Family Interaction in Clinical and Nonclinical Families: An Appraisal of the Research Domain

by

Frances B. Ravinsky

A thesis
presented to the University of Manitoba
in partial fulfillment of the
requirements for the degree of
Master of Arts
in
Psychology

Winnipeg, Manitoba

FAMILY INTERACTION IN CLINICAL AND NONCLINICAL FAMILIES: AN APPRAISAL OF THE RESEARCH DOMAIN

BY

FRANCES B. RAVINSKY

A thesis submitted to the Faculty of Graduate Studies of the University of Manitoba in partial fulfillment of the requirements of the degree of

MASTER OF ARTS

ACKNOWLEDGEMENTS

There are a number of people that I would like to thank for their contribution to this thesis.

I would like to first extend my thanks to my advisor, Dr. John Schallow, for persisting with me through the many muddles and frustrations of the task, and for sharing his thoughts and insights in a manner that helped clarify issues both large and small.

Thanks also are extended to my committee members, Drs. Barry Trute and Mary Warmbrod for their careful readings of the manuscript and for their thoughtful suggestions.

I would also like to thank all those who assisted me in conducting the research; in particular Jill Duncan, for her methodical bibliographic work and Hans Beihl, for his diligent efforts as second coder. Particularly warm thanks to Laurel Lee Mayo for contributing her skill and for acting in solidarity.

Finally, I would like to thank my jewel of a little daughter, Sarah Brenner Miller, for helping me keep perspective during the project's grimer moments.

Financial assistance came from Manitoba Mental Health Foundation Grant 387-1665-22.

ABSTRACT

Since its emergence nearly three decades ago, family systems theory has attracted growing numbers of theoreticians and practitioners who champion the validity and effectiveness of the approach. While research efforts directed at validating the principles of systems theory have been evident from the start, appraisals of the relevant literature have not been favourable. Such appraisals are, however, easily discounted. Whether highly pessimistic or tentatively optimistic in tone, they have generally confronted the reader with a mass of contradictory information that is difficult, if not impossible to process. Such confusion can be traced to the use of integrative strategies that are not appropriate for the literature of interest.

The purpose of the present research was to correct for past confusion, by integrating the family interaction literature in a clear and orderly fashion. To meet this objective, I initiated an alternative review strategy that I hoped would be more appropriate, namely, meta-analysis (Glass, 1976). In light of the breadth of this task, a decision was made to restrict this thesis to a number of 'preparatory' projects. These were designed to serve three interrelated functions. The thesis's first function was to prepare the groundwork for further meta-analytic steps by 1) selecting a set of 35 study features potentially predictive of study outcome; and 2) by creating a set of coding conventions that allows for the measurement and quantification of these features. According to a number of specified criteria, the selection of study features proved to be sound. In addition, the coding conventions were applied to the family process literature with a reasonable level of reliability. The second function of the research was a descriptive one, involving the categorization and summarization of methodological and procedural features of interest. What is suggested by the study's data is a literature decidedly uniform in its choice of methodological and procedural practices, though less so in choice of research foci. Thus, family process researchers have been decidedly rigid in their use of methodology and paradigm -- inappropriately so in light of the complex nature of the subject matter. A final objective of the present research was to evaluate the preparedness of the family process literature for integration, and more specifically, to evaluate the appropriateness of meta-analysis as the integrative strategy of choice. Concerns regarding issues

of conceptual and statistical adequacy emerged out of the evaluation.

CONTENTS

Appendix page

THE FAMILY INTERACTION LITERATURE: THE STATE OF THE ART

Since its emergence nearly three decades ago, family systems theory has attracted growing numbers of theoreticians and practitioners who champion the validity and effectiveness of the approach. Research efforts directed towards validating this enthusiasm have been evident from the start. Two major bodies of literature have emerged from these efforts; 1) the family interaction literature (also referred to as family process) that compares patterns of interaction and styles of communication in clinical families with those of nonclinical families and 2) the family outcome literature that compares the efficacy of family therapy with that of other clinical approaches. The present thesis is concerned with the former.

Evaluations of the family process literature have not been favorable. With the exception of Doane (1978a), and to a more moderate degree, Liem (1980), reviewers of the process literature have called either for a sharp dampening of enthusiasm (Frank, 1965; Goldstein & Rodnick,1975; Jacob, 1975; Jacob & Grounds, 1978) or, at most, a cautious support of systems principles (Fontana, 1966; Haley, 1972; Riskin & Faunce, 1972; Sanua, 1961; Wynne, 1981). The sharp decline in interactional research during recent years, noted by

Goldstein and Rodnick in 1975 and again by Liem in 1980, likely reflects such negative appraisal. Abandonment of the research domain may, however, be premature, insofar as past appraisals are open to serious criticism. Whether highly pessimistic or tenatively optimistic in tone, they have generally confronted the reader with a mass of contradictory information that is difficult, if not impossible to process. Such confusion can be traced to the use of integrative strategies that are not appropriate for the literature of interest.

The purpose of the present research was to correct for past confusion, by integrating the family process literature in a clear and orderly fashion. To meet this objective, I initiated an alternative review strategy that I hoped would be appropriate, namely, meta_analysis (Glass, 1976).

The nature of the inadequacies of past reviews will be clarified in the assessment to follow, as well as in the following chapters on integrative techniques. Before proceeding in this direction, my intent is to briefly present some concepts and principles basic to the family systems perspective.[1] This will be followed by a discussion of the difficulties involved in conducting primary research

[1] Since most family theorists and researchers refer at some point to the basic systems principles, it would be exceedingly cumbersome to include all such references in this section of the thesis. I have, therefore, been decidedly spartan in my referencing style here, and have mainly selected one or two examples of what I consider to be clear presentations of the subject matter.

on family interactions and of the implications of these difficulties for the integration of findings.[2]

Family Systems Principles

The family systems approach contains a broad set of assumptions about human behavior in general and psychopathology in particular (Watzlawick, 1977). A number of perspectives have contributed to the general family interaction model. The earliest work emerged out of the clinical tradition, with research on families of schizophrenics conducted by Bateson, Lidz, Wynne and Bowen, among others, providing the initial thrust. In their search for theoretical foundations, these investigators utilized the sociological and social psychological formulations of

---

[2] A number of points about the terminology used in this thesis are in order. First, the terms "family systems", "family interaction", and "family process" are by and large used interchangeably to refer to the family process literature. Second, the term "families" preceded by a diagnostic label, is used as a shorthand for "families that contain a member who has been so diagnosed". For instance, the term "schizophrenic families" refers to families that contain a member(s) who has received a diagnosis of schizophrenia. Conversely, the term "nonclinical families" is used to refer to families in which no member has received a clinical diagnosis and/or been in contact with the mental health profession for a designated time period. The fact that nonclinical families are defined in terms of what they are not rather than in terms of what they are underscores the bias in favor of pathology that permeates psychology, as well as the absence of attention to the attributes of health (Rabkin, 1965; Seeman, 1983). Since family interaction researchers have as yet, failed to select their nonclinical families according to health-related criteria, the terms "normal" and "healthy" families are avoided in this thesis.

communication theory, information theory and role theory, along with psychiatric principles and general system theory principles (Haley, 1967; Olson, 1970; Riskin & Faunce, 1972). Once articulated, the general model divided into a number of branches sufficiently diverse to warrant the proposal of a variety of schemes for classifying family theory. Most of the suggested schemes differentiate branches of the general model in reference to aspects of therapeutic practice (Okun & Rappaport, 1980, p. 68); specifically, the unit of treatment (Olson, 1970), the role of the therapist (Feber, Mendelsohn & Napier, 1973), and the goal of therapy (Group for the Advancement of Psychiatry, 1970). Underlying all such schemes is a concern with the degree to which theorists consider intrapsychic and interactional paradigms to be compatible. There are two other classification schemes of which I am aware that ignore the issue of paradigmatic compatibility and instead base distinctions on substantive emphases, those of Haley (1967) and of Okun and Rappaport (1980).

Despite large theoretical and practical differences, all of the branches of the general family systems model, regardless of how classified, are linked by their emphasis on family life as the prime determinant of psychopathology. As well, they ascribe to a common set of underlying principles derived from general system theory (Bertalanfly, 1956). The first of these principles concerns the nature of

causality. Consistent with current epistemology, all systems thinkers reject the linear model of causality in favour of the cybernetic model (Watzlawick, 1977). Rather than cause preceding effect, the two events are seen as operating within a circular feedback loop and hence as being interchangeable. In regard to psychopathology, family systems thinkers reject the traditional unidirectional model of influence in favour of a bidirectional model. The identified patient (often the child in the family literature) is no longer viewed as a passive recipient of adverse family conditions but as an influential co-participant in the relationships in which s/he is involved (Haley, 1963).

Systems thinkers are not alone in their rejection of the unidirectional model of influence. Both child study (Fox, 1978; Lytton, 1971) and behaviorally-oriented investigators (Wahler, 1976) have also begun to revaluate family relationships within a bidirectional framework. There is a central feature of the systems approach that distinguishes it from these other schools of thought, however, and that is the unit of analysis utilized both in theory and in clinical practice. While behaviorists have by and large limited their analysis to simple dyadic interchanges (Haley, 1971) involving husband and wife in the case of marital conflict, and parent (generally mother) and child in the case of child psychopathology, systems proponents focus on the entire

network of interchanges within the family. The emphasis, moreover, is on the interdependency of these interchanges. The word "system" in fact refers to just this quality of interdependency (Walrond-Skinner, 1976). In more concrete terms, systems proponents maintain that any one component of the family system, for instance, the mother-child relationship, can only fully be understood within the context of all other components, such as the father-child relationship and the marital relationship. In this sense, the approach can perhaps best be described as one that operates within a multi-directional model of influence.

Apart from providing a distinctive unit of analysis, the family system approach can be further differentiated in terms of the level of analysis that it imposes on interactional material. Rather than explaining family life in terms of specific interactional episodes, the approach associated with both behavioral and developmental schools of thought, family systems thinkers direct their interest to the operation of higher-order rules or principles that concisely govern a wide range of specific interactional episodes. The operation of such principles has been fully elucidated by Don Jackson in one of his early influential papers on the family titled <u>The Study of the Family</u> (1965). In this paper, Jackson asserts that the family is rule governed in the sense that its members interact in an organized repetitive manner. That is, from the mass of

exchanges that occur between family members, recurrent patterns and sequences emerge that can be identified and labelled by observers. More important than the assertion of rule government is Jackson's further assertion that one concise body of rules consistently directs all aspects of family life.[3] Thus, only a relatively small number of higher-order rules need be inferred to explain the uncountable number of specific instances of family interaction.[4]

Rules emerge during the course of development of a family, as members go through the process of defining for themselves a range of mutually acceptable behaviors and expectations. These rules dictate family life in the broadest sense. On one level there are rules that specify the range of activity and of emotional expression considered

---

[3] This initial assumption of rule government may be seen as less important than the further assumption of higher order rule government insofar as the existence of rules is automatically assumed for all phenomena that are considered amenable to scientific investigation. What must of course be recognized is that the family has only relatively recently been conceptualized "as a meaningful unit which can be strategically studied in relation to psychopathology" (Wynne, 1966) and that Jackson's description of the family as rule governed reflects the emergence of this new conceptualization.

[4] This premise of higher order principles lends a certain economy to interactional material which mitigates to some extent the impression of its overwhelming complexity. It is important to note that such a premise is only held in certain theories or models. Bateson's conceptualization of "deutero-learning" or "learning to learn" and biologists conceptualization of genetic codes, for instance, share with systems theory the premise of higher-order principles (Jackson, 1977).

appropriate for family participants. On another level there are rules about rules--whether they can be introduced, commented upon, disagreed with, clarified, negotiated, modified, violated, etc., and if so, through what kind of processes (Foley, 1978; Haley, 1963). Family rules are often not articulated or even known to the participants (Haley, 1963; Jackson, 1977). Once "defined", however, they are not readily dislodged. Even in the face of persistent distress and minimal satisfaction, the rule-governed family system is characterized by a "tough stability" (Bateson, 1959, p. 128), a tendency that is referred to as homeostasis. There are, however, times in the family life cycle, be they in the form of natural life events (e.g., birth of the first child, adolescence, and the "empty nest" period) or idiosyncratic life events (e.g., birth of a handicapped child, premature death of a family member) during which internal and external factors converge in a way that threatens the system's stability (Haley, 1973, pp. 41-64). These threats take the form of any behavior on the part of a family participant that either deviates significantly from the prescribed set of rules or interferes with the rule-governed behavior of other participants. In response to such deviations, regulatory mechanisms that serve to restore equilibrium may be brought into play (Feldman, 1976; Jackson, 1977; Walrond-Skinner, 1976). The symptom that brings the family into therapy, be it depression, psychosis or psychosomatic disorder, is thought

to function in just such a regulatory fashion (Ackerman, 1958; Framo, 1972; Minuchin, Baker, Rosman, Liebman, Milman & Todd, 1975).

This view of the symptom as a homeostatic agent was introduced in the seminal work on family interactions conducted by Bateson, Jackson, Haley and Weakland (1956).[5] These clinicians/researchers were alerted to the operation of such powerful maintenance mechanisms when their attempts to change the behavior of schizophrenic patients were consistently met with strong resistance by the family group, or resulted in the transfer of symptomatic behavior from the identified patient to another family member. Family interactionists have since extended this conceptualization of schizophrenia to a wide range of disorders. Steinglass (1976) and Marshall and Neill (1977), for instance, discuss alcoholism and obesity respectively within a homeostatic framework, while Minuchin and his colleagues discuss diabetes mellites, asthma, and anorexia nervosa along the

---

[5] The idea that the symptom serves a homeostatic function predates the emergence of the family systems model. Psychodynamic theory also regards the symptom as a homeostatic agent in that it functions to maintain the intrapsychic balance. Psychodynamic theory does not, however, extend this perception of the symptom in reference to the family. Although the relationship between symptom and family life is not the focus of any explicit theorizing, psychodynamic theory implies that, at the same time as it stabalizes intrapsychic processes, the symptom destabalizes family processes. Thus, family system and psychodynamic theories differ in their conceptualization of the symptom partially in reference to the identity of the structure on which it operates homeostatically.

same lines (Minuchin, Baker, Rosman, Liebman, Milman & Todd, 1975).

Systems theorists are by no means alone in their recognition of the paradoxical opposition of family members to therapeutic intervention. Case histories documenting the transfer of symptomatic behavior from one family member to another appeared in journals as early as 1942 (Burgum, 1942). Freud (1917, p. 459), moreover, commented on the unexpected resistance of the family to the recovery of the identified patient, though he failed to incorporate the family unit into psychoanalytic theory and practice.[6] Current concern with family opposition appears in the form of widespread reports by practitioners of the refusal or failure of parents to follow recommended treatment procedures. Attempts to address the issue on an empirical level have produced the parental compliance literature.

---

[6] Apart from the use of the words 'healthy', 'sick', 'neurosis' and 'analyst', the following statement by Freud has a decidedly "family systems" tone to it.

> No one who has any experience of the rifts which so often divide a family's will, if he is an analyst, be surprised to find that the patient's closest relatives sometimes betray less interest in his recovery than in his remaining as he is. When, as so often, neurosis is related to conflicts between members of a family, the healthly party will not hesitate long in choosing between his own interest and the sick party's recovery.

Despite this widely shared concern, the various schools of psychotherapy respond to resistance to change in markedly different manners. A common behavioral response, for instance, is to view parental opposition as an impediment to treatment. Attempts to modify a treatment package into a more "manageable" format as a way of marshalling parental compliance is a frequently encountered strategy. Alternatively, factors aimed at enhancing generalization from treatment to home setting are incorporated into therapy in an effort to protect against the remission that often occurs when the hospitalized patient returns to the family. In contrast, family systems adherents accept resistance to change (i.e., tendency towards homeostasis) as a given. They do not expect the family to carry out a treatment program designed to eliminate symptomatic behavior, given the social context in which the symptom operates at the onset of therapy. For instance, husband and wife may fail to encourage the autonomy of an overly dependent adolescent if the resulting increased contact between spouses forces them to confront issues of sexual intimacy and marital disharmony. Rather than attempting to extricate a particular symptom, family therapists work at reconstructing the social context in which it operates.

In sum, systems theorists perceive the family as an organization whose members constantly interact and influence each other in complex ways. The form of this influence is

determined by a set of rules, while the enforcement of these rules is accomplished through the operation of homeostatic mechanisms. Apart from incorporating a novel multi-directional approach to causality, this view of the family is distinct both in its unit and level of analysis.

Implications of the Systems Perspective

The interactional view of the family has far reaching implications for theory and therapeutic practice, as well as for research. In regards to theory, it generates an understanding of personality and symptom formation that diverges sharply from that of individually focused perspectives. While acknowledging the critical role played by the family, individually focused theorists ultimately view intrapsychic processes and conflicts as the chief determinants of personality and psychopathology. Systems theorists, in contrast, locate the determinants of personality in external processes, namely in the interactions of family members. Through the repeated occurrence of particular interactions, patterns of relating emerge which in turn generate implicit expectations about how family members behave towards one another (Beihl, 1979). It is these patterns of relating and expectations about behavior that constitute personality. Symptomatic behavior is similarly linked to the interactional process in that it emerges from and is an expression of widespread disturbances

in family relationships (Feldman, 1976; Haley, 1976; Minuchin, 1974; Whitaker,1975).

