels of social complexity-- individuals, the interactions between individuals, the relationships that are the result of these interactions between individuals, and then the groups within which all of this occurs. So rather than dealing with main effects from individual to misconduct or antisocial behavior, or from interactions to the prediction of antisocial behavior or relationships, we ought to be marrying all of these in some conceptually rich way. I listed a whole set of questions that I am sure people have wonderful data for, that address each of these levels in a mixed fashion. I will share them with John a little later. You could probably fund a slew of graduate students to do all your work for you rather cheaply.

Dr. Jensen: Dr. Suppe?

Dr. Suppe: I think I am going to make my parting comments something systematic with respect to epistemology and epistemological issues, because that is something that we have talked about, at least systematically. I think the question you have to ask is, at what level will an epistemological analysis or asking epistemological questions are going to make a difference in how the science is done. When I ask that question, my answer is -- the epistemological question is going to be like the following: What would happen if we used multiple indicators in collecting our data in such a way that the mathematics of overdetermined systems would kick in, and then we exploited that mathematical power in our analysis. Would it make a difference in the reality of our model? If I look at a number of branches of the physical sciences, particle physics, geophysics, and take a couple, you will find that if you get an actual fit to your data, the models almost certainly is approximately correct, even though the data underdetermined the model. Why is this? Well, in geophysics it has to do with the measurement properties themselves are such, that full representation of a finite set of measures requires an interdimensional Hilbert space, and you have to fit the full properties of the interdimensional Hilbert space in order to have a fit to your data. If you go into particle physics you will find all the experimental data are in the form of homomorphic functions. Those are real valid functions that have to hold in the complex space, their underdetermined the real space, but in the complex space, but additional structure gives you a semi-unique fit and it is down the line like this.

The most successful branches of science in modeling are ones where you have very rich mathematical structures to the models and the demand is that -- I mean to the measures, and the demand is that the fit not be to the scalar value, but to the full mathematic properties. Now, do we have potential for developing that kind of measures? And, if so, would we thereby then get models that are not just fitting central tendency, but are actually fitting the data set. Similarly, the questions that Nathan was talking about on looking at the extreme cases, looking at the outliers. One reason why you want to look at the outliers is, in aggression analysis they drive the analysis. Do you want it in there or not? What if we us two key lines instead of aggression lines? These seem to me the useful kinds of epistemological questions that will impact on how the science is done. I would like to suggest that I

do not think there is much in the philosophy of science, or the epistemology literature that will be of any value at that level.

Dr. Hinshaw: I will make four points in closing. One is about contextualization, those contextual factors that several people have mentioned in their concluding comments. The very nature of the behavior we are discussing is highly contextualized, and paying attention to culture and ethnicity, as well as to gender (more than just giving these factors lip service at the end of a grant) is going to be central for NIMH in future years. In other words, we need powerful investigations of the roles of these moderating factors and of mechanisms and processes that diverge across socioeconomic or ethnic groups or between males and females. Second, training of students, post-docs, and all of ourselves in hypothesis generation, rather than just hypothesis confirmation and testing, a point that Bill McGuire has made in many of his writings, is crucial. Textbooks fail to even mention this phase of the research process. Third, I think an issue we largely ducked--that is, we took it up implicitly rather than explicitly--is one that John Richters challenged us about, which pertains to the nature of our evidentiary basis in the field. If our acid tests must transcend the weak ones of rejection of null hypotheses, what sorts of persuasion, what sorts of looking at graphical data, what sorts of converging- operations arguments will we believe?

Finally, I think that the task in front of us is really Sisyphean, if I can coin an adjective. The topic of interest is a form of social deviance, and it is impossible to imagine, given both our evolutionary past and our current social structures, that we could ever eliminate antisocial behavior. There is going to be, in a Durkheim's sense, deviance in any society. We can promote seemingly straightforward methods, like gun control, to reduce some of the devastating effects of certain forms of modern antisocial behavior; but whether we can fully understand the complexity of the topic and fully prevent it is a near impossibility. More reasonable and proximal goals can and will occupy us over the next decades.