The systems perspective also has a direct bearing on the nature of the therapeutic encounter. To begin with, the shift in perspective from dysfunctional behavior to dysfunctional interbehavior requires all family members to enter therapy together (Framo, 1976; Minuchin,1974). Thus, the identified patient is no longer the individual who happens to manifest pathology, but the entire family unit. Apart from the social unit of treatment, the interactional perspective also has implications for the goals of treatment. Broadly speaking, all system adherents regard the reconstruction of the entire pattern of family interactions as the basic clinical objective. Before such reconstruction can occur, however, a preliminary objective must be met, namely the disruption of the family system's existing homeostasis, with the ultimate aim of re-establishing homeostasis at a level that is more satisfying and less disruptive to family members (Foley, 1978). A wide variety of techniques have been employed towards this end: hypnosis, paradoxical procedures, symptom amplification, and explicit prohibition of certain interactions, to name a few (Beihl, 1979).

Finally, the systems perspective has profound implications for research insofar as it demands the creation of new methodologies that are compatible with its basic

tenets. The enormity of this demand becomes apparent when one considers the unique position of family researchers in psychology. First, they have been unable to take advantage of methodological skills that have emerged out of monadic models. In 1963 Haley noted: "There is no adequate precedent for this type of experiment; the methodology which has been developed to experiment with individuals ... does not apply to the measurement of typical patterns of an ongoing system." This sentiment was again expressed by Alexander in 1973. In spite of some obvious methodological advances that have been made in family interaction research since 1963, it is still largely relevant today. Second, researchers have had to continually confront and resolve discrepancies between a deeply entrenched intrapsychic perspective and a relatively undeveloped interactional perspective. In light of such demands, it is not surprising that the links between systems theory and research are often tenuous.

To begin with, the central systems tenet of interdependency has by and large been ignored in the literature. The approach used in the vast majority of interactional studies basically involves the tallying and averaging of various coded behaviors, followed by the statistical comparison of the resulting measures of central tendency across diagnostic categories, settings, or interactional tasks (Fox, 1978; Krahn, 1980; Mishler &

Waxler, 1975). Since these summary measures fail to take into account the position of behavior in ongoing sequences of events, they ignore the interrelatedness of social exchange. Thus, while providing general descriptive information (e.g., the comparative frequency of conflict in clinical and nonclinical families), the typical aggregate analysis fails to reflect the dynamic quality of family process (i.e., the contingencies within which such conflict operates) (Mishler & Waxler, 1975). In contrast are those few studies that have applied sequential procedures to interactional material. Taking advantage of complex computer technology, these studies (Fox, 1978; Krahn, 1980; Mishler & Waxler, 1975) examine streams of behavior in terms of their interdependent and sequential nature.

The systems tenet of rule government is similarly not well reflected in the research. Three general strategies for extracting rules from interactions have been suggested by Jackson (1977): 1) through repeated observation of family exchanges, with careful noting of redundancies; 2) through repeated observation of family exchanges, with careful noting of interactional features that do not appear to operate and 3) through the introduction into the family system of destabilizing forces, followed by careful observation of measures taken by family members to re-establish equilibrium. In other words, observers may be able to infer the operation of rules by noting the family's

reaction to their violation. Jackson has suggested that this latter strategy may be more efficient than those that depend on "tiresome long term observation" insofar as the family's reaction to rule violations tends to be immediate and rather violent.

To my mind, family process researchers have not as yet successfully incorporated any of these rule-extraction strategies into their investigations. In regards to the first two strategies, researchers have relied on frequency counts of interactional features that occur within a single experimental context as a basis from which to infer the operation of governing principles. This approach places distinct limits on the process of rule extraction in that it only allows the individual researcher to comment on those interactions that occur within his/her chosen context, rather than on higher order principles that operate across a range of contexts associated with family life. Of course, the practical exigencies of clinical research often preclude the incorporation of multiple contexts into a single study.[7] This being the case, the onus is on individuals concerned with the notion of rule government to systematically compare various contexts across studies that are identical in all

[7] A small number of researchers have manipulated context by examining the differential effects of experimental tasks (Haley, 1964; Jacob & Davis, 1973; Murrell, 1971, Klopper, Tittler, Friedman & Hughes, 1978; Zuckerman & Jacob, 1979) and settings (Fox, 1978; O'Rourke, 1963) on interaction patterns. Apart from Haley, these researchers have restricted their investigations to nonclinical populations.

other respects and to provide clear evidence of interactional stability across these contexts. Unfortunately, the strong tendency to shun replication research in favour of original research operates as much in the systems field as in other areas of psychology (Riskin & Faunce, 1972).

In regards to the third strategy, attempts to introduce destabilizing manoeuvres are infrequent in family research. (See Reiss, 1978; and Solvberg and Blakar, 1975 for exceptions.) This, I feel, reflects the widely voiced concern that the behavior sampled during research represents, as much as possible, the behavior that occurs during the 'everyday life' of the family. Researchers, moreover, seem to have equated the notion of everyday-family-life with stable-family-life. Such an equation is, for example, clearly reflected in Wynne's (1981) statement of methodological goals, that calls for the use of interactional tasks that allow the family to "feel free and comfortable enough to communicate in a 'natural manner' that is likely to be representative of enduring family patterns."

There are a number of problems with the natural/stable equation. First, it is not consistent with the understanding that the lives of all families, both clinical and nonclinical, are punctuated with recurrent threats to stability, often in the form of developmental crises (see

Haley, 1973, pp. 41-64).[8] Thus, family behavior in the face of stress or crisis is as much a part of 'everyday life' as family behavior under stable conditions. In addition, the equation likely impacts on the power of the typical process research to yield significant interactional differences. The fact that psychopathology often acquires clinical expression during periods of crisis strongly suggests that differences between clinical and nonclinical groups would more likely emrge during stress-induction experimental conditions than during 'everyday life' conditions. Results from a study by Solveberg and Blakar (1975) that incorporates a stressful/stable manipulation lend support to such speculations.

There are various other problems specific to family interaction research apart from those related to the major systems tenets. Some exploration of their nature is in order, particularly for the reviewer of the literature, insofar as these problems have direct bearing on the choice of procedures used to integrate research findings.

---

[8] While both clinical and nonclinical families experience similar threats to stability, family interaction theorists maintain that they can be clearly differentiated in terms of the effect of such threats on the integrity of the family unit and in terms of the nature of the homeostatic mechanisms that are called into play (e.g., symptomatic expression versus the recalibration of the homeostatic mechanism to a more satisfying level).

Difficulties Inherent in the Study of Family Interactions

The root of many of the problems encountered in family research lies in the choice of subject matter, namely, the entire network of interactions that occur in the daily life of the family. Not only are there an infinite number of potentially observable behaviors emitted by each member of the family, but also a complex web of simultaneously occurring exchanges involving the various dyads and triads that make up the family unit. Interactions, moreover, occur across a number of channels of communication, and the type of information that emerges often varies with the choice of channel (Fox, 1978; Riskin & Faunce, 1972). That is, interactions may contain multiple and at times conflicting levels of meaning, some conveyed verbally, others tonally, and still others posturally (Haley, 1967). Another aspect of family interaction that poses problems for investigation is its idiosyncratic nature. While other forms of social exchange involve some idiosyncratic aspects, the intense and enduring quality of family life promotes the widespread development of privately understood codes. Unique forms of expression readily emerge that convey information about a unique present and a unique past common only to members of a particular family (Haley, 1963; Winter & Ferreira, 1967). Furthermore, the form of family communication tends to become increasingly economical over time in the sense that small gestures come to convey a lot of meaning. The

family's reliance on such interactional shorthands is of course inversely related both to the effort required and the capacity of the researcher to accurately interpret interactional data (Jackson, 1977). Finally, the fact that numerous factors external to the family, in particular those of a sociocultural nature, have been found to influence interaction patterns (Sanua, 1961) introduces further complexities into the field of inquiry.

Thus, systems researchers are faced with a decidedly unwieldly subject matter. Add to this the problems associated with clinical research in general (see Haley, 1963; Rabkin, 1965) and the task of examining interaction patterns may seem overwhelming. A number of measures operate to reduce the complexity of interactional research to manageable proportions. To begin with, researchers often restrict their focus in any one study to one dimension of interaction considered relevant to psychological well-being. Certain dimensions have received particular attention in the literature; namely patterns of dominance or power distribution, conflict, affect and communication clarity (Jacob, 1975). In addition, the efficiency and effectiveness of the family-as-a-working-unit and the flexibility/rigidity of the family system have been identified as important in discriminating clinical from nonclinical populations (Doane, 1978a). Further reduction of the complexity of interactional material generally

results from the imposition of some structure on the free flow of family process. While there is almost unanimous agreement on the need for such structure, the question of what form and degree is most appropriate remains unresolved. Common to researchers and critics alike, however, is a concern that the fantastic richness of family interaction not be entirely sacrificed to experimental managability and rigor (Fontana, 1966; Haley, 1963; 1967; Lytton, 1971; Olson, 1970; Riskin & Faunce, 1972; Watzlawick & Weakland, 1977, p. 70). Thus, an essential goal in systems research is to achieve a reasonable balance between demands for manageability and demands for data that reflects family life in a vivid and meaningful way.[9] Phrased in somewhat more concrete terms, the goal of systems research is to gain access to representative (i.e., not highly atypical) interbehavior of family members and to then find the tools with which to perceive, organize, record and quantify it without introducing undue distortion. Further translated, the task becomes that of choosing methodological tools appropriate for interactional material and of applying statistical techniques appropriate for the resulting data. In light of the breadth and complexity of the focus of inquiry, what may be deemed "appropriate" covers considerable ground. Consequently, one would expect

[9] The troublesome nature of this objective has been vividly described by Watzlawick and Weakland (1977, p. 87) in terms of "steering a middle course between the Scylla of overwhelming complexity and the Charybdis of irrelevant detail."

tremendous variability in the methodological features utilized in process research; variability extending to experimental tasks used to generate interactional data, to settings, recording procedures and outcome measures, as well as to ranges of behavior and populations sampled. The impression of an unwieldly research area is thus further heightened.

Three additional difficulties associated with family interaction research warrant discussion. The first concerns the development and use of terminology. Although theorists and researchers have accepted the tenets of the systems perspective, their efforts to conceptualize issues in interactional terms have been met with only partial success. The problem has been that of abandoning the traditional psychological terminology, which reflects a deeply entrenched focus on the individual, and replacing it with the terminology that adequately describes interactional data. While a new terminology has been developed in response to this need (e.g., double bind, communication deviance, martial schism/skew, pseudomutuality, homeostasis, etc.), it has not been used with a high degree of precision (Mishler & Waxler, 1965). The correspondence between terms and constructs has been far too weak in that different terms have been used to refer to identical or overlapping constructs while similar or identical terms have been used to refer to dissimilar constructs (Riskin & Faunce, 1972).

A second difficulty concerns the use of operational procedures. An obvious feature of the family interaction literature is that researchers have not been successful in arriving at a sufficient level of consensus regarding the use of operations. Thus, identical operations have been used to reflect what appear to be distinct constructs (Doane, 1978a, 1978b; Liem, 1980). Moreover, the fact that researchers have applied a relatively large number of labels to a limited set of operations introduces unnecessary confusion into the field.

Finally, researchers have not always succeeded in incorporating their interactional perspective into compatible research designs. Specifically, they have at times focused on the behavior of individuals that make up the family unit when the substantive interest clearly demands that the network of interactions among individuals be of primary consideration. For example, in order to study the distribution of power in disturbed and normal families, it is necessary first to quantify the behavior of each member of the family in relation to that of other members, and then to compare diagnostic categories in terms of the resulting ratios. Instead, some researchers interested in power distributions have restricted their comparisons to the behavior of individual family members (e.g., normal father vs. clinical father). Similarly, in order to investigate coalitions it is necessary first to compare parent-parent

and parent-child interactions within each family unit, and then to compare diagnostic categories in terms of the resulting ratios. Some researchers have instead restricted their analyses to the interaction of either parent-parent or parent-child dyads (Doane, 1978a, 1978b; Jacob & Grounds, 1978).

We are thus confronted with a body of research that incorporates a range of conceptual foci, that utilizes a wide variety of methodological features, and that generates multivariate data that is likely interactive in form. Moreover, it is marked by a terminology that is imprecise as well as by serious problems with operational specification.

Needless to say, all of these characteristics impose considerable demands upon the reviewer who is determined to sort out the unwieldly literature and to extract some meaningful information from it. As a first requirement, s/he must come to terms with the weak correspondences found in family process research between terms, operations and constructs, and between constructs and designs. In order to do so, the reviewer might well follow Doane's (1978a, 1978b) recommendation; namely to classify each study in terms of his/her understanding of its conceptual focus rather than in terms of the conceptual focus specified by the primary researcher. For instance, if a study on "coalitions" compares the supportive interactions of the parent-parent dyad across diagnostic categories without attending to

interfamilial ratios, it would be more appropriate to consider it along with studies on, say, family affect.

As a second requirement, the reviewer must come to terms with the variability in research features. Due to the multivariate nature of the subject matter, no single investigation focuses on more than a small number of the relevant factors and/or utilizes more than a small number of the relevant research strategies. Moreover, given that family research is not directed and coordinated by a central authority, the choice of factors and methodological features varies considerably from study to study. Given such variability in research features, the reviewer cannot expect convergent outcomes to emerge out of his/her cross-study comparisons. In other words, coming to terms with variability in research features and coming to terms with variability in research findings are one and the same requirement. This is no mean requirement, given that the traditional function of research integration has been primarily to evaluate literatures in terms of whether or not findings do converge. Once the reviewer regards outcome variability as an accepted feature of the family process literature rather than as a sign of unproductive research, s/he is free to explore its sources by taking into account the variability in research features.

Reviewers of the family literature have not met either of these demands with a high degree of success. Though a

number of critics have addressed the issue of imprecise terminology and incompatible designs (Jacob & Grounds, 1978; Mishler & Waxler, 1965; Riskin & Faunce, 1972), only Doane (1978a) has attempted to incorporate relevant strategies into her review procedures. (i.e., classifications strategies meant to impose some consistency among terms, operations, constructs and designs). The degree to which she has successfully applied such strategies cannot, however, be easily evaluated since she has unfortunately failed to describe her system for reorganizing the literature under appropriate substantive headings in sufficient detail (Jacob & Grounds, 1978).[10]

The issue of variability in findings has been approached in a particularly unproductive fashion. Reviewers have responded to it by emphasizing two aspects of the research, first its methodological inadequacy, and second, the large number of relevant features and the degree to which they vary from study to study. The focus on methodology has as its basis the understanding that a body of studies must be methodologically rigorous before it can produce consistent and reproducible findings. Conversely, it is understood that a failure to attain such rigour insures that findings will be inconsistent and unpredictable. Thus, the contradictory quality of the family process literature has

[10] Doane has responded to my request for the specifics of her classification strategy by simply reiterating her caution regarding the face-value acceptance of the labels of research foci provided by primary researchers.

been attributed to a variety of design weaknesses: namely widespread failure to insure comparability across experimental groups in setting, demand characteristics and demographic factors, an insufficient use of blind coding procedures coupled with inadequate interrater agreement rates, and a general lack of statistical proficiency (Jacob, 1975; Jones, 1973; Riskin & Faunce, 1972). Accepting the link between methodological adequacy and consistency of outcome, some critics have chosen to integrate only that small subset of the available literature that fully satisfies their methodological criteria. (e.g., Fontana, 1966) This strategy can be criticized on at least two accounts. First, it generally results in a drastic reduction of the data base coupled with a loss of potentially valuable information. Second, the link between methodological rigour and consistent findings in the family process literature is an assumed one that has not as yet been empirically validated.

As an alternative response to the issue of methodology, critics have entered into their reviews studies of varying degrees of rigour and have then differentiated among them on methodological grounds. (e.g., Fontana, 1966; Jacob, 1975). Though never directly expressed, the purpose behind this approach seems to be to afford the reader the chance to place greater weight on the sounder research. Given, however, that the methodological evaluations (usually in

table form) and the summary of findings are presented separately, it is unlikely that the reader is able to coordinate the two sets of information. Of all the critics in the area, only Doane (1978a) has chosen not to differentiate studies on methodological grounds, though she fails to explain why.

In regard to the number of relevant factors and their lack of comparability across studies, reviewers have responded with either pessimism or confusion. Frank (1965) exemplifies the former response by concluding that family process factors "defy being investigated scientifically and defy one's attempts to draw meaningful generalizations from the explorations that have already been done". The more typical response is a series of shuffles that assumes something of the following form: (1) the reviewer demonstrates his/her awareness of the importance of relevent factors during a lengthy discussion of methodological and substantive issues; (2) in the presentation of empirical findings that follows, s/he proceeds to ignore the operation of most factors by relating outcome to only one or two study characteristics; (3) on the basis of apparent inconsistency in the data the reviewer concludes that the literature has failed to differentiate diagnostic groups in terms of their interaction patterns; (4) the reviewer then seems to imply that such conclusions may be unreliable, that different (i.e., consistent) findings might have emerged were there

techniques available with which to take into account cross-study differences in sample, design and analytic features; (5) depending on his/her personal investment in the systems perspective, the reviewer concludes on one of two notes. S/he may reiterate that the literature has been largely unproductive. Alternatively, s/he may highlight any pockets of consistency and promising trends that emerge in the review, and at the same time recommend that methodological improvements and innovations be incorporated into future research.[11] With the exception of Sanua's (1961) suggestion that a central authority be set up to coordinate research efforts, the issue of alternative approaches to variability in research features has gone largely unexplored.

The reviewer's inadequate response to variability is often reflected in the way s/he organizes studies and presents findings. The most common approach has been to present verbal synopses of studies that either examine

[11] Jacob (1975) provides a good example of this approach. While the general tone of his review emphasizes the failure of family research to isolate interaction patterns that reliably differentiate disturbed from nonclinical groups, he is by no means comfortable with this evaluation of the literature. Specifically, he wonders whether "many of the observed inconsistencies across family interaction studies may be a function of differences in diagnostic status of experimental groups, measurement techniques used in assessment of particular domains, data analysis procedures by which results are evaluated and demographic characteristics of family groups." Finally, after promoting the need for improved methodological standards, Jacob concludes with a statement that mirrors the pessimism expressed by Frank: "it would seem impossible to untangle the inconsistencies among studies which differ along a number of potentially influential variables."

hypotheses of interest or illuminate certain methodological or theoretical issues (e.g., Frank, 1965; Goldstein & Doane, 1982; Haley, 1972; Riskin & Faunce, 1972, Wynne, 1981). These synopses are followed by impressionistic statements that summarize the body of research as a whole. Since no attempts are made to organize studies in terms of diagnostic, demographic and methodological characteristics, what emerges is a picture of unaccountable and hence confusing variation.

While incorporating many of the inadequacies just described, two recent integration efforts in the area, by Jacob (1975) and Doane (1978a), differ from previous reviews in three important respects.[12] First, rather than simply selecting studies that illustrate methodological and/or

[12] A number of reviews have since emerged, authored by Liem (1980), Wynne (1981), and Goldstein and Doane (1982). These have been rejected as a base against which to compare substantive conclusions and review methodologies for the following reasons: First, Wynne, and Goldstein and Doane did not intend their reviews to be exhaustive in scope. Rather, these critics have selected studies that illustrate methodological and conceptual interests and concerns, and that identify the types of questions that are currently under empirical focus. Moreover, these critics have relied on boundary guidelines quite different from those selected for the present integrative effort. For instance, a high portion of both sets of studies rely on individual test protocols or standard interview responses to draw inferences about interactional material rather than on direct observational procedures. In addition, many of the selected studies do not focus on clinical/nonclinical comparisons but rather on comparisons within schizophrenic samples. Particular attention is paid to EE (expressed emotion) studies that attempt to identify additudinal factors predictive of schizophrenic relapse. Liem's review, while exhaustive in its coverage, is too restrictive in its time boundaries to provide a

conceptual issues, or that meet rigorous methodological criteria, both Jacob and Doane have aimed at an exhaustive coverage of substantive findings. Second, both Jacob and Doane have attempted to organize the diverse literature by focusing on particular features potentially related to outcome, a strategy used by only one other family process critic (i.e., Liem, 1980).[13] Jacob has organized his review into four substantive domains, namely dominance, conflict, affect and communication. He has further subdivided studies according to the diagnostic category of the clinical population (disturbed or schizophrenic) and the nature of the dependent variable (quantitative process, qualitative process or outcome measures). Thus, rather than simply emphasizing the importance of diagnostic and measurement comparability in primary research, Jacob has appropriately extended these concerns to the integration procedures. Unfortunately, by limiting himself to only three aspects of

comparative focus, in that it only contains studies published between 1975-1979. Its restrictiveness is further ensured by the fact that only 17 empirical studies were published during that time period, the majority of which emerged out of only five discrete research projects. Moreover, the selection guidelines used by Liem are similar to those used by Wynne, and Goldstein and Doane, rather than to those used in the present research effort.

[13] Liem has divided her review into two major substantive areas: deviant role relationships and disordered communication process. These areas are further divided into six specific topics. In addition, Liem differentiates communication studies in terms of whether they examine interactional process or task performance in a manner consistent with Jacob. In my opinion, Liem has used a clearer organizational scheme than either Jacob or Doane.

the data he has failed to attend to a host of other descriptive and methodological factors that, by his own account, influence the nature of family interactions.

Doane's review is, by comparison, slightly less orderly insofar as she has organized her literature on the basis of only two study features, namely substantive focus and type of dependent measure. In regards to substantive focus, she has been thorough. Two broad substantive categories , conflict and dominance, and communication, are further subdivided in her review into seven specific aspects of interaction. In regards to dependent measures, Doane has been less rigorous. She has disaggregated subsets of measures considered by her either to be too narrowly defined or to be tied to constructs in ambiguous and imprecise ways without fully defining the nature of such measurement problems. Moreover, like Jacobs, Doane has not taken into account a variety of factors potentially predictive of outcome. Consequently, both reviewers have failed in their efforts to tackle cross-study variation by organizing the literature into homogeneous subsets.

Both Jacob and Doane differ from previous reviewers in another respect, namely that their conclusions are based on quantified information rather than on impression alone. Specifically, after defining each outcome as either positively significant, negatively significant or nonsignificant, a tally is taken to determine whether the

plurality of studies falls into any one category. If a modal category does emerge it is accepted as the best reflection of the relationship between the independent and the dependent variable under examination. There are a number of problems associated with organizing the literature on the basis of significance levels. Setting these aside for the time being, it is important to recognize that the method of tallying outcomes used by Jacob and Doane signifies a welcomed first step away from purely subjective evaluations.

Though Jacob and Doane have utilized similar approaches and have accessed data bases that overlap to a large extent[14] they have arrived at conflicting conclusions as to the status of the literature. Doane is in fact the only reviewer in the area to unequivocally conclude that the interactions of normal and disturbed families can be reliably differentiated from one another.

After wading through the more vitriolic aspects of their attempts to address their differences (see Doane (1978b) and Jacob & Grounds (1978)), what emerges are criticisms that centre around issues of organization. Both reviewers

---

[14] The bodies of literature reviewed by Jacob and Doane differ in three respects: 1) Only Doane's review incorporates the communication deviance research conducted by Wynne and his colleagues 2) Doane's review includes research conducted after June, 1973, Jacob's cut-off date and 3) there are differences in the theoretical and methodological writings referenced by the two critics.

criticize each other for failing to choose dimensions and to organize studies around them in a manner that insures that comparisons are being made among similar sets of data.

Two responses to the Jacob-Doane debate come to mind. First, it is a valuable exchange insofar as it highlights a number of important review requirements as well as the difficulties involved in meeting them. In particular, the need to classify studies in terms of relevant features and the need to correct for inappropriate use of terminology and design come to the fore. Second, the question of whose conclusions best reflect the data has not and cannot be resolved, given the type of information presented in both reviews. Although both Jacob and Doane have made efforts to organize and quantify their material, their failure to sufficiently explore sources of variation prevents the reader from untangling real from apparent inconsistencies. The summary of findings offered by both Jacob and Doane actually differ very little in their effect from the narrative and impressionistic style of review, in that the reader is confronted with a huge amount of contradictory information. The following extract from Jacob's summary of findings illustrates this well and leaves little doubt as to the need for further improvements in the integration of the family interaction literature. The fact that this extract occupies only one-half page out of 11 1/2 similar pages of journal text underscores the extent of the burden still imposed on the reader by this approach to integration.

Quantitative process measures. Five studies assessed dominance by means of verbal frequency measures such as successful interruptions, talking time, statement length, and to whom acts are directed (Cheek, 1964a, 1964b, 1965a; Farina & Holzberg, 1968; Ferreira et al., 1966; Lennard et al., 1965; Mishler & Waxler, 1968). Of nine total comparisons, four indicated no reliable differences between schizophrenic and control families (Farina & Holzberg, 1968 - talking time; Lennard et al, 1965- number of communications; Mishler & Waxler, 1968 - successful interruptions and being successfully interrupted - for these two measures the authors reported significant group differences only when a "well sibling" was present). Although the remaining five comparisons reported significant differences between schizophrenic and control families, the direction of such differences was found to be rather inconsistent across studies. Cheek (1964a, 1964b, 1965b) noted that normal sons participated more than schizophrenic sons, that schizophrenic daughters participated more than normal daughters, and that normal mothers participated more than schizophrenic mothers (interacting with male or female children). Ferreira et al. (1966) observed that normal children talked more than schizophrenic children, although interactions containing male and female children were not analyzed separately. In seeming contrast with Cheek's results, Mishler and Waxler (1968) reported that normal daughters participated more than poor premorbid daughters and spoke in longer intervals than good premorbid daughters, whereas good premorbid fathers spoke in longer intervals than normal fathers (with male children only). For "acts directed to" data, the same investigators reported that normal fathers received more acts than schizophrenic fathers, which was largely a function of the differential rate with which schizophrenic and normal sons directed acts to fathers; and schizophrenic sons received more acts than normal sons from their parents, fathers, and mothers.

Many of the issues that have been referred to in the above discussion of problems for integration are expanded upon in the chapter on integrative strategies that follows.

RESEARCH INTEGRATION TECHNIQUES

Over the past decade, a number of critics in the fields of educational and psychological research have begun to clearly articulate their dissatisfaction with the slow accumulation of knowledge in their disciplines (Glass, 1976, 1977; Kavale & Glass, 1981; Light, 1978; Light & Smith, 1971; Pillemer & Light, 1981; Rosenthal, 1978; Smith, Glass & Miller, 1980). Basically, they have presented the problem as a failure to use review techniques suitable for extracting knowledge from an ever-expanding body of information. In the present chapter their criticisms of traditional review techniques are presented, along with descriptions of the alternative techniques that they have developed for integrating educational and psychological literatures.

Traditional Approaches to Research Integration

Narrative Reviews

During the 1940's and 1950's and well into the 1960's, scholars primarily reviewed the literature narratively by listing and summarizing those studies that examined the effects of any of a number of factors on a dependent

variable of interest (Glass, 1976).[15] Such an approach was and is adequate only under limited circumstances: when the number of studies is very small and/or when the pattern of outcomes is uniform (Light & Smith, 1971). Research in education and psychology is, however, most often characterized by large numbers and nonuniformity. The literature is nonuniform in two respects; first, studies in a particular problem area tend to vary widely in choice of experimental designs, measurements, classes of subjects and contexts to name but a few properties. Second, variation in findings abound. Critics have responded to the welter of findings by abandoning the straight narrative style of review in favor of two approaches that attempt to impose some order on divergent information.

The Favorite Studies Approach to Research Integration

In what can be termed the "favorite studies" approach, the reviewer essentially attempts to create consistency by examining only a small subset of the available literature (Light & Smith, 1971). This subset is chosen on the basis of experimental rigor, the assumption being that methodological violations are responsible for much of the lack of convergence among findings. The expectation is that rigorous pruning of studies will bear unmarred fruits.[16]

[15] A number of reviews of the family process literature are essentially narrative in form, for instance those by Frank (1965), Haley (1972) and Riskin and Faunce (1972).

This expectation would be somewhat reasonable if literatures could be neatly divided into acceptable and unacceptable subsets. However, while "good" and "bad" studies do exist, most could more accurately be placed at some midpoint on a scale of excellence. Glass (1977) presents a hypothetical but easily recognizable group of 100 studies in which studies 1 to 10 are weak in representative sampling but strong in other respects, studies 11 to 20 are weak in measurement but otherwise strong, studies 21 to 30 are weak in internal validity alone and so on. Given such a pattern, Glass argues that the choice of which studies to include in a review is largely left to the idiosyncratic leanings of the individual critic. Even if sound guidelines were available for such a choice, contradictory results might still emerge. The only way to circumvent such an outcome would be to reduce the literature to one "trustworthy" study and to then abide by its results, the logic of which need not be commented upon (Light & Smith, 1971). The favorite studies approach to research integration has been subjected to one other perhaps more essential criticism. Glass and his colleagues (Glass, 1976, 1977; Smith, Glass & Miller,

---

[16] Fontana (1966) provides a clear example of the favorite studies approach in his review of the family process literature. Other such examples in the family systems field are provided by Wells and Dezen (1978) and Wells, Dilkes and Trivelli (1972). The latter review of the family therapy literature is a particularly extreme example of the favorite studies approach in that its authors were only able to unearth two outcome studies published in major journals in social work, psychiatry and psychology from 1950 through 1970 that met their methodological standards.

1980) repeatedly observe that it is an empirical question whether the results from studies that meet the reviewers chosen set of methodological standards are at variance with results from studies that violate these standards. The phenomenon under review may or may not be sensitive to any of a number of methodological deficiencies. It is an important task for the research integrater to determine whether specific design and analytic features that characterize a body of literature do in fact covary with research findings and, if so, to what extent. "Assuming studies out of existence" (Smith & Glass, 1977) in the absence of empirical substantiation results at best in a marked reduction in the reviewer's data base and at worst in distorted conclusions.

One final point concerning the "favorite studies" approach: along with the expectation of consistency emerging out of tight designs, is the understanding that consistency emerging out of any other set of circumstances is spurious (see Eysenck, 1978). Glass (1977) invokes the law of parsimony in his attack of this position. Referring to his hypothetical group of 100 studies, he maintains that it is not unreasonable that such a collection lend support to one conclusion in spite of the incorporated admixture of design flaws. Efforts to counter such a conclusion would involve invoking an embarrassingly large number of multiple causes (i.e., the observed effect can be caused by weakness

in measurement, or by maturation and testing effects, or by weakness in data analysis and so on).

The Voting Method

A more flexible approach to the integration of large and seemingly patternless literatures involves classifying each study according to outcome (i.e., statistically significant or nonsignificant) and then examining the relationship between outcome and one or two chosen properties of the literature via crosstabulation procedures. Glass (1976) provides a reference list of such reviews. While labelling these efforts as "praiseworthy", he considers their methodology inadequate on three accounts, all of which result in large reductions in information: 1) Continuous quantitative information regarding differences in outcome is reduced to dichotomous measurement. 2) Information on all but one or two properties of the studies as well as interactions among properties is excluded from consideration. 3) Multivariate information can not be accessed via simple two-factor crosstabulation procedures.

A variant of the crosstabulation approach, and the methodology currently in widest use, has been labelled the "voting method" by Light and Smith (1971).[17] This method can actually be considered a hybrid between the crosstabulation

[17] The most recent reviews of the family process literature, those by Jacob (1975) and Doane (1978), utilize the voting method of integration.

procedure and the straight narrative review. While an association is drawn between one property of a study (occasionally two) and the statistical significance of the outcome, it is often reported in terms of verbal synopses followed by a verbal tally, rather than in terms of statistical operations. The voting method is described by Light and Smith (1971, p. 433) in the following manner:

> All studies which have data on a dependent variable of interest are examined. Three possible outcomes are defined. The relationship between the independent variable and the dependent variable is either significantly positive, significantly negative, or there is no significant relationship in either direction. The number of studies falling into each of these three categories is then simply tallied. If a plurality of studies falls into any one of these three categories with fewer falling into the other two, the modal category is declared the winner. This modal categorization is then assumed to give the best estimate of the direction of the true relationship between the independent and dependent variable.

The voting method is not considered adequate to the task of depicting, organizing, or extracting meaningful associations from confusing bodies of data. A number of drawbacks in addition to (and overlapping) those discussed in reference to the general crosstabulation procedure have been commented upon by its critics (Glass, 1977; Kavale & Glass, 1981; Light & Smith, 1971; Smith, Glass & Miller, 1980). To begin with are those problems resulting from the use of significance levels as a basis for organizing study outcomes. Such organization builds in an automatic bias in favor of large sample size, as large-sample studies are more

likely to attain significance than their small-sample counterparts. Kavales and Glass (1981) illustrate this weakness of the voting method by describing a hypothetical body of literature consisting of nine small-sample findings that approach but do not quite reach significance and one large-sample finding that does reach significance. They argue that a conclusion based on a straight tally of one for and nine against would be decidely counterintuitive. There is a further limitation to basing decisions solely on significance levels; such a figure relate little about the strength or importance of empirical relationships under review. The informative value of significance levels has in general been highly overrated by researchers (Glass, 1977).

A second criticism of the voting method centers around a false assumption, namely, that any finding that does not fall into the modal category is merely a consequence of sampling variation or undetected experimental error (Light, 1978; Light & Smith, 1971; Pillemer & Light, 1981). Light argues that an examination of the distribution of significance levels from a literature under review reveals whether or not such an assumption is justified. He asks the reader to consider 189 outcomes of which 58 are positively significant, 50 are negatively significant and 81 are nonsignificant. The user of the voting method would likely point to the absence of directional consistency among these outcomes. S/he would then conclude that positive findings

largely cancel out negative ones, thereby providing little in the way of evidence in either direction. However, the large number of positive and negative findings in this example clearly contradicts a conclusion of random variation around a true effect of zero. With a two standard error criterion, one would expect sample variation to yield approximately five significant outcomes in each direction rather than 108 such outcomes.

Critics who utilize the traditional techniques of research integration that have just been described generally end their reviews (as do their readers) on one of two notes: The more pessimistic among the ranks decry the state of the art and suggest that the particular area under examination be abandoned in favor of more fruitful avenues.[18] The more optimistic contingent (and one that seems to be the better represented) sounds the call for more and better research.[19] Proposed guidelines for future research usually focus on two sets of criteria, one related to greater methodological

---

[18] Frank (1965) concludes his review of the family process literature on a particularly pessimistic note: "Apparently, the factors which play a part in the development of behavior in humans are so complex that it would appear that they almost defy one's attempts to draw meaningful generalizations from the exploration which has already been done."