Dr. Loftus: Okay, as I said, my comments are going to be probably a little prosaic and random, and I will be echoing things that other people have said. The thing that struck me the most is the emphasis that has come up in various guises on understanding versus prediction. I wish that Reuven Dar could have been here, because he brought up most explicitly. I think that comment is important along with all it implies, which includes things that we have been talking about, such as using graphical techniques, eschewing null hypothesis testing, playing with our data as much as possible, and so on.

The second thing is that I have benefitted a lot from this meeting and I hope we will have additional ones, both perhaps meetings with us, and meetings that would include specific subareas within psychology to inquire more specifically about how the issues we have raised here apply to the different diverse areas within our fields. Third, somehow or another fostering more mathematical sophistication within students of psychology, as I said in my email messages, would be good. Finally, I really like the idea of some sort of yearly or semi-yearly journal, in which there would be maybe even a competition for the best articles across all of psychology. This would allow people in one area to see what is happening in other areas, which would perhaps promote the kind of cross-fertilization that is becoming increasingly absent at the moment. I would also like to thank John Richters for initiating the whole thing; I think it has been great.

Dr. Robinson: On the assumption that we will not be allowed to abandon the paradigm, or even retreat from it, I will just leave you with this joke making the rounds in Dublin. Ireland, you may know, has a holy hour -- from 2:30 to 3:30 the bars are closed -- and at 3:00 o'clock an Englishman goes in and orders up something, and the bartender says to him, he says, "Well, apparently you don't know about the holy hour, and the bar will be closed till 3:30. But would you like a drink while you wait?"

Dr. Trickett: I know a statistician who sort of you know does shoots the bullseye four inches to the right, shoots the bullseye four inches to the left and said, "Bullseye."

Dr. Appelbaum: That is a variation on the statistician being a person whose head is in the oven and foot is in the refrigerator and says, "On average I feel --"

Dr. Trickett: Exactly. Just two or three kinds of different agendas. One is in terms of the theory development agenda. I think the two big emphases that came over today for me, were the concepts of discovery and diversity. I think the idea of looking at data as heuristics, and trying to develop methods that focus on understanding rather than prediction are very, very helpful. With respect to the diversity, I think methodological pluralism seems

reasonably important to pursue, and population diversity, and levels of analysis diversity also. I mean, it seems like all those different things are in the hopper. I have been struck a number of times about how of behavioral science is being seen more and more as a constraint, and some kind of an analysis of how NIMH participates in that, how we in journals participate in that seems quite important to do. I was thinking about the difference between looking at our field as studying people versus variables, as another kind of way of thinking about things, is important. Finally, I do think there is an increasing emphasis on specialization in a whole bunch of ways. I think Mark Appelbaum emphasized this point earlier. I think the idea of how to foster some kind of structures that can create sort of general overviews of things, such as you are doing here, is a very useful way to go. If you cannot create generalists as people, maybe we can at least create some structures that bring specialists together and a bunch of pointy-headed people to develop a well-rounded idea of something or other.

Dr. Mitnick: I am just intrigued by the variety of comments, and how my colleagues are going to put them all together.

Dr. Jensen: Well, we better turn next to Ellen, because she is the one with all the money.

Dr. Stover: I basically probably ought to begin by thanking everyone. I am quite struck with the progress actually that was made from this morning. These are always challenging sorts of endeavors, and to see whether they can come together. I being a very practical type like to see very concrete solutions emerging around 1:30, 2:00 o'clock in a day meeting. I think you gave us some very, very, very good suggestions. There are many which Len and I could tell you, we either have in the hopper, or on the burner, or somewhere within the Institute that we can draw from. The other thing I wanted to share with you is, some of you may have participated in this process, and maybe you did not -- but Norman Anderson -- a different Norman Anderson, I think -- do you know what I am talking about? That process of sort of identifying who is with you - I mean, maybe you could describe it better than I could, but to me, that this kind of process -- another step in that might be to move to that level of trying to identify the forces. They range from political to within NIMH.