[19] Fontana (1966) and Riskin and Faunce (1972) are two reviewers of the family process literature who recommend a blending of optimism and critical awareness. Jacob (1975) stresses the need for, and feasibility of, methodological improvements in the field. His general tone, however, is decidedly more pessimistic than that of most process reviewers.

rigor and the other to greater cross-study standardization. The issue of methodology has already been discussed. In regards to greater standardization, critics of traditional research integration methods have downplayed its importance both on practical and heuristic grounds. Along wth questioning the feasibility of a centralized authority, Glass (1977), for one, feels that such efforts would ultimately "exert a conservative drag on inquiry."[20]

The Model of Scientific Growth Underlying Traditional Reviews

As pointed out by Glass and Light and their respective colleagues (Glass, 1976, 1978; Light, 1978; Light & Smith, 1971; Pillemer & Light, 1981) the traditional approaches to literature review arise from and are maintained by a particular way of conceptualizing scientific growth; namely that knowledge accumulates in a regular hierarchical fashion, with the results from one study building on those of a preceding study. Implicit in this conceptualization is the belief in the correspondence between one true study and one true outcome. If well designed and executed, that 'true' study is expected to yield consistent results, regardless of the number of times it is replicated. In this

---

[20] In the family process field, Sanua (1965) does explicitly recommend that a central authority be set up to coordinate research efforts. Most other process reviewers carp on the lack of cross-study standardization but, in contrast to Glass and Sanua, fail to take any specific stance on the issue.

scheme, inconsistencies are viewed solely as products of experimental error and/or sampling variation. Now, such a paradigm is more or less accurate, depending on the field of scientific inquiry as well as on the focus of inquiry within a particular field. Glass (1978) maintains that it is important to differentiate among areas of inquiry with regard to their predictability. The more predictable the area, the more orderly the accumulation of knowledge and hence the more fitting the paradigm. The less predictable the area, the more variable the outcome and irregular the accumulation of knowledge; hence the need to modify the model of scientific growth. While research in celestial mechanics or biology typifies the former pattern, as already indicated, much of the research in education and psychology falls into the latter. Similarly, within psychology, laboratory research on a particular learning theory is more predictable than family interaction research or therapy outcome research.

Individuals who conduct research in education and psychology have long been aware of the nature of the inquiry in which they are involved (see Light & Smith, 1971 for a discussion of the distinguishing features of educational research). A highly variable response pattern is an expected feature of the data. In fact, a marked lack of variation is suggestive of relationships so well known that they are no longer interesting as topics for study. The

task that the researcher sets for him/herslf as a route to comprehending phenomena is to search for and explain variation in the data (i.e., to partition the variance). This search takes place at the design stage (by incorporating factors likely to account for significant portions of the variance) and at the analysis stage (by choosing multivariate techniques that afford full exploration of main effects and interactions among factors).

Integration of research bodies can and ought to be approached in a similar fashion. When literatures yield highly consistent findings, there is little need for synthesis; the phenomena under investigation are no doubt straightforward and well understood (Pillemer & Light, 1981). When, however, literatures yield highly variable findings, there is a need to thoroughly examine the sources of the discrepancies. While there is no room to manipulate design features of completed studies, the full range of statistical procedures can and ought to be employed to study variation in findings (Glass, 1976, 1978). Reviewers can only appreciate the need for such an approach when they begin to perceive such variation as an index of data that is multivariate and interactive in form rather than as error; or, as phrased by Light (1978), when they begin to view variation as a signal rather than as noise. Critics of traditional review methods have been calling for just such a cognitive switch. At the same time, they have been

developing approaches to research integration more appropriate to problems addressed in psychology and education. Three such approaches will now be discussed, with particular attention paid to the approach referred to as meta-analysis.

Alternative Approaches to Research Integration

The Combined Probabilities Approach

The aspect of research integration that has received the most persistent attention is one that has long concerned statisticians and only recently reviewers, namely the problem of obtaining an overall level of significance from an aggregate of studies. Rosenthal (1978) references the work done in the area from the 1930's onward. He also summarizes nine methods of combining probabilities from two or more studies testing the same directional hypotheses and specifies each one's advantages and limitation. The combined probabilities approach to research integration has two objectives. The first, which is common to all integration procedures, is to characterize a large mass of information about relationships in terms of some primary conclusion(s) (Light & Smith, 1971). The second objective concerns the gain in power attained through an increase in the size of the overall sample. Light and Smith discuss the approach in terms of a marshalling of statistical power. Gage (1976) conceptualizes it as a means of determining

whether individual studies, many of which fail to demonstrate a significant effect, can as an aggregate acquire sufficient power to support the rejection of the null at a high level of significance. A problem develops, however, when attempts are made to combine probabilities of a group of studies which contain many instances of significance but fail to demonstrate any directional consistency. Failure to take into account and adjust for sources of such inconsistencies either prior to or during the aggregation procedure can result in a combined probability value which does not accurately reflect either the whole composite or any of its parts. The cluster approach to research integration designed by Light and Smith (to be discussed) provides techniques for taking sources of variation into account prior to combining probability values in order to insure against such inadvertent distortion of synthesized findings.

Critics who accept the goals of the combined probabilities approach recognize a number of drawbacks associated with its use (Cook & Leviton, 1980; Light & Smith, 1971; Pillemer & Light, 1981; Rosenthal, 1978). The first one is common to all statistical manoeuvres that focus only on significance; that is, the positive relationship between attainment of significance and sample size. Because of such a dependency, these critics caution against the use of such procedures without concurrent use of integration

techniques that focus on other aspects of the data. Glass (1977) carries the point a step further by questioning whether the consideration of statistical significance has any place in research integration. Given that literatures under review usually involve large numbers of studies with correspondingly larger numbers of subjects, he argues that one can expect the null hypothesis to be routinely rejected. Consequently, he recommends that the concern with inferential issues be preempted in research integration by a primary concern with descriptive matters.

Another problem for the combined probabilities approach which is related to the issue of sample size is the well-known bias in published material in favor of positive findings.[21] This has been labelled by Rosenthal (1978) as the "file drawer problem". The solution recommended by Rosenthal involves calculating the number of unpublished studies showing no effect that would have to be "on file" in order that the positive published outcomes reflect nothing more than sampling error. This calculation has been dubbed a fail-safe N (Cooper, 1979). If the resulting estimate is larger than what might be reasonably expected, the reviewer can feel somewhat confident that an obtained combined probability value is a valid way of characterizing the

---

[21] A case has also been made for the possibility of bias operating in favor of the null. When research is being conducted on a theory that has fallen out of favor, journal editors may be more interested in publishing damning than supportive evidence (Rosenthal, 1978).

relationship under review, rather than a mere artifact of the publishers' selection procedure. Rosenthal cites as an example of this procedure his synthesis of studies on the effects of interpersonal expectancy, in which a file drawer of 50,000 unpublished studies would be needed in order that the obtained combined p be nonsignificant. Recently Strube and Hartmann (1983) have recommended that the fail-safe N be interpreted with a measure of caution. These critics point out that the informational value of the calculation is largely dependent on the correctness of its underlying assumption; namely that the sample values used in its computation are representative of population parameters that at no point become known.

The Cluster Approach

The cluster approach to the review of literatures was first presented by Light and Smith in 1971. Glass (1977) later referred to this presentation as the first instance of serious attention to the methodological and technical problems of research integration. The paper's distinction rests on the fact that 1) it offers the first cogent criticism of traditional methods of research integration and 2) it addresses the problem of variation among findings directly. Rather than dismissing such variation as an indication of confusion, the authors present guidelines that enable the reviewer to tease out its sources and hence make it understandable.

Light and Smith begin by questioning a standard premise of reviewers, namely that studies which examine identical hypotheses utilize samples that are randomly drawn from a common population. They argue that a more profitable starting position is to assume that such studies utilize natural aggregations within a population that often differ from one another in discernible and systematic ways. These aggregations are labelled clusters. The characteristics on which they differ may be those related to the selection of subjects, the specific details of the experimental treatment, or the context within which the treatment occurs.

The task that Light and Smith set for reviewers is to determine whether clusters do in fact differ from one another, and if so, to pinpoint the sources of such differences. If clusters are found to be similar on all of a number of specified criteria,[22] Light and Smith recommend that their data be combined to allow a maximally powerful test of the relationship under study. If clusters are found to differ on any one of the specified criteria, they recommend that appropriate statistical adjustments precede aggregation procedures. Failure to detect and adjust for such differences not only insures conflict among findings but more important, prevents the reviewer from untangling

[22] Light and Smith (1971) differentiate among clusters in terms of means of variables, variances, relationships between dependent variable and one or more covariates, subject-by-treatment interactions and contextual effects (i.e., features that are uniform for all subjects within any given cluster).

real from apparent contradictions.

Although Light and Smith state that an increase in power attained by combining data is the "ultimate purpose" of their procedures, their presentation of the cluster approach clearly focuses on another objective: to study variations in patterns of findings that emerge from studies that differ on a variety of characteristics. This objective is shared by what clearly has become the most widely used of the innovative integrative techniques, namely Glass and his associates' meta-analysis. Before discussing Glass's procedures, it is worthwhile to differentiate between cluster analysis and meta-analysis on the following grounds: 1) Light and Smith are more concerned with methodological quality than are Glass and his colleagues. Each study included in a cluster analysis must meet a variety of standards related to design and instrumentation. In contrast, Glass and his colleagues reject any a priori elimination of studies from a meta-analysis. Instead they recommend that each study be rated on methodological rigor and that the relationship between degree of rigor and outcome be empirically examined.[23] 2) Light and Smith's

[23] Arguements in favor of modifying Glass's full-inclusion-sampling-strategy have recently been articulated by Cook and Leviton (1980) and Strube and Hartman (1983). While continuing to promote the classification of studies on quality, the critics suggest that renewed attention be paid to basic validity standards. Such a recommendation does not, it must be noted, undermine the basic meta-analytic sampling approach that emphasizes comprehensiveness. Further discussion of issues relating to meta-analytic sampling strategies is contained on pp.

procedures require the use of raw data while Glass's require the use of group data. The question of whether the increase in information obtained through the use of original data justifies the time and effort involved in its accumulation and processing must remain unanswered for now insofar as raw data has largely been unaccessible. This situation may gradually change as journal editors come to value integrations of research findings as much as they do original research. They may then stipulate that researchers submit original data along with written reports, to be preserved for the use of both reviewers, and researchers engaged in secondary analysis. At this time, however, the lack of available raw data detracts from Light and Smith's recommended procedures. (This is not to say that Glass's reliance on group data insures the easy integration of findings. On the contrary, primary researchers too often fail to report the most basic of summary measures utilized in meta-analysis.) 3) While cluster analysis attempts to integrate studies that use identical measures, meta-analysis attempts to integrate studies that use conceptually-related measures (Cook & Leviton, 1980). This distinction has implications for the types of conclusions that can be drawn from the integrative procedures. Light and Smith's use of identical measures restricts their conclusions to relatively specific statements about the effects of the treatment examined in any one cluster (e.g., Head Start programs

promote reading readiness to a greater extent than other pre-school programs, under a specified set of conditions). In contrast, the use of a range of conceptually related measures by Glass and his colleagues generates broad summary statements about whole areas of research (e.g., psychotherapy does have a beneficial effect). Such broad summary statements can and ought to then be systematically narrowed down in a meta-analysis by focusing on progressively more precise questions involving mediating variables (e.g., the benefits of psychotherapy are more pronounced with subjective as opposed to behavioral outcome measures) (Kavale & Glass, 1981). 4) The statistical procedures used by Light and Smith to study the relationships between outcome and various characteristics of studies are more cumbersome than those used by Glass and his colleagues. Light and Smith examine the properties of each individual cluster independently and then compare a set of clusters in terms of these properties. Glass characterizes each study in terms of a chosen set of properties and then simultaneously enters these properties into a single regression equation to study their effect on the magnitude of the outcome. Glass (1978) champions the use of statistical regression analysis because it is "efficient, flexible and altogether handy". 5) While critical evaluation of the cluster approach is still missing (Cook & Leviton, 1980), meta-analysis has been repeatedly evaluated since its introduction. Apart from theoretical and

statistical critiques, a growing number of applications of meta-analysis to substantive bodies of literature provide empirical demonstrations of the utility of this integrative approach.

Meta-Analysis

Meta-analysis is defined by Glass (1977) as "the statistical analysis of the findings of many individual analyses". Its general goal is to describe in quantitative terms 1) an aggregate of findings and 2) the relationships between findings and characteristics that best define a body of studies. This goal is accomplished by means of a series of techniques, beginning with the quantification of study outcomes and study characteristics and followed by the quantification of relationships between outcome and characteristics. This latter step involves entering study characteristics into a regression equation as predictor variables and regressing them onto study outcomes which serves as the criterion in the equation.[24] In the following section, the major tasks of a meta-analysis are described in some detail. Particular attention is directed to problems

[24] The repetoire of available meta-analytic techniques has grown since the integrative approach was first introduced. Such development has taken place in response to unique problems associated with each application of meta-analysis to a new substantive area. Some of the recent innovations outlined by Strube and Hartmann (1983) include multidimensional scaling techniques applied to the construction of independent variables, and curve fitting procedures appropriate for studies with quantitative independent variables.

inherent in the approach, all of which have been ackowledged by Glass and his colleagues. These problems have recently been elaborated upon by a number of review methodologists who essentially promote meta-analysis but are concerned about its irresponsible usage. Recent papers by Cook and Leviton (1980) and Strube and Hartmann (1983) were relied upon heavily to bring order and direction to this discussion. These critiques appear to based on a thorough grasp of both statistical procedures and review objectives.

Sample selection procedures. As a first step to a meta-analysis, the reviewer must decide on a set of guidelines that will enable him/her to accumulate a body of studies representative of the research domain of interest. When the literature is of small or moderate size, the reviewer may be able to examine and integrate all relevant material. In most cases, however, literatures are too large to allow such comprehensive coverage and sampling procedures must be introduced. While such procedures should ideally guard against the operation of selection bias, current editorial practices prevent such a condition from being realized. Editorial policy is first of all biased in the favor of significant findings (Chase & Chase, 1976). Such a bias has a number of implications for meta-analysis; first, it results in overestimations of average magnitudes of study outcomes and second, it generates a data base marked by overrepresentation of Type I errors of inference (Strube & Hartmann, 1983).

In addition to the bias in favor of significance, editorial policy also promotes a bias in favor of studies that are compatible with the current Zeitgeist, both in terms of methodology and theory (Strube & Hartmann, 1983). With respect to methodology, the ready entrenchment of methodological tools into published literatures prior to any clear evidence of their worth has been identified as a cause for concern (Goldfried, 1977). With respect to theory, the tendancy to favor findings supportive of the prevailing ideological climate while ignoring the less popular results was first vigorously reviled by Eysenck in 1964. It has recently come under renewed attack by a number of review methodologists (e.g., Rosenthal, 1978; Strube & Hartmann, 1983) who have all commented on the political nature of the publication process.[25] Zeitgeist bias affects research integration insofar as it restricts the range of methodological and theoretical variables open to consideration (Strube & Hartmann, 1983).

Problems of selection bias cannot be linked solely to editorial policy. The practices of primary researchers also hamper unbiased research integration. Strube and Hartmann have drawn attention to the known but oftentimes ignored

[25] In a recent meta-analysis of the literature on the topical issue of sex bias in psychotherapy, Smith (1980) demonstrated that published and unpublished studies can yield average effect sizes of different magnitudes. While published studies showed a tendency for therapists to hold more negative stereotypes of women relative to men clients, unpublished studies failed to reveal any such sex-related bias.

fact that researchers generally fail to include in their reports all the outcome information needed for an unbiased data base. Such incomplete reporting may involve actual suppression of the null or unexplainable results that detract from a "clean" presentation of findings. Or (more defensibly to my mind) it may involve a discrepancy between the amount of statistical information attached to reports of significance and nonsignificance. The strong tendency to restrict actual statistical data to reports of significance while relying on verbal reports of nonsignificance precludes the incorporation of accurate outcome estimates for nonsignificant findings into a meta-analysis (Strube & Hartmann, 1983). In addition, primary researchers share with journal editors a reluctance to work outside of popular theoretical and methodological boundaries and, in so doing, contribute to Zeitgeist bias. As a consequence, tired but reliable research paradigms tend to dominate journal space long after their informational value is spent.[26]

In sum, current editorial and academic practice more or less insures that research integrators are confronted with educational and psychological literatures that are biased in terms of informational, ideological, and methodological content. There are no easy solutions to this problem. With

---

[26] While such restricted research practice provides a measure of safety to individuals who are forced to respond to the publish-or-die dictate that holds sway in the current academic community, it has an obviously unhealthy effect on scientific inquiry.

respect to significance bias, Strube and Hartmann have recommended the establishment of central repositories for unpublished as well as published material, coupled with efforts directed at sensitizing editors and researchers to the value of complete and detailed reports of findings. Such sensitization will occur more readily once the call for some reconsideration of the informational value of nonsignificance is finally heeded (see Smart, 1964). With respect to Zeitgeist bias, rather profound changes in the operating contingencies and values underlying psychological and educational research are in order. While awaiting such remediation, the research integrator must be content to rely on the following strictly palliative measures: first, s/he must fully detail for the reader sampling procedures and selection guidelines. Second, s/he must make some assessment of the impact of selection bias on the interpretability and generalizability of the conclusions drawn from the integrative procedures.

Specification of inclusion data. After deciding on an appropriate sampling strategy, the meta-analyst must direct his/her attention to the second phase of sample selection, namely, the precise specification of inclusion criteria. There is considerable debate among the ranks over the issue of inclusion criteria (Strube & Hartmann, 1983). On the one hand there are reviewers who promote the use of restrictive quality standards. Most meta-analysts, however, have tended

to follow Glass's guidelines and argue in favor of full inclusion. Support for this position rests on an assumption of equal methodological bias across studies. Such support has recently been somewhat weakened by Cook and Leviton (1980) who question the certainty of such an assumption, particularly in reference to literatures that are biased predominantly in one direction. Under such conditions, Cook and Leviton argue, the inclusiveness of meta-analysis relative to typical qualitative reviews may be of limited value, insofar as the number of studies then becomes a "conceptual irrelevance". To illustrate the problem of unidirectional bias, Cook and Leviton cite the work of Campbell and Erlebacher (1970) and Director (1979) that traces the ostensibly harmful effects of compensatory programs on the widespread use of treatment subjects who have lower pretest scores than their notreatment controls.

For less vulnerable literatures, Cook and Leviton recommend continued support for the standard meta-analytic practice of examining the effect of bias by coding studies on quality. This approach is most readily executed with respect to methodological quality (see Glass et. al., 1981). Its success is, of course, dependent on the reviewer's capacity to ferret out methodological and procedural invalidities. Less easily addressed is the issue of conceptual quality. Although intuitively it would seem that the appropriateness of a test of theory (i.e., its

conceptual validity) would have greater impact on the interpretability of the findings than the strength of a test (i.e., its methodological validity), meta-analysts have by and large restricted their attention to nominal, demographic, and methodological characteristics at the expense of theoretical characteristics (Cook & Leviton, 1980). This bias no doubt reflects the problems involved in operationalizing the feature 'theory relevance' or 'theory appropriateness'. Problems are particularly likely to arise in substantive areas marked by the absence of a well-delineated conceptual focus.

Regardless of the difficulty imposed on the meta-analyst, issues of construct validity can no longer be overlooked. Indeed the position promoted by Strube and Hartmann warrants strong consideration; namely that an inability to evaluate a literature in terms of construct validity should lead the reviewer to question the preparedness of a research area for integration.

Quantification of study features. As a second step to a meta-analysis, each study must be characterized in terms of a set of features thought to be potentially predictive of study outcome. Thus, unlike other integrative techniques that characterize literatures on the basis of one or occasionally two properties, meta-analysis provides the means for describing literatures in terms of as many or as few features as deemed profitable by the reviewer. The one

constraint on the number of features is the size of the sample of findings available for review. If the number of features is too large relative to the sample size, the data is stretched too thin to afford sufficient power to the test of relationships among outcome and study characteristics via multiple regression procedures. The choice of features used to explore a substantive area is based on both theoretical and methodological considerations. What is apparent is that the appropriateness of such choices is dependent not only on the astuteness of the reviewer, but also on the nature of the literature under review. Appropriate choices are most readily made with literatures that are well delineated with respect to conceptual foci and the methodological strategies. The adequacy of feature selection is less reliable with empirical literatures that lack clearly articulated conceptual foci and that lack a firmly established consensus among researchers regarding the appropriate use of research paradigms and methodological tools. In such cases it is difficult for the reviewer to make intelligent a priori decisions on either logical or theoretical grounds about the potential utility of study characteristics.

Glass (1977) points to another practical problem associated with the process of quantifying study features, one that centers on definition. Although some study features can always be defined without ambiguity, others

defy straightforward definition (Glass, 1977). Features can be ambiguous first with respect to their operational specifications. Difficulties associated with developing unambiguous operational schemes have, to my mind, been glossed over in critiques of meta-analysis. The degree to which unambiguous schemes can be developed is, as with feature selection, dependent as much on the nature of the literature under review as on the ingenuity of the reviewer. While some literatures are best described in terms of features that are readily operationalized and readily translated into quantifiable units, others may be more usefully described in terms of features that demand subjectively-toned definitions and that cannot be automatically linked to the natural number system. Some assessment of the ambiguity of the definitional scheme applied to a substantive area is always in order. The use of at least two independent coders followed by the calculation of inter-observer reliability coefficients allows for such assessment.

Study features can also be ambiguous with respect to their definitional boundaries. This problem is best elucidated through example. It would likely arise, for instance, when attempting to characterize a therapy outcome literature in terms of the feature 'time spent in therapy'. The reviewer may be uncertain whether to operationalize 'time' in terms of number of actual hours of therapist

contact, number of hours spent in a 'therapeutic' setting (e.g., hospital), number of weeks during which a therapeutic contact is maintained, etc. With respect to this source of ambiguity, Glass suggests classifying each feature in as many ways as needed to capture it fully and to later eliminate distinctions that prove to be irrelevant.

A study feature may be defined as irrelevant under a number of conditions. First, it may be so defined when it proves to be stongly related to another study feature. Such redundancy can be readily assessed through the construction of a correlational matrix into which is entered the entire set of predictor variables. A feature can also be defined as irrelevant if information needed for quantification is characteristically too scanty to accommodate the use of missing data techniques in correlational procedures utilized during meta-analysis.[27] Finally, irrelevance can arise if there is a marked lack of variation across studies with respect to feature categories. An assessment of the latter two sources of irrelevance can readily be accomplished through an examination of the frequency distributions of the various categories of each study feature. A reviewer can routinely expect the weeding out of some of his/her selected

[27] Incomplete and imprecise reporting styles affect the quantification of both study features and study outcomes. However, while strategies exist for estimating outcome magnitudes on the basis of partial information (see following section) as well as for assessing the effects of such estimation procedures on the interpretability of findings, no such compensatory or protective strategies exist with respect to study features.

study characteristics following such assessment.[28]

In sum, the quantification of study features in a meta-analysis entails, first, the selection of a set of features potentially predictive of study outcome and second, the creation of coding conventions that allow for the operationalization of such features as well as for their translation into numerical units. While the adequacy of feature selection is ultimately revealed during the lattermost stages of a meta-analysis (e.g., by the results of the regression analyses), some initial appraisal is in order before further integrative procedures are brought into play. Such an appraisal entails 1) an assessment of feature redundancy expressed in terms of intercorrelations among study features and 2) an assessment of feature irrelevance expressed in terms of frequency distributions of feature categories. The serviceability (i.e. reliability) of the coding conventions must also be appraised before further integration takes place.

Quantification of study outcomes. The third major step to a meta-analysis involves the translation of each finding of a study into a common metric.[29] For experimental studies,

[28] A point worth noting here: a thorough examination of the frequency distributions of feature characteristics functions not only to tease out irrelevance; it also allows the reviewer to take into account the effect of the shape of predictor variables on the interpretation of regression coefficients used in the meta-analysis to study the relationship between study characteristics and outcome.

each finding is expressed in terms of Cohen's "d" statistics (1977), which is known as the effect size (ES) in meta-analytic parlance. The ES is defined as the difference between the mean of an experimental and a control group divided by the within-group standard deviation which is assumed to be homogeneous. In other words, the ES defines in standard score fashion the position of an experimental group relative to that of a control. Once an ES is calculated for each finding, the distribution of such measures can be readily described in terms of its mean, variance and skew. Furthermore, an average ES can be graphically depicted in terms of the overlap of average control and average treatment distributions. If normality of distribution of scores within individual studies is assumed, the reviewer is able to compare performances between two groups in terms of percentages under the curve. Average ESs can be calculated for entire data sets as well as for various cross-sections of the data.

In the meta-analysis of correlational studies involving continuous variables, each finding is expressed in terms of a Pearson product-moment correlation coefficient. When, however, correlational studies involve independent variables that assume a dichotomous or polychotomous form, findings

[29] The basic unit on which meta-analysis is carried out is the subject of some debate. While Glass supports the use of each finding of a study as the chosen unit, a concern with violating the assumption of independence in the data has prompted Rosenthal (1979), for one, to endorse the use of the study as the unit of choice.

must be algebraically translated into ES measurement. Expressing such findings in terms of an association between several levels of an independent variable and a continuous dependent variable is unsatisfactory. Such a measure may reflect a complete set of conditions specific only to a particular study or group of studies, and is therefore not directly comparable to measures from studies which examine only a subset of those conditions (Glass, 1977).

Proponents of meta-analysis have supported the use of both the effect size and the product moment correlation coefficient for aggregation purposes on two accounts: 1) They index the strength of the relationship between dependent and independent variables and 2) Their magnitude is not affected by the size of a study's sample. This latter advantage has recently been disputed by Hedges (1981) who has argued that sample size does influence ES in two respects. The first concerns the degree of bias incorporated into ES calculations. Hedges has demonstrated that small samples yield ES statistics that overestimate their population parameters. Sample size furthermore influences the precision of ES calculations insofar as large samples generate more precise information than small samples. Fortunately, Hedges has provided meta-analysts with means of controlling for the influences of sample size. First, a correction factor which is based on the degrees of freedom of a sample's standard deviation is the suggested

treatment for sample-size bias. Second, the differential weighting of studies by sample size is the suggested treatment for the unequal precision of ES estimates. By incorporating Hedges's suggestions into the quantification of study outcomes, the reviewer is able to retain his/her confidence in the ES as the common metric of choice.

Although the ES provides a basis for aggregating data that is conceptually straightforward, problems with its execution frequently arise (Glass, 1977; Smith, Glass & Miller, 1980). First, studies often fail to contain the means and standard deviations that permit its easy calculation. Fortunately, however, various algebraic manipulations compensate for incomplete data reports. Provided that the n's of experimental groups are known, a variety of data features such as F and t ratios as well as some multiple comparison statistics can be used to calculate accurate effect sizes. When the information reported is too scanty to allow the data to be recast into exact ES terms, Glass and his colleagues provide methods for determining conservative estimates (Glass, 1979; McGaw & Glass, 1980; Smith, Glass & Miller, 1980). Of course the more complete the reports of findings, the greater the degree of precision built into these estimates. As already indicated (p. 60-61), the strong tendency on the part of primary researchers to restrict reports of nonsignificance to verbal statements impacts particularly strongly on the accuracy of a meta-analytic data base.

A second calculation problem arises when the within-group variances of experimental and control group prove to be heterogeneous. The reviewer is then forced to decide which variance is most appropriate for the ES denominator, knowing that the choice can have a pronounced impact on the magnitude of the calculation. The problem of heterogeneous variances is particularly irksome for the reviewer insofar as it does not lend itself to any definitive solution. A discussion of all the conditions and complications associated with heterogeneous variances is beyond the scope of this paper. The reader is referred to McGaw and Glass (1980) who address such matters thoroughly. The choice of an approach to heterogeneous variances most appropriate for any one substantive area is best made with reference to the data characteristics of that area.

Apart from calculation concerns, the meta-analyst is confronted with issues of statistical validity during this phase of integration. Problems emerge under two sets of conditions (Strube & Hartmann, 1983): 1) when the primary researcher has applied statistical tests that are inappropriate for the question at hand; and 2) when the primary researcher has violated the statistical assumptions underlying his/her chosen procedures. With respect to inappropriate statistical tests, it is sometimes possible to reconstruct appropriate comparisons from the data. The feasibility of such secondary analysis is of course a direct

function of the completeness of the report of findings. The impact of statistical violations on ES calculations is generally less open to amelioration, particularly when inferential statistics must be relied upon for these calculations.

Thus, as is necessary at the single-study level, the meta-analyst must be concerned when quantifying study features and outcomes with issues of methodological, conceptual, and statistical validity.

Application of inferential techniques to meta-analytic data. Once the reviewer has addressed validity-related concerns, s/he can feel free to implement the full range of statistical methods subsumed under the rubric of meta-analysis in an effort to tease out meaningful patterns from the data. The first set of procedures generally brought into play are descriptive in nature (e.g., calculation of average ESs, calculation of means, variances and skews of ES distributions). It is strongly recommended by Glass and his colleagues that the descriptive phase of a meta-analysis be followed by an inferential phase (e.g., use of ANOVA, multiple-regression techniques, etc.). The introduction of this second phase of statistical analysis gives rise to a number of concerns, the most pressing of which centers on the violation of a statistical assumption underlying inferential techniques, namely that of independence in the data. At least five different sources of violation in Smith

and Glass's (1977) integration of the psychotherapeutic outcome literature have been identified by Landman and Dawes (1982): 1) multiple measures taken from the same subject; 2) measures taken at multiple points in time from the same subjects; 3) nonindependence of scores within composite outcome measures that incorporate single measures taken from the same subjects; 4) nonindependence of studies that results when identical or overlapping samples participate in more than one article. Meta-analysts can routinely expect any or all of these sources of nonindependence to plague their efforts. Particularly vulnerable are those research areas in which samples are hard to acquire and in which experimental procedures are highly time consuming; hence the desirability and/or necessity of overusing samples and measures.

The problem of nonindependent results in meta-analysis unfortunately presents no definitive solution. Landman and Dawes (1982) describe a number of strategies that correct for the dependence that results from the use of multiple measures, none of which are without obvious drawbacks. Two of these strategies involve the use of each finding of a study as the basic unit on which meta-analysis is carried out, and two involve the use of a single "representative" outcome measure per study. Each finding can be incorporated into a meta-analysis if the reviewer 1) introduces a complex statistical procedure that adjusts for

nonindependence (e.g., the Tukey jackknife method, the generalized least squares method) or 2) is willing to accept a "simple but risky" solution in which statistical independence is falsely assumed. Glass (1977), for one, promotes the use of the latter strategy, while advising that possible distortion resulting from the statistical violation be considered in the interpretaton of findings. Reasons for his choice, apart from the benefit accrued from limiting the complexity of an already unwieldy task, are never clearly articulated by him.

The two strategies that extract a single measure from each study involve 1) the pooling of all the individual ESs within a study in order to arrive at an average ES figure for each study (e.g., Kulik, Kulik, & Cohen, 1979) or 2) the selection of a single independent measure from each study that, according to some predetermined basis, best represents the study's objective (e.g, Cooper, 1979). The use of a pooled ES has been criticized because of the associated likelihood of collapsing across study features of interest (Glass, 1977). As to the selection of the single "best" measure, there is often no clear logical or theoretical grounds on which to base such a choice (Landman & Dawes, 1982).

The choice of an 'appropriate' solution for violations of independence in a meta-analysis can best be made with reference to the nature and the degree of the violation that

exists in the literature under review. There are likely instances in which the violation is so extreme that the interpretations of any inferential statistics that emerge during meta-analysis must be made with extreme caution.

An additional less frequently addressed concern regarding the application of inferential techniques to meta-analytic data also centers on violations of an underlying assumption, in this case that of randomization. Strube and Hartmann identify two sources of this violation; first, as already indicated, meta-analytic samples are not randomly selected from a known universe of published and unpublished findings and 2) subjects of a meta-analysis (i.e., study findings) are not randomly assigned to levels of independent variables (i.e., categories of study features). Such nonrandomization places meta-analysis squarely within the boudaries of quasi-experimental and nonexperimental designs, with the latter designator likely being the more appropriate one. Distinct limitations are consequently placed on the nature of the conclusions that can be drawn. In particular, restraint must be exercised with respect to unwarranted causal conclusions. Put into more concrete terms, the reviewer must take care to avoid equating relationships between study features and outcome magnitude that emerge during a meta-analysis, with relationships that emerge from studies in which the manipulation of such features has been incorporated into the experimental design. What meta-

analytic relationships do provide are prompts for further exploration within a true experimental context. Thus, the incorporation of inferential techniques into meta-analysis serves an important hypothsis-generating function. Moreover, it affords the reviewer the opportunity to make relatively precise statements that would not otherwise be possible. In recognition of such benifits, review methodologists generally support its judicious use.

Criticisms of meta-analysis. Apart from the methodological and theoretical limitations and hurdles that have been acknowledged by its inventors and promoters, meta-analysis has been the target of a number of criticisms directed at its very foundation. Both Gallo (1978) and Rimland (1979), for instance, have challenged the interpretations that Smith and Glass (1977) have gleaned from their statistical manoeuvres. Gallo and Rimland have in turn been criticized for their misuse of statistical information by both Kalat (1980) and Glass and Smith (1980) respectively. Of more general relevance are those criticisms directed at the logic of meta-analysis. A strong objection is taken by Gallo (1978) to what he considers an inappropriate equating of a "hodgepodge" of dependent measures in meta-analysis through the translation of highly divergent terms into a standardized metric. Gallo argues that such a procedure ignores the different contexts within which measures are taken, and by so doing yields aggregated

findings that are stripped of any meaning. True, an ES of .68 on a pencil-and-paper test of manifest anxiety may have markedly different implications from an identical score on a behavioral measure of anxiety. Combining the two scores into an average ES clearly does result in a loss of information, though not necessarily in a loss of meaning. This loss of information is of course the cost to be paid whenever large bodies of findings are collapsed into more manageable and hopefully informative terms. Furthermore, while the nature of the dependent measures may be ignored in the calculation of an average ES, meta-analysis can readily study such differences by regressing "type of measure" onto study outcome.

An attack on meta-analysis by Eysenck (1978) also warrants attention if only because of its prominence in the literature. Relabelling the approach as "meta-silliness", Eysenck proceeds in a characteristically acerbic but uncharacteristically obtuse fashion. To my mind, he seems either to be unfamiliar with, to ignore, or to misunderstand the purpose of the statistical operations utilized by Glass to integrate findings. The thrust of his criticism centers on the perceived failure of meta-analysis to attend either to the quality of design or the quality of outcome measures (in particular the subjectivity of measures). When reviewing literatures, complex statistical operations are consequently wasted, according to Eysenck, on what is likely

to be valueless data. A similar criticism of meta-analysis is expressed by Mansfield and Busse (1978). As already indicated, however, Glass and his colleagues repeatedly discuss the very issue of quality variation in considerable detail in their writings as well as provide procedures for assessing and controlling for such variation through the use of regression equations. Moreover, they clearly state that the recommendation that less than perfect studies be incorporated into a review does not imply in any sense support for low standards of research (Glass, 1976, 1977; Glass & Smith, 1978; Kavall & Glass, 1981; Smith, Glass & Miller, 1980). In their response to Eysenck, Glass and Smith (1978) are thus put in the embarrassing position of having to address what by and large can be reduced to an ad hominem attack.