Dr. Appelbaum: They were people, they were structural, they were --

Dr. Stover: Right. To try to get at how we would make use of our product and our effort, and who ought to be, not only at the table, but for whom are we generating all this. I found that an extremely illuminating couple of days.

Dr. Appelbaum: And somewhat mind-boggling.

Dr. Stover: Very. But there were parts of today that reminded me of that.

Dr. Appelbaum: The mind-boggling part.

Dr. Stover: The need for the structure, and when you were talking about trying to define the clusters, and what actually makes sense in terms of the scaling process, it might lend itself. So you could think about that.

Dr. McGuire: If you send us report on this, could you have an attachment of the chart that Norman Anderson had.

Dr. Stover: Yes. It was like a year-long process, which involved major players in the field, and major associations, and then people from the government. So again, another thought.

Dr. Robinson: There is a lusty history behind science bashing, and I do not think we should be put off by it. It is a form of respect really. You know, when Swift wrote *Gulliver's Travels*, it was a broadside against the Royal Society. And in fact the Island of Laputa, which is the floating island connected to earth by strings, is run chiefly by a scientific community that has promised the king that they have been able to extract sunlight from green plants, and, if the funding keeps up, they might be able to get enough of it to sell it off cheaply. So I would not worry so much about that. At our very, very worst we can get Swift to do something very good.

Dr. McGuire: I thought that was an island in the Caribbean.

Dr. Jensen: Any comments from the NIMH staff?

Dr. Nottelmann: Well, I am very happy to have been here, and I think it is tremendous what this group has come up with in terms of laying out all the things that need to be done.

Dr. Hann: Oh, I second Edith's comment! I think some very provocative ideas have been put out today, and I think that some of them, the ones that are in various stages, different pockets for different problems. It will be interesting to see how they might be applicable to an area.

Dr. Maser: Well, I was glad to be invited to this intellectual challenge, and entertained really in many ways by it all. The couple of things that I would comment on: One is, if you want to change how the NIMH operates, you are doing it. If you want to change how publication boards operate, the way to do it is not offer them a bribe, I think, but to go to them and say how would you change things in order for this to happen. And, believe me, they will come up with the same suggestions you did, but it came from them. As far as the group versus the idiographic data convergence, I think it would depend -- if you could figure out the characteristics of the group very finely, and then find two people who had those same characteristics, you could then expect the data to converge. If they are the outliers of the group, you are not going to find that convergence, but I think that is a nifty idea. Finally, just like review committees, the proliferation of journals is subject to the pugilism. We have met the enemy, and it is us. It is *us* that has to tell the library not to subscribe to some journal.

Dr. Appelbaum : Except for mine!

Dr. Maser: We have to refuse to serve on those editorial review boards of second-rate journals, and we have to not submit articles to those journals. We have to insist on not rushing into publication the moment we have some piece of data. So those are my comments.

Dr. Jensen: Kimberly?

Dr. Hoagwood: Yes. I would like to second Ellen's comment, and thank you very much for coming. I think the recommendations you gave are concrete, and I can assure you that there will be some new initiatives forthcoming very soon. When we first started out organizing this workshop, we were using the Kuhnian notion of a paradigm shift as a central framework, but I think after hearing the discussion today that David Hull's notion of science as a selection process is preferable. Hull in *Science as a Process* in 1988 talked about scientific ideas as being part of a conceptual lineage, much along the same lines that we have biological lineages. Ideas themselves resemble or differ from each other as genes do, and they evolve. This is a notion that is much more evolutionary and appropriate I think for conceptualizing the changes necessary for advancing science on antisocial behavior, because it includes an interactional and transactional dimension. I think we have converged on some conceptually coherent ways of thinking about the problems of theory generation, and have made some practical recommendations as well. So I want to thank everybody very much for having participated in this workshop, and I hope it will not be the last.

Dr. Jensen: Well, before John sends us on our way, let me make one or two final comments. I guess the thing I have been impressed by--- and actually Ed Trickett summarized it very nicely-- we need multiple ways to triangulate inferentially, if you will, to converge upon "knowing well enough". If our measures of the phenomena are such moving targets, we are going to need converging operations, much as we do with the CAT scan that maps in the three dimensional space to find out what is inside the brain. As with CAP scans, with converging operations we can get there.