To meta-analyze or not to meta-analyze. Meta-analysis has, at this point, acquired enough credibility to justify its place among innovative integrative approaches. What now needs to be addressed centers less on issues of rationale and more on issues of appropriate application. Specifically, there is a need to delineate conditions under which meta-analysis becomes the integrative strategy of choice. Particularly useful to this task is a recent paper by Cook and Leviton (1980). This paper's importance rests on the fact that it not only enhances an understanding of meta-analysis, but also unexpectedly prompts a

reconsideration of traditional techniques. Cook and Leviton organize their paper around the three major criticisms of traditional reviews that have been repeatedly expressed by meta-analysts: 1) that the magnitude and direction of effects are ignored, 2) that a strong selection bias characterizes samples of studies under review, and 3) that the importance of contingency-specifying interactions are ignored. Cook and Leviton's primary objective is to evaluate the nature of these criticisms: whether they are intrinsic to traditional reviews or whether they merely reflect the poor practices of reviewers who use qualitative rather than quantitative techniques. They furthermore are concerned with comparing meta-analysis with traditional reviews in terms of these same criticisms. Cook and Leviton conclude that the disenchantment with traditional reviews can to a large measure be attributed to poor practices. The persuasiveness of their conclusion rests not only on logical considerations but also on citations of excellent qualitative reviews that do not incorporate the aforementioned inadequacies (e.g., that make extensive use of magnitude estimates, that do use information about the direction of results, that utilize exhaustive samples or that are self-corrective with respect to sample bias, and that do detect interactions).

Cook and Leviton furthermore conclude that meta-analysis is not without its unique limitations. Specifically they

focus on limitations that can be traced to: 1) faulty data bases (e.g., unequal methodological bias across studies, bias in favor of methodological and procedural considerations at the expense of conceptual considerations) and 2) faulty practices by its users (e.g., inappropriate selection procedures, inadequate coding conventions). While such limitations do not suggest that meta-analysis be abandoned, they do present substantial risks for integration. These risks are reinforced, moreover, by the fact that the limitations of meta-analysis tend to be shadowed by the apparent "objectivity", "precision" and "scientifism" of the approach. In other words, the use of rigorous quantitative techniques is as likely to have "mischievous" as beneficial consequences if readers are lured into accepting statistical information that is based on procedural invalidity. This concern is shared by Strube and Hartmann (1983) also recommend that vigilance be exercised in response to the "seductive" nature of quantification.

In light of the potential drawbacks of meta-analysis as well as the potential utility of traditional techniques, the reviewer is once again faced with a choice between qualitative and quantitative reviews. The basis for such a choice rests, according to Cook and Leviton, on the perceived complexity of the area under review. In areas in which multiple variables with multiple levels operate, the

reviewer may find it worthwhile to utilize meta-analysis, both for the sake of efficiency and to avoid "cognitive overload". Given that the reality of cognitive overload in psychology and education was the major thrust behind the initial work on quantitative reviews, it seems safe to assume that Cook and Leviton support the use of meta-analysis for many of the literatures in these disciplines.

In addition to the criterion of complexity, criteria of methodological and conceptual adequacy are strongly implicated in Cook and Leviton's discussion. These issues have been elaborated upon by Strube and Hartmann (1983) who focus as well on the extent to which statistical validity is upheld in a research domain. More specifically, both Cook and Leviton and Strube and Hartman maintain that, in cases in which extreme sorts of methodological and/or conceptual and/or statistical invalidity pervade the data base, the otherwise sound meta-analytic practice of categorizing studies with respect to quality considerations is of limited, if any, utility. Thus, at this point, meta-analysis can be recommended for complex multivariate literatures that satisfy some minimal validity standards. While deviating somewhat from Glass's original guidelines, such boundary specifications remain consistent with the original conception of meta-analysis.

To summarize, meta-analysis provides a number of advantages over traditional styles of research integration

as well as over some of the more progressive approaches. First, the use of statistical procedures on either entire literatures or, as is more often the case, on selected subsets of the literature reduces the subjective bias that operates in both the narrative and the favorite studies approach to review. Second, the use of a range of complex statistical tools allows the examination of multiple factors and of interactions among factors. In contrast, the use of simple crosstabulation procedures places limits on the number of factors that can be examined as well as on the type of statistical information that can be extracted from that data. Third, unlike reviews that define study outcome in terms of significance levels, meta-analysis translates outcome into ES terms. The strength of the relationships between study features and study outcome is thereby indexed and the bias in favor of large sample sizes reduced. Fourth, unlike the cluster approach which depends on the availability of raw data, meta-analysis can use any of a variety of inferential statistics in its calculations. As well, the statistical procedures employed in meta-analysis to study complex data patterns are more elegant than those used in cluster analysis.

In spite of these attractive features, meta-analysis is not without its inadequacies. Specifically, it provides at very best only partial solutions to the problems of heterogeneous variances and nonindependence in the data. As

well, it retains a bias in favor of statistically significant findings. Furthermore, meta-analytic procedures cannot override limitations inherent either in the data or in the practices of its users. As to the criticisms that have been advanced by its detractors, many seem to reflect inadequate comprehension of statistical properties and techniques as well as misinterpretations of the objectives of reviews. The recent critiques of meta-analysis advanced by Cook and Leviton and by Strube and Hartmann are exceptions. In addressing the potential limitations of the approach, these critics stress that quantification in and of itself does not protect against the procedural invalidity often found in traditional reviews. Having expressed this reservation, they go on to recommend that meta-analysis take its place among rigorous integrative techniques, to be used selectively and intelligently on research areas that readily induce cognitive overload.

Purpose of Investigation

The object of this research was to integrate the family interaction literature in a clear and orderly fashion. Past reviews of the literature have fared poorly on both characteristics. Whether highly discouraging or promising in tone, they have generally confronted the reader with a mass of contradictory information that is difficult, if not impossible to process. The conclusion that must be

extracted from these reviews is that family interaction research has failed to yield any evidence that differences exist between the patterns of interaction of clinical and nonclinical families. Two possible reasons for this failure can be identified: 1) both the complexity of the subject matter and the lack of a well-established methodological tradition have placed unprecedented demands on the primary researcher. Hence, studies do not often adequately tap theoretical tenets; and 2) the integrative strategies that have been relied on by past reviewers have been inappropriate for the area of research. In this thesis I began to address the lattermost source of difficulty by applying an alternative strategy that I hoped would be more appropriate for the family interaction literature. This broad objective was broken down into a number of parts: 1) the identification and collection of all published studies that compare the interaction patterns and styles of communication of clinical families with those of nonclinical controls; 2) the identification and quantification of a set of study features potentially related to the magnitude of study findings; 3) the translation of each study outcome into a standardized metric that reflects the magnitude of the statistical effect; and 4) the examination of the relationships between the selected study features and the magnitude of study outcomes by means of a variety of statistical techniques subsumed under the rubric of "meta-analysis".

In light of the breadth of this objective, I chose to restrict this portion of the research to the first and second of the designated tasks. The choice of tasks was made on the basis of my understanding of the procedural sequence underlying meta-analysis. That is, I felt that important decisions regarding tasks 3 and 4 (i.e., re the nature of the selected statistical strategies; re the appropriateness of introducing further integrative measures) could best be made with reference to the data that emerged from tasks 1 and 2. Concerns with task manageability lead me to restrict the scope of inquiry in yet another sense. While fully recognizing the importance of incorporating statistical and conceptual concerns into the selection of study features, the primary focus of this thesis was methodological and procedural. More specifically, while a small portion of the selected features have some bearing on conceptual validity (these will identified as such in the Method section to follow), no attempt was made to disaggregate the literature on the basis of theory relevance. In light of the problems family interaction researchers have encountered in developing research paradigms reflective of systems tenets and in reaching consensus about the appropriateness of the available paradigms, I felt that the task of operationalizing 'theory relevance' would constitute a thesis in and of itself. I did hope, however, that a careful reading of the literature during this phase of the research would provide some

direction for further work on conceptual issues by helping me assess the feasibility of devising an operational scheme for theory relevance. Similarly, with respect to statistical issues, I hoped that a perusal of the literature would reveal the nature and extent of the statistical violations in the family interaction data base, and at the same time provide direction for possible compensatory strategies.

The present research was designed to serve three interrelated functions. Its first function was to prepare the groundwork for further integrative steps. Such preparation entailed: 1) the selection of a set of study features potentially predictive of study outcome, coupled with some initial assessment of the usefulness of the selection, and 2) the creation of a set of coding conventions that allows for the measurement and quantification of study features, coupled with some assessment of the degree to which the conventions can be reliably applied to the family process literature.

The second function of the research was a descriptive one, involving the categorization and summarization of methodological and procedural features of interest. Such information is important for those who intend to study family process by empirical means; it helps pinpoint those aspects of research that have been both over and underused, thereby allowing for the efficient use of research resources

while protecting against the accumulation of redundant information.

A final objective of the present research was to evaluate the preparedness of the family process literature for integration, and more specifically, to evaluate the appropriateness of meta-analysis as the integrative strategy of choice. This lattermost function was as much dependent on impressionistic information as on quantitative results.

METHOD AND RESULTS

Boundaries of Review

Selection Strategy. The family interaction literature, as already indicated, is a research domain marked by weak correspondences between constructs, operational procedures and research designs, as well as by the unsystematic use of labels of constructs and operations. As a consequence, no reliable guideposts in the form of standard operations and/or research paradigms are available to direct the identification and collection of relevant studies. Faced with such a literature, the reviewer can proceed in one of two ways. First, s/he can evaluate each potential study in terms of the extent to which it incorporates systems tenets into its operational procedures and research designs (i.e., in terms of its theory-relevance and theory-appropriateness) and then use that evaluation as a basis on which to separate out "illegitimate" from "legitimate" systems research. There are a number of disadvantages associated with this strategy which are similar to those outlined by Glass (1976, 1977) in reference to the use of methodological rigour as a selection criterion. Apart from resulting in a loss of potentially valuable data, the use of theory-relevance as a selection criterion places undue reliance on the reviewer's

idiosyncratic evaluation of the incorporation of systems tenets into research efforts. In addition, it precludes any empirical exploration of the effects of theory-relevance on study outcome.

As an alternative strategy, the reviewer can include in his/her survey all studies that have been designated by the primary investigator as "family process research". There is an obvious risk associated with accepting the reviewer's appraisal of his/her work at face value. What it produces is a collection net that readily traps appropriately-labelled but theoretically irrelevant studies, while readily excluding innappropriately-labelled but theoretically-relevant ones. In spite of this risk, the label-based strategy was chosen to direct the collection of studies in this review because of the protection it offered against data loss. At the same time the following compensatory measures were introduced to help offset the strategy's inherent disadvantage. First, studies were classified in terms of a number of methodological and design features that have some bearing on theory relevance: namely, the composition of the family unit participating in research, the family unit on which measurement is taken, and the incorporation (or lack thereof) of between-subsystem comparisons in measures of family interaction. respectively. Unfortunately, no attempt was made to incorporate a direct measure of theory relevance into the

coding scheme in spite of the feature's obvious importance. As already indicated, such a procedure would constitute a thesis in itself. As an additional protection against the face-value acceptance of the designations of primary researchers, Doane's (1978a) recommended classification strategy was adopted; that is, an attempt was made to reclassify studies when necessary under designations other than those supplied by the primary researchers, in an effort to impose some consistency among constructs, operational procedures, and research designs. The details of this procedure are provided in the section on Substantive Domain to follow.

To reiterate, in the present review the collection procedure was guided by the designation of "family process" research (or an equivalent term such as "family interaction" or "family systems") as supplied by the primary investigator. This strategy is consistent with the guidelines for integration recommended by Glass and his colleagues insofar as it does not involve any a priori decisions regarding the legitimacy of individual studies. A variety of secondary selection criteria were then applied to the resulting body of literature in an attempt to mold the integrative effort into manageable proportions. These can be divided into two categories; criteria for inclusion and criteria for exclusion. They are as follows:

Criteria for Inclusion. 1) In keeping with the most recent reviews of the family process literature, the present integrative effort was restricted to studies that assess interactional data, as opposed to individual, retrospective, or attitudinal data. Studies included in the review were consequently those that utilize direct observational procedures associated with systems research, rather than clinical histories, case reports, psychodiagnostic material[30] or self-report measures. Such procedures involve the systematic observation and recording of patterns of interaction of family members as they engage in some task designed to generate interactional data.

2) Only those studies that compare the interaction patterns of a clinical category with those of a nonclinical control were considered for integration. Studies that restrict their comparisons to clinical subgroups were therefore not assessed.

---

[30] A distinction must be made between two types of "psychodynamic" studies that have been identified as family research. First there are those studies that employ projective techniques for the personality assessment of individual family members. Data from such studies are subjected to traditional projective scoring procedures. On the basis of the resulting individual protocols, global judgements concerning psychodynamic processes are then extended to the family as a whole. In contrast are those studies that employ projective procedures to stimulate interactions and that approach the resulting data in terms of interactional coding schemes. Only the latter type of projective study was incorporated into the thesis.

3) The review was in addition limited to studies that present clinical information in discrete rather than continuous terms, thereby excluding the small number of studies that correlate patterns of interaction with degree of adjustment or degree of risk. A point worth mentioning: studies that fail to include a nonclinical comparison group or that define psychopathology as a continuum, do not in and of themselves fail to qualify either as systems research or as material for meta-analysis. They were excluded from the review because the common metric deriveable from their data is not readily compared to the common metric most appropriate for the majority of available family process studies, namely ESs that express differences between a discrete clinical category and a nonclinical control in standard deviation units.

4) To be incorporated into the review, studies had to present their findings in quantified terms. A number of family process studies report the use of coding systems that lend themselves to quantification, only to then present their findings in terms of verbal transcripts followed by impressionistic interpretations. These were deemed unsuitable. One further stipulation regarding quantification: the review was restricted to those studies that report statistical information in terms that allow ESs to be calculated, namely as means and standard deviations or as any of the alternative statistical measures that are

appropriate for the solution of equations provided by Glass and his colleagues (Glass, 1980; Smith, Glass & Miller, 1980). There was one exception to this rule, made in reference to verbal reports of findings. Although it is sometimes possible to translate simple verbal reports of significance into conservative ES estimates (by setting the p value at .50 in cases of nonsignificance and at .05 in cases of significance), this thesis was restricted to outcomes expressed in numerical terms. No attempt was made to correct for incomplete reporting styles by obtaining unpublished statistical information; evidence suggests that such efforts tend to be both highly time consuming and highly unproductive (see Glass, 1977). Finally, in an effort to simplify the task at hand, no attempt was made to take advantage of instances in which the reported information allows for additional analyses and/or for analyses more appropriate to the research question of interest. The above decisions regarding outcome information were made with the understanding that the judicious use of verbal reports of findings and of secondary analyses could always be introduced at a later stage of integration.

5) One final inclusion criterion: while recognizing its effect on the representativeness of the sample of studies, concerns with both feasibility and expense dictated that the review be restricted to published material contained in English language journals.

Criteria for Exclusion. The following classes of family studies were not incorporated into the review:

1) Studies that are not concerned with a comparison of interaction patterns across diagnostic groups were excluded from consideration. Five categories of such studies were identified: a) studies that examine the validity of systems principles (e.g., the stability of interaction patterns) independent of psychopathology; b) studies that develop or refine interactional tasks or coding procedures within the context of pilot research; c) studies that compare interactional tasks, settings or data collection procedures independent of diagnostic considerations; d) studies that compare the efficacy of family therapy with other forms of clinical intervention, and e) the large body of studies that examine parent-child (generally mother-child) interactions from a child development perspective. This lattermost exclusion warrants some comment, particularly in light of the fact that previous critics have strongly recommended the incorporation of the child development and family interaction literatures into a single review (see Fox, 1978). While recognizing that the two fields share a concern with interactions, and indeed focus on variables that overlap to some degree, the methodologies with which they examine family interaction are too distinct to allow for any tidy integration of findings. While systems research is most often designed to test specific hypotheses

about dysfunctional interaction within the family, child development research is designed to increase understanding of the process of child development through the accumulation of normative data (Krahn, 1980). My decision to exclude interaction studies that emerge out of the child development perspective was thus based on methodological considerations; it in no way suggests that I oppose the pleas for greater cross-discipline fertilization that have been voiced by critics of both the child development and family interaction fields of study (e.g., Fox, 1978; Olson, 1970; Riskin & Faunce, 1972).

2) The body of research that examines communication deviance by means of what Wynne and Singer term the "Predictive Method" was not assessed. While such research is concerned with interaction patterns, the dependent measure is the extent to which blind ratings of interactions of various diagnostic categories are correct beyond chance expectancy, rather than the extent to which interaction patterns differ across diagnostic categories.

3) The few studies that analyze interaction patterns by sequential procedures were set aside because their outcomes (in the form of time lag correlations) are not readily compared with those derived from the more common aggregate procedures. This exclusion is unfortunate because of the obvious importance of this line of research (see Introduction, pp. 16-17).

In sum, to qualify for the integration, empirical studies had to 1) be labelled in a manner suggestive of the researcher's identification with a family systems perspective; 2) compare the interaction patterns of clinical families with those of nonclinical families; 3) collect data by means of direct observational procedures; and 4) contain statistical information needed for the computation of effect sizes.

Search Procedures. A combination of computer and manual search procedures was used to access the relevant literature. A free text search of the Psychological Abstracts data base executed by computer yielded disappointing results, both in terms of an abundance of irrelevant references and in terms of its failure to generate known, let alone unknown, data sources. In contrast, the reference lists of previous reviews, in particular those of Riskin and Faunce, Jacob, and Doane, served as major data sources. A manual search of the Psychological Abstracts accessed a number of studies published after the cutoff date of Doane's 1978 review (i.e., 1977).

Seventy-seven references that met the specified content-related inclusion criteria (as opposed to statistically related) were retrieved through these search procedures. Subsequent screening of these studies in terms of the adequacy of their statistical reports, reduced the initial

set to a final set of 46 studies. A list of these studies is contained in Appendix A. To be included in the final set, a study needed to contain only one usable outcome measure; the presence of nonusable reports of findings (however many) did not serve as grounds for exclusion.

Thus, the family interaction data base was substantially reduced on statistical grounds. While such a reduction is not unusual in the meta-analytic literature, it was particularly unwelcomed here because of its anticipated effect on the capacity to meet some of the research objectives associated with later stages of integration; specifically to put some closure on the Jacob/Doane debate and, at the same time, to compare the traditional review strategy utilized by Jacob and Doane with a meta-analytic one. In order to meet these objectives, the data bases of the three relevant reviews would have to overlap to a reasonable degree. While the operation of similar content-related inclusion criteria in all three reviews insured some degree of overlap, the idiosyncratic operation of statistical criteria in the present review imposed an upper limit on such overlap. A comparison of relevant bibliographies reveals that 52.6% of the 57 studies reviewed by Jacob, and 47.1% of the 51 studies reviewed by Doane have been incorporated into the present integrative effort. While far less than desirable, such percentages, I feel, allow for valid (though perhaps tentative) comparisons of both review conclusions and methodologies.

Selection of Relevant Study Features. The final set of 45 studies yielded 809 units appropriate for use in the present research (i.e., outome measures that lend themselves to ES calculations). Each of these units was characterized according to a variety of features that potentially influence the magnitude of study findings. These features were chosen to serve three major integrative functions, only the first of which was of immediate concern: 1) to provide descriptive information by allowing for the categorization and summarization of the characteristics of the family process literature; 2) to provide dimensions around which to group studies during the calculation of average ESs performed on various cross-sections of the literature; and 3) to serve as predictors in the regression equations utilized during the latter stages of the meta-analysis.

Guidelines for the selection of study characteristics took a number of forms. First, a preliminary survey of the literature suggested distinctions both in theoretical foci and methodologies adopted by primary researchers. The use of review papers helped clarify these distinctions. In particular, Mishler and Waxler's (1965) excellent review of principle theories of family interaction helped delineate the range of substantive foci that come under investigation. Review papers that emphasize methodological considerations (Fontana, 1966; Frank, 1965; Jacob, 1975; Jones, 1973; Rabkin, 1965) were also helpful in suggesting ways to

distinguish among studies on the basis of data collection procedures and validity and reliability standards. Riskin and Faunce's (1972) comprehensive methodological critique proved to be the essential reference for the task of selecting features for investigation.

A couple of points should be made regarding the list of study features. First, the reader will no doubt be struck by the large number of features that entered into consideration. While this large number reflects the multivariate nature of family process, it presents obvious problems for the stability of the statistical operations involved in a meta-analysis. This is particularly true in light of the fact that my survey of the family process literature yielded at best adequate rather than plentiful data with which to perform such operations. My reasons for selecting a large set of features for initial consideration were primarily practical. To begin with, it was difficult to make intelligent a priori predictions as to which and how many of the selected features would prove to be either redundant, nonvariable, or nonscorable -- attributes that would result in their exclusion from the feature set. Moreover, the relatively unknown and decidedly tortuous nature of the research territory made it difficult to separate out essential from nonessential features on either logical or theoretical grounds. Thus, rather than risk being left with either too few features to afford an

adequate exploration of the family interaction literature, or with a nonessential set of features, I decided to err, if at all, in the direction of excess.

A second point concerning the list of study features: no attempt was made to select features that are mutually exclusive. Two kinds of interdependency can be identified in the feature set: first, there is methodological interdependency caused by research objectives that dictate methodological choices across a variety of study characteristics. A primary concern with experimental precision, for instance, is often associated with the joint use of experimental tasks that are high in structure, and outcome measures that are low in inference. Thus, some overlap between the features Task Structure and Inference Level could be expected. Second, a certain amount of interdependency was built into the coding scheme. For example, a 'nonscorable' code on the feature Sex of the Focal Child automatically results in a 'no' code on the feature Sex Specificity. Similarly, a 'noninferential' code on the feature Inference Level results in a 'nonapplicable' code on the feature Presence of Reliability Coefficient. An important statistical need was thus introduced; namely to assess interdependency among study features. Such assessment would be particularly important during later stages of integration, when concerns with statistical efficiency and power dictate that redundant features be

eliminated from inferential analyses, and when information about interdependency guides both the choice of regression solutions and the interpretation of resulting semi-partials.

Relevant Study Features. The selected study features were divided into four groups: those relating to 1) methodological tools, 2) sample composition, 3) methodological rigour and 4) research focus. A description of each feature, the rationale for its inclusion, and the data on the distribution of its categories will be presented shortly, along with the conventions that were used to assist in its coding and the reliability data pertaining to such conventions. Thus, contrary to standard format, Method and Result sections are combined in the present manuscript. This alternative format was chosen for the sake of improved readability.

One characteristic of the coding scheme warrants a comment here; namely its oftentimes simplistic form. In particular, dichotomous variables have been constructed in cases in which logical considerations might well support the construction of more finely graduated scales. Demands for such simplification came from two sources: 1) a need to reduce the complexity of a decidedly unwielding task and 2) a concern with the statistical stability of data generated from the coding scheme. Since such stability is in part dependent on a sufficient number of instances of each level of the study features, it was important to limit the number

of levels associated with each feature. A copy of the coding form and of the coding appendix can be found in Appendices B and C respectively.

Training in the Use of the Coding Scheme. Training in the use of the coding scheme involved its independent application by two coders to four family interaction studies. These studies were eliminated from the present review on statistical grounds. They were chosen for training purposes because they incorporate features typical of the interactional literature, as well as a number of features idiosyncratic enough to broaden the boundaries of training.

Independent application of the coding scheme was followed by meetings between the two coders, during which requests for clarification by the second coder were answered, and during which negotiation over divergent judgements took place. Guideline revisions were often made on the basis of these negotiations. Thus, construction of the coding scheme and training in its application unfortunately overlapped to some extent. Concerns with time and expense introduced additional limitations into the procedures. To begin with, criterion protocols were at no point established for training purposes. As a consequence, training sessions were less systematic than desirable. In addition, retraining sessions were not introduced at any point to protect against observer drift.

Reliability of Coding Conventions. A subset of 25 studies was used to assess the degree to which the judgements of the two independent coders converged. This subset is identified in Appendix A. For the purposes of reliability assessment study features were differentiated on the basis of two considerations: 1) the nature of their numerical properties and 2) whether they are study-wide or measure-specific. The former distinction determined the choice of statistical operations, while the latter determined the number of coding occurrences that entered into the reliability calculations. With respect to numerical distinctions, features with underlying distributional properties were assessed in terms of both Pearson product-moment correlation coefficients (r) and interrater agreement ratios. Features with nominal properties were assessed only in terms of interrater agreement ratios. For the purposes of such ratios, the judgements of one coder on any particular feature were charted against the judgements of the second coder on that same feature. This procedure generated matrices whose constituent cells reflect both convergent and divergent judgements. Such matrices were particularly useful because they allowed me to readily compute simple agreement ratios, by dividing the combined frequency of convergent-judgement cells by the total cell frequency. Moreover, they allowed me to readily track down sources of interrater nonreliability by examining the row and column values of

heavily populated divergent-judgement cells. An example of one such matrix is presented in Appendix D for clarification purposes.

With respect to the study-wide/measure-specific distinction, it emerged out of a concern over the fact that for some features, a single judgement pertains to all outcome measures contained in any one study, while for other features independent judgements have to be made for each outcome. This distinction is important insofar as a true assessment of the reliability of a coding scheme should be based on the number of actual judgements that are made, rather than the number of data points that emerge from such judgements. Consequently, an attempt was made to separate features into study-wide and measure-specific categories. This separation, however, proved to be somewhat problematic. There are, to begin with, a number of features that can be classified as study-wide for most but not all of the studies under review. For example, for the many studies that rely on a single experimental task to generate interactional data, judgements regarding the features Task Structure, Emotional Tone and Task Characteristics remain constant across all outcome measures. However, for the small number of studies that utilize more than one interactional task, judgements pertaining to the task-related features may vary across outcomes derived from the distinct tasks. There are in addition, a number of features (e.g., Demographic

Comparability, Subject Expectancies, Setting) for which a single judgement pertains to all outcomes contained in studies that examine only one clinical sample in comparison to a nonclinical control. When, however, studies examine more than one clinical category, and when moreover, these categories differ with respect to the features of concern, coding judgements may vary across outcomes derived from the different clinical samples. One further problem entered into the study-wide/measure-specific distinction. Since only 25 studies were coded for reliability purposes, there was limited data available for the study-wide assessment.

In response to the above difficulties, a decision was made to assess the reliability of features that are clearly measure-specific with reference to the entire data set (n=410). Features that can by and large be classified as study-wide but that at times function as measure-specific were, on the other hand, assessed both with reference to 'actual' judgements (n=25) and with reference to the entire data set. This double assessment was introduced partially to protect against a small N assessment by providing a large N comparison, and partially to arrive at some estimation of the degree to which one could expect nonreliability to impact on the actual data base. In addition, it generated reliability boundaries for those features that defy easy classification with respect to the study-wide/measure-specific distinction. Lists of study-wide and measure-specific features are presented in Appendix E.

Some comments regarding the treatment of missing data in the reliability assessment is in order here. Data was, in some instances, designated as missing when a feature was either not reported in the study or when, for reasons specified in the coding conventions, information pertaining to the feature was judged to be nonscorable. The resulting 'unknown' and 'nonscorable' ratings functioned as missing data during the calculation of Pearson _r_s where their inclusion would have disrupted the distributional quality of the features of concern. Such ratings did not, however, function as missing data during the calculation of interrater agreement ratios, where they were readily incorporated into the existing nominal framework.

There was a second circumstance under which data was designated as missing, namely when only one of the two coders judged a particular study outcome as appropriate for ES purposes. (This resulted in the exclusion of an entire case from the data set.) Of the 529 outcome measures coded by at least one coder, only 410 were coded by both coders. Thus there was more than a slight measure of disagreement between coders regarding the appropriateness of the available statistical information. The extent of such disagreement could doubtless have been reduced had ES guidelines been more fully specified. While both coders relied on a common set of guidelines contained in Glass's 1977 document (Integrating Findings: The Meta-Analysis of

Research), instructions pertaining to the use of the document were left open-ended (i.e., With the help of the guidelines, select all outcome measures that you feel can likely be translated into ES terms.) Moreover, no attempt was made to equate the two coders on statistical acumen. Thus, the many subtleties of the statistical guidelines may not have been equally accessible to both coders.[31] Finally, the fact that coders were not provided with written definitions or instructions concerning 'redundant' measures added to the idiosyncratic usage of outcome information.[32] Data designated as missing on statistical grounds was not incorporated into reliability calculations. Had it been included, the resulting coefficients would have been difficult to interpret, insofar as they would have simultaneously reflected the adequacy of the coding conventions and of the statistical guidelines.

---

[31] I have every reason to believe that the statistical grasp of the second rater, Dr. Hans Beihl, exceeds that of my own.

[32] The second rater, for instance, consistently excluded the 'inferior' member of pairs of measures contained in some of the studies on the grounds of redundancy, with inferior/superior being defined in terms of whether or not covariate procedures were applied to correct for the presence of unmatched samples. I, on the other hand, included both 'inferior' and 'superior' members of such pairs in the data base. This latter inclusion pattern should have been built into the research guidelines because it provides data needed to test for the effect of sample comparability via meta-analytic procedures. A number of such decisions regarding measure redundancy only emerged with any clarity during the course of coding, as I became more familiar with the subtle characteristics of the literature.

In sum, the presence of missing data in the present study can be attributed to three sources: 1) the incomplete reporting styles of primary researchers; 2) the nature of the coding conventions; and 3) the use of insufficiently detailed guidelines concerning the appropriateness of statistical information. Missing data was not incorporated into any of the reliability computations.

Features Relating to Methodological Tools

In selecting a method for measuring family interaction, the researcher must confront choices relating to experimental tasks, outcome measures, and research settings. Such choices are based mainly on professional judgements concerning the capacity of methodological features to yield data that validly reflects the theoretical premises and substantive interests of the researcher. Empirical foundations for such judgements are at this point largely undeveloped.

A. Tasks Used to Generate Interactional Data

a) Task Structure. The nature of the experimental tasks utilized in family process research can be described in terms of a number of dimensions. First, tasks vary in their degree of structure. At one extreme are highly circumscribed tasks that limit the form and range of subjects' interchange. An often cited example of a highly

structured task is the coalition game designed by Haley (1963), in which the communication of family members is limited to button pushing. The pattern recognition task devised by Reiss (1967), the map task reported by Solvberg and Blakhar (1975) and the twenty question task developed by Mosher and Hornby (1976) are similarly high in structure. Researchers advocate the use of highly structured tasks for a number of reasons. Of greatest importance is the fact that such tasks are thought to increase the probability of occurrence of theory-relevant behavior (Bell, 1964). In addition, their capacity to readily generate low-inference measures has been viewed with favour. In spite of such advantages, the choice of highly structured tasks is not made without reservations. Their sensitivity to clinical phenomena as well as their capacity to reflect the real-life exchanges of family members has been repeatedly questioned (e.g., Riskin & Faunce, 1972; Smith, 1971), even by those who endorse their use. Haley (1963), for instance, emphasizes their potential to either obscure or distort the interactional process. He nonetheless has preferred the risks involved in testing within a "restricted communication setting" to those involved in the collection of more "naturalistic" forms of data. Given that even elementary bits of behavior such as who-speaks-to-whom are open to multiple interpretation, Haley argues in favour of strategies that rigourously reduce the complexities of family exchange, such as those that simplify the data in

advance through the assignment of highly circumscribed tasks.

Researchers who prefer semi-structured or loosely structured tasks do so because of the presumed capacity of such tasks to elicit relatively natural forms of exchange. In making such a choice, the researcher may be willing to sacrifice some experimental control in order to capture some of the richness of family interaction. (Researchers always have the option when using loosely structured tasks of simplifying observations post hoc by focusing on limited aspects of interaction and/or by utilizing coding schemes that consist of low-inference measures.) While examples of loosely structured tasks can be found in the family process literature (e.g., Hansen, 1968; Lennard, 1961), there is little if any use of purely naturalistic observations. The experimental tasks most readily associated with family process research are the semi-structured ones. The various forms of the Resolution of Difference Techniques (RDT) initially developed by Strodbeck (1951), the Plan Something Together task from the longer standardized interview designed by Watzlawick (1966), and the Create a Story around TAT Cards used by Winter, Ferreira and Olson (1966) and Stabenau, Tupin and Pollen (1965), among others, are all examples of semi-structured tasks.

In the present research studies were classified in terms of the degree of structure of their interactional tasks according to the following set of conventions:

High Structure. Tasks are coded as high structure when a variety of restrictions placed on family participants result in a type of interaction quite unlike that which occurs in the family's everyday life, both in terms of its form and content. Restrictions may be placed on the physical positioning or movement of family members by, for instance, prescribing seating arrangements. Or the number of channels of communication generally available to human interactants may be limited. Visual communication, for instance, may be blocked by the appropriate placement of visual barriers or by having family members communicate from separate rooms by microphone and receiver. Alternatively, verbal communication may be prohibited, or restricted to only one aspect of speech such as questions or yes/no responses. Strict control over family interactions may also take the form of directives concerning the number of family members able to participate at any one point during the experimental proceedings or the order in which they participate.

Medium Structure. Tasks that are coded as medium structure are characterized by a specific set of instructions that direct the content of family exchanges and by a specific set of stimuli that provide a basis around which to focus exchanges. Such tasks do not, however, apply any restrictions to the form of family exchange.

Low Structure. Tasks are coded as low structure when they leave research subjects free to direct both the form and content of family interactions. If experimental stimuli are introduced (e.g., toys, juice and cookies, blackboard), they serve largely to "naturalize" the experimental context rather than to provide a basis around which to direct exchanges.

The coding of the feature Task Structure was reasonably straightforward. The problems that did arise can be linked to the few idiosyncracies in the literature that went unnoticed during the development of the coding conventions. Scoring problems were created, for instance, by those studies that utilize a number of interactional tasks that differ with respect to structure while failing to analyze the data in terms of task. One obvious solution to such a problem would be to rely on a 'nonscorable' code. No such

alternative had, however, been incorporated into the coding scheme, although coders were free at any point to create such a category to correct for the inadequacies of the available conventions.

A second problem arose with respect to those studies that rely on videotape to collect interactional material. To fully benefit from the use of videotape it is necessary that all family members remain within the camera's lens range at all times. This can only be accomplished by placing restrictions on the physical movements of study participants. Insofar as physical restrictiveness is incorporated into the definition of 'high structure', the use of videotape as a data collection devise automatically results in a 'high structure' classification. However, in cases in which the use of videotape is accompanied by the use of tasks that in all respects (save physical restrictiveness) meet the definition of 'low' or 'medium' task structure, strict adherence to the coding conventions seems inappropriate. Thus, at this point, two revisions of the coding conventions can be recommended: 1) A nonscorable category should be added to the feature Task Structure and 2) The 'high structure' definition should be amended to take into account the unique features of videotape usage.