So whether through methodological pluralism or by demonstrating similar effects across diverse populations, or using at different processes, or through different levels of analysis, we must bring scientists together who present different data, so that converging operations can happen around the table. This is what we at NIMH are going to have to do. As we have been talking about this within the Developmental Psychopathology Research Branch, we need a strategic rethinking out where of where the gaps and the holes are, so we can better identify and target critical knowledge gaps, so that we can demonstrate to the public-- and on behalf of the public-- that what we are doing is in the long run serving the public good. I thank you all for being here.

Dr. Richters: We were talking last night about Bill McGuire's concern that this Roundtable might play into the hands of those who would like to close down the social and behavioral sciences. And Dan said *My God, if you look around the room at how seldom we cite each other, and at how many of us have never met, we certainly can't be a cabal*. I think we left the problems in better shape than we approached them this morning, which is all we expected from this meeting. It was a form of venture intellectualism that has already paid dividends, an investment that will continue to yield into the foreseeable future. After everyone has had a chance to look at a transcript of our dialogue, we can talk about the possibility posting it in some edited form on the Internet. There are a number of alternatives available to us. We will also try to draft many of you on an ongoing basis for follow-up efforts, realizing that some of you will sort yourselves out of the process because of other commitments and others will be recruited as the foci and requirements of our efforts change. I will keep you all on this little e-mail list that I have been using for the past couple of months.

Dr. Jensen : The hit list!

Dr. Richters: The hit list, that's right. You all have that backspace/delete key on your computers in some form, but I appreciate that so many of you have chosen not to use it so far.

Dr. McGuire: Could I ask you to consider doing the report of this meeting as a two-step procedure? First, you would send to each of us only his or her own remarks, giving us a chance to say more coherently what we intended to say. Then the various sets of revised remarks would be put back together as a more coherent whole, the way the Congressional Record is done.

Dr. Richters: I think that makes a lot of sense.

Dr. McGuire: If you have time.

Dr. Richters: Well, we will make time. I think for the group of people we brought together, their varying interests, perspectives, and histories, it is remarkable and indeed encouraging that we had as much consensus as we did about the nature of the problems --- the outlines of the beast --- and what we might do to track it better and try to bring about some change. And for that I cannot thank you enough. You honored the issues, you honored the founders of our field(s), and you made an important contribution to stimulating and shaping a long-overdue constructive dialogue about reform and retooling. So, our people will be in touch with your people (*Whereupon, conference was adjourned at 4:30 p.m.*)

Postscript

The National Institute of Mental Health encourages and welcomes comments concerning the issues and recommendations discussed during the Roundtable meeting. Please forward comments to jrichter@mail.nih.gov

Selected References

Bhaskar, R. (1997). *A realist theory of science* (2nd ed). London: Verso Classics.

Cairns, R. B. (1986). Phenomena lost: Issues in the study of development. In J. Valsiner (Ed.), *The individual subject and scientific psychology* (pp. 97-112). New York: Plenum Press.

Cartwright, N. (1997). What is a causal structure?. In V. R. McKim & S. P. Turner (Eds.) *Causality in crisis?: Statistical methods and the search for causal knowledge in the social sciences* (pp. 343-357). Indiana: University

of Notre Dame Press

Dar, R. (in press). Null hypothesis tests and theory corroboration: Defending NHSTP out of context. *Behavioral and Brain Sciences*.

Dar, R., Serlin, R. C., & Omer, H. (1994). Misuse of statistical tests in three decades of psychotherapy research. *Journal of Consulting and Clinical Psychology, 62*, 75-82.

Dar, R. (1990). Theory corroboration and football: Measuring progress. *Psychological Inquiry, 1*, 149-151.

Dar, R. (1987). Another look at Meehl, Lakatos, and the scientific practices of psychologists. *American Psychologist, 42*, 145-151.

Freedman, D. A. (1997). From association to causation via regression. In V. R. McKim & S. P. Turner (Eds.) *Causality in crisis?: Statistical methods and the search for causal knowledge in the social sciences* (pp. 113-161). Indiana: University of Notre Dame Press.

Freedman, D. A. (1997). Rejoinder to Spirtes and Scheines. In V.R. McKim & S. P. Turner (Eds.) *Causality in crisis?: Statistical methods and the search for causal knowledge in the social sciences* (pp. 177-182). Indiana: University of Notre Dame Press.

Hinshaw, S. & Park, T. (in press). Research problems and issues: Toward a more definitive science of disruptive behavior disorders. In H. C. Quay & A. E. Hogan (Eds.), *Handbook of disruptive behavior disorders*. New York: Plenum Press.

Kagan, J. (1997). Conceptualizing psychopathology: The importance of developmental profiles. *Development and Psychopathology, 9*, 321-334.

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

Kuhn, T. S. (1977). The essential tension: Tradition and innovation in scientific research. In T. S. Kuhn (Ed.), *The essential tension: Selected readings in scientific tradition and change*. Chicago, IL.: University of Chicago Press.

Loftus, G. R. (in press). Why psychology will be a much better science when we change the way we analyze data. *Current Directions in Psychological Science*.

Loftus, G. R. & Masson, M. E. J. (1994) Using confidence intervals in within-subjects designs. *Psychonomic Bulletin & Review, 1*, 476-490

Loftus, G. R. (1993). Editorial Comment. *Memory & Cognition, 21*, 1-3.

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

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.

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

Maltz, M. D. (1995). Criminality in space and time: Life course analysis and the micro-ecology of crime. In J. E. Eck & D. Weisburd (Eds.), *Crime and place* (pp. 315-347). New York: Criminal Justice Press.

Maltz, M. D. (1994). Deviating from the mean: The declining significance of significance. *Journal of Research in*

Crime and Delinquency, 31, 434-436.

Manicas, P. T., & Secord, P. F. (1983). Implications for psychology of the new philosophy of science. *American Psychologist*, 4, 399-413.

McGuire, W.J. (1997). Creative hypothesis generating psychology: Some useful heuristics. *Annual Review of Psychology*, 48, 1-30.

McGuire, W.J. (1989). A perspectivist approach to the strategic planning of programmatic scientific research. In B. Gholson, W. R. Shadish, R. A. Neimeyer, & A. C. Houts (Eds.), *Psychology of science: Contributions to metascience* (pp. 214-245). New York: Cambridge University Press.

McGuire, W. J. (1983). A contextualist theory of knowledge: Its implications for innovation and reform in psychological research. *Advances in Experimental Social Psychology*, 16, 2-47.

McKim, V. (1997). Introduction. In V. R. McKim & S. P. Turner (Eds.) *Causality in crisis?: Statistical methods and the search for causal knowledge in the social sciences* (pp. 1-19). Indiana: University of Notre Dame Press.

Meehl, P. E. (1990). Appraising and amending theories: The strategy of Lakatosian defense and two principles that warrant it. *Psychological Inquiry*, 1, 108-141.

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, 806-834.

Overton, W. E., & Horowitz, H. A. (1991). Developmental psychopathology: Integrations and differentiations (pp. 1-42). In D. Cicchetti & S. Toth (Eds.), *Rochester Symposium on Developmental Psychopathology, Vol. 3: Models and integrations*.

Richters, J. E. (1997). The Hubble hypothesis and the developmentalist's dilemma. *Development and Psychopathology*, 9, 193-229.

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

Robinson, D. N. (1995). The logic of reductionistic models. *New Ideas in Psychology*, 13, 1-8.

Robinson, D. N. (1993). Is there a Jamesian tradition in psychology? *American Psychologist*, 48, 638-643.

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

Turner, S. (1997). "Net Effects": A short history. In V. R. McKim & S. P. Turner (Eds.) *Causality in crisis?: Statistical methods and the search for causal knowledge in the social sciences* (pp. 23-45). Indiana: University of Notre Dame Press.