The frequency count of Task Structure yielded the following information. Of the 809 coding occurrences, there are 101 instances (12.5%) of high structure, 594 (73.4%) of

medium structure and 111 (13.7%) of low structure. Thus, much of the family process literature can be characterized in terms of medium structure. This imbalanced distribution suggests one of two possibilities. It may point to the failure of the coding conventions to capture relevant distinctions in the literature. Or, more likely, it may point to an overuse of a methodological strategy, and to an incomplete exploration of a methodological feature that has received considerable attention in family process critiques. In either case, the usefulness on Task Structure for the purposes on a meta-analysis of the collected literature is somewhat restricted by the uneven representation of its categories. An examination of the feature's effect on study outcome remains viable, however, particularly in light of the fact that it was coded with a reasonable level of reliability. For the measure-specific calculations a Pearson r of .84 and an interrater agreement rate of 91.2% emerged. The study-wide calculation yielded an agreement rate of 84%.

b) Emotional Eone of Tasks. Apart from degree of structure, interactional tasks can be differentiated in terms of their emotional tone. Efforts to modulate emotional tone reflect a concern with the capacity of tasks to elicit uninhibited and meaningful exchange. While this criterion has been discussed largely in reference to the RDT and its variations, it can be applied to other interactional

tasks as well. Some investigators have favoured the use of emotionally charged RDT items because of their presumed capacity to highlight areas of both friction and harmony that might otherwise go undetected under less evocative stimuli (see Haley, 1967). Others, however, have criticized the use of "controversial" items on the gounds that they call into play family taboos and defenses, thereby inhibiting a free flow of communication. Ferreira and Winter (1968), for instance, support the use of ostensibly "neutral" items to encourage maximal freedom of expression. These investigators are furthermore of the opinion that truly neutral issues rarely exist in family life, and that highly revealing exchanges occur around the most trivial of subjects.

Given a preference for a particular emotional level, an investigator must anticipate difficulties in creating an interactional situation that assumes the desired tone. Without knowledge of the family's unique past and present, s/he might easily label as neutral an issue that serves as a major battleground on which relationship-related struggles are fought. Alternatively, an ostensibly controversial issue may go uncontested in an otherwise disharmonious family environment. To circumvent this problem, Goldstein, Judd, Rodnick, Alkire and Gould (1968) have suggested that highly relevant idiosyncratic issues be designed for each family that participates in a study. While such a strategy

may be useful for the purpose of clinical assessment, it poses obvious problems for cross-family comparisons made within the context of research.

Many investigators who use particular sets of RDT items fail to state the reasons for their choice. While this may simply suggest an incomplete reporting style, it more likely reflects an assumption that is widely held by family interactionists; namely that family behavior is so basic and pervasive that it is likely to emerge regardless of the particular content of an exchange (see Riskin & Faunce, 1972). While this assumption has been questioned (Fontana, 1966) only Haley (1967) has subjected it to empirical assessment, at least in regards to the RDT. Using a variety of low-inference measures (i.e., calculations based on frequencies of audible sounds and who-speaks-to-whom sequences), Haley concluded that families respond similarly to neutral and controversial questionnaire items.

There is another possible explanation for the cavalier attitude regarding RDT item selection. It may reflect the fact that investigators who use the RDT insist that the resulting conversations remain within the context of "everyday" exchange, regardless of the choice of items. Thus, built into the procedure are distinct limits that preclude the exploration of the full range of emotional tone in family process research.

Studies included in the present research were classified in terms of their emotional tone according to the following set of conventions:

Highly Emotional. Tasks are coded as highly emotional if they engage research families in interactions around issues that are profoundly important in family life: issues concerning parental control, children's compliance, spousal roles and the like. Such tasks can also be identified in terms of their emphasis on interpersonal and affective issues relating to family life as opposed to instrumental issues. Moreover, the relevance of these issues to the research families is explicitely acknowledged in task instructions.

Moderately Emotional. Tasks are coded as moderately emotional if, like highly emotional tasks, they engage research families in interactions around interpersonal and affective concerns. Moderately emotional tasks can be distinguished from highly emotional tasks insofar as their instructions allow respondents to distance themselves from the issues in hand. That is, the personal relevance of such issues to the research families is neither directly acknowledged in task instructions nor directly reflected in task stimuli. Examples of moderately emotional tasks are ones that use projective stimuli depicting family experiences, and tasks that require subjects to comment on norms and values relating to family life rather than on their actual experiences.

Emotionally Neutral. Tasks are coded as emotionally neutral if they focus on experiences and issues that are not generally thought of as central to family life (e.g., questionnaire items that request subjects' opinions on car colors, dinner menus and the like, games involving identification of visual designs). Of course labelling a task as emotionally neutral does not imply that the experience of participating in the task is neutral. Regardless of the nature of the task stimuli, family members' reactions to being observed and evaluated may be highly emotional.

Problems encountered during the coding of Emotional Tone can be traced to three souces: 1) incomplete reporting styles of primary researchers; 2) insufficiently detailed coding conventions and 3) flaws in the logic underlying the

choice of conventions. Reporting styles are incomplete in two respects. First, descriptions of task instructions are often omitted, making it difficult, in the case of the RDT, to determine whether families are asked to focus on actual experiences or on abstract values. As a consequence the high/moderate distinction becomes problematic. Second, RDT items are often described very generally in terms of "family problem situations". In such cases it is difficult to determine whether the term 'problem' is meant to serve as a general label for RDT items or as a specific label for emotionally charged items. While the second coder seems to have chosen the latter interpretation, my choice (in retrospect possibly a poor one) was to avoid the interpretive dilemma by creating a 'nonscorable' coding category.

With respect to the comprehensiveness of the coding conventions, it was only during the course of coding that I realized that, in cases of low task structure, it is difficult to decide on an appropriate category for Emotional Tone. Insofar as the primary researcher leaves the experimental context by and large open-ended under such circumstances, emotional tone remains undefined. Two possible responses to this situation come to mind: 1) to decide that Emotional Tone becomes irrelevant under such conditions and that a nonscorable code is most appropriate and 2) to amend the 'neutral' definition to include open-

ended or unspecified conditions. The fact that I chose the first alternative while the second coder chose the latter of course lowered interrater reliability. There is a second sense in which coding conventions are insufficiently detailed. As in the case of Task Structure, coding problems arose with studies that collapse the data across a number of interactional tasks that differ with respect to Emotional Tone. Again, the use of a 'nonscorable' code seems appropriate under such circumstances.

With respect to the logic underlying the coding distinctions, exposure to the literature during coding led me to question the legitimacy of the family value/family experience distinction incorporated into the definitions of 'high' and 'moderate' emotional tone. Apart from the fact that the scanty reporting of task instructions often makes it difficult to detect this distinction, it seems that whether a family relates to a task in terms of actual experiences or in terms of abstract values depends not so much on the instructional focus as on the overall experimental context. Moreover, what remains an abstract issue for one clinical category may readily be responded to in light of actual experience by another. Thus, it may be more important to classify tasks in terms of their relevance to research families than with reference to the real experience/abstract value distinction.

The above impressions that emerged during the process of coding can be translated into the following recommendations: 1) Explicit instructions for coding Emotional Tone under conditions of low task structure should be incorporated into the coding conventions to protect against idiosyncratic responding. In this regard expansion of the neutral category to include instances in which emotional tone is left open-ended might be preferred over the use of a nonscorable category. 2) Some reconsideration of the real experience/abstract value distinction is in order. Modifications of the high and moderate categories would likely follow such reconsideration, along with the introduction of an addition feature(s) that measures the relevance of the interactional task to research participants.

Results of the frequency and interrater reliability analyses support these impressions. Of the 809 coding occurrences, there are 197 instances (24.4%) of highly emotional, 205 instances (25.3%) of moderately emotional, and 198 instances (24.5%) of emotionally neutral tasks, along with 209 (25.8%) nonscorable codes. This distribution suggests that, within the boundaries imposed by the current family process methodology, all levels of the feature Emotional Tone have received adequate empirical representation. The distribution at the same time suggests the need to look for ways of reducing usage of the

nonscorable code through an elaboration of the coding scheme.

Problems with the use of the nonscorable category are also reflected in the reliability data. While a comfortable Pearson r of .94 emerged, interrater agreement rates using both the measure-specific and the study-wide data base are only of moderate size, namely, 60.5% and 68% respectively. The discrepancy between these figures warrants some explanation. The high r value can be accounted for by the fact that the nonscorable category was excluded from the calculation in order to retain the distributional quality of the variable under examination. The moderate agreement ratios, on the other hand, can be accounted for in terms of disagreements between coders regarding the use of the nonscorable code. Examination of the matrix designed to display both convergent and divergent judgements in fact reveals that all divergent cells involve the nonscorable code.

c) Categories of Task Stimuli. Along with structure and emotional tone, interactional tasks were categorized in terms of their stimuli characteristics. Four categories were incorporated into the coding scheme. One category consists of those tasks that utilize verbal stimuli to generate verbally expressed differences (e.g., RDT and its variations), while another consists of tasks that utilize projective stimuli to elicit such differences (e.g., joint

construction of TAT stories and joint interpretation of Rorschach cards). A third category consists of more general verbal exercises (e.g., Plan Something Together task), while a fourth consists of efforts to engage family members in a game situation (e.g., Strauss's SIMFAM, Haley's coalition game).

As with the feature Emotional Tone, problems were caused for the coding of Task Characteristics by 1) studies that collapse data across a number of interactional tasks that differ with respect to the feature and 2) studies that utilize a task that is low in structure, thereby leaving the characteristics of the task undefined or open-ended. The fact that I responded to these situations by relying on a nonscorable code while the second coder stayed within the boundaries of the available categories, led me to anticipate problems with reliability. The recommended solution is similar to that presented for the feature Emotional Tone; namely that explicit instructions for the coding of Task Characteristics under multiple-task and low-structure conditions should be incorporated into the coding conventions via the use of 'nonscorable' and 'open-ended' categories.

My own efforts to apply the coding conventions for Task Characteristics lead me, in addition, to recommend a general fleshing out of the coding guidelines. Unlike those features whose characteristics are fully specified in the

Coding Appendix, all information pertaining to Task Characteristics is contained on the actual coding form. More details, particularly with respect to the categories 'general verbal exercises' and 'game situations', would facilitate the coding task.

An additional impression emerged during the course of coding that ran contrary to expectations; namely that, in spite of talk of abundant methodological variability by critics of family process research, the majority of researchers have restricted their attention to one of two procedures: the RDT and the Consensual Projective. Thus, two of the features categories (i.e., verbal stimuli, and projective stimuli designed to elicit verbally expressed differences) are highly uniform in composition.

An examination of both the frequency and the interrater reliability data support the above impressions. The frequency count of Task Characteristics yielded the following distribution. Of the 809 coding occurrences, there are 297 instances (36.7%) of verbal stimuli and 193 instances (23.9%) of projective stimuli designed to elicit verbally expressed differences, along with 25 instances (3.1%) of general verbal exercises, 111 instances (13.7%) of game situations and 183 (22.6%) nonscorable instances. A number of comments can be made about this distribution, apart from the obvious overuse of the nonscorable category. First, the marked underrepresentation of the general verbal

exercise category, reflecting either insufficiently detailed conventions or, more likely, the scanty use of the category by family researchers, poses a problem for confident interpretation of the effect of the category on study outcome during later stages of integration. Nonetheless, adequate representation by the remaining three categories suggests that Task Characteristics remains a viable feature for integrative purposes. Interest in it is, however, limited by the boundaries imposed by the restrictive exploration of methodology by family researchers. In other words, the fact that two interactional procedures dominate the literature has implications for the breadth of the conclusions that can be drawn regarding the effect of Task Characteristics on family interaction.

With respect to the feature's reliability, moderate agreement rates emerged both from the study-wide (72%) and measure-specific (74.4%) computations in spite of the poorly detailed coding conventions. Examination of the appropriate judgement matrices revealed that, as anticipated, all of the well-populated divergent cells involve nonscorable judgements, save one that involves general verbal exercise judgements.

B. Outcome Measures

a) Inference Level of Outcome Measures. The form of data generated by interactional tasks provides a number of dimensions on which to differentiate individual outcomes. First, outcomes can be categorized on the basis of the level of inference of their measurement technique. As already implied in the section pertaining to task structure, reliance on noninferential measures tends to reflect the researcher's concern with experimental rigor, while reliance on inferential forms of data suggests a concern with the pithiness of interactional material. While a number of fine-grained distinctions can no doubt be made with respect to inference level, a simple two-category classification was adopted for the purposes of the present research. Its conventions were as follows:

Noninferential. Outcome measures that require the observer to simply conduct a frequency count of an objective event without making any decisions as to its quality are coded as 'noninferential' (e.g., interruptions, statements received, talking time, number of correspondences between individual members' questionnaire responses and group responses, number of adjectives incorporated into a story).

Inferential. Outcome measures that require the observer not only to determine the occurrence of an event but also to evaluate its nature subjectively are coded as 'inferential' (e.g., agreements, yielding, empathy, degree of vagueness and confusion in joint family TAT).

The coding of the feature Inference Level posed few problems. Relevant information was always reported and the coding requirements did little to tax the decision-making capacities of the coders. The healthy Pearson r of .91 and

interrater agreement rate of 96.1%, along with the absence of nonscorable judgements, attest to this. Grounds for indecision arose only with respect to a few process measures that incorporate both inferential and noninferential measurement phases. An example of such a measure is one employed by (Farina, 1960) that requires the coder first to evaluate the leniency of father's and mother's responses to offspring prior and subsequent to a joint interactional session by means of a rating scale (inferential phase), and second, to compare the degree to which spouses yield to one another following the joint session, with yielding being defined in terms of pre-post differences in the spouses respective leniency ratings (noninferential phase). While both coders responded appropriately to the subjective component of this measure (and others like it) by selecting the inferential code, explicit instructions pertaining to 'mixed' measures should be incorporated into the coding conventions to protect against idiosyncratic responding.

The frequency count of Inference Level indicates a somewhat surprising preference for inferential measures on the part of family researchers (58.3%), coupled with adequate representation by noninferential measures (41.7%).

Thus, both frequency and reliability data suggests that Inference Level is an appropriate candidate for further analyses. Whether its coding conventions need to be elaborated upon to allow for finer distinctions among measures requires some further consideration.

b) Process Versus Product Measures. Outcome measures can in addition be described as dealing either with the process or the product of an interaction. This distinction may be important insofar as the two types of measures may yield very different sorts of information. Though it can be taken for granted that the process of an interaction has an effect on the product, the form of the effect may not be readily predictable. For example, a family interaction that is characterized by frequent disagreements need not result in a group product whose form directly mirrors those disagreements. The product of such an interaction may be one that can best be described as a nonproduct due to the failure on the part of family members to reach any agreement, or one that can be described as dictatorial insofar as it reflects the opinions and decisions of only one member of the family, or one that can be described as democratic in that it closely reflects the opinions and decisions of the majority of family members.

In addition to incorporating the above rationale into the coding guidelines, an attempt was made to further clarify the process/product distinction by way of the following examples:

Examples of process measures: agreements, disagreements, yielding, acknowledgments, interruptions, statements received, talking time, button pushes.

Examples of product measures: number of correspondences between individual members' questionnaire responses and group responses, number of adjectives incorporated into a story, degree of vagueness and confusion in family TAT.

In retrospect, these examples were misleading. Although researchers have indeed tended to apply different sets of measures to process and product material, the process/product distinction logically rests not so much on the choice of measure as on the locus of the interactional information. It also became apparent during coding that the distinction is only straightforward under specific conditions, namely, when task instructions explicitely call for the creation of a joint family product, or when the outcome of a game situation is defined in terms of wins and loses.

In spite of problems inherent both in the coding guidelines and in the feature's conceptualization, Process/Product yielded a respectable 89.3% interrater agreement rate. With regard to frequency information, a strong preference for process measures emerged (82.6% of the 809 coding occurrances), coupled with very moderate usage of product measures (17.3%). Such a distribution reflects the emphasis on ongoing interactions and the rejection of 'static' indications of family exchange that has held sway in family research during the last two decades. What the distribution suggests is the need to explore more fully sources of information apart from the ongoing interactions of family members. A glaring inadequacy of family research is the absence of any empirical exploration of the

relationship between the process of family interaction and the product of such interaction.

c) Global Rating Measures. Outcome measures can also be differentiated in terms of their underlying measurement operations. Two broad categories of operations can be identified in family process research: 1) estimation of the quantity of an interactional attribute of interest in a specified unit of analysis (generally the entire interaction) in terms of a position on a rating scale and 2) evaluation of a specified unit of analysis (generally an utterance, speech, or interaction) in terms of the presence or absence of an interactional attribute of interest. This distinction was incorporated into the present research under the feature Global Rating in acknowledgement of the distinct requirements the two forms of measurement impose on the coder of behavioral data (see Wiggins, 1973). Given that rating scales used in family process research are all inferential, a decision was made to exclude noninferential measures from consideration when examining this feature. Thus, three feature categories were incorporated into the coding scheme: yes-rating, no-rating and nonapplicable (for use with noninferential measures).

While I experienced very few problems coding Global Rating, an unsatisfactory 64.4% interrater agreement figure emerged from the data. This figure prompted close inspection of the appropriate judgement matrix. Two heavily

populated cells were revealed, the first accounting for 15.9% of the 410 coding occurrences, the second for 14.4%. The first of these divergent cells crosstabulates my nonapplicable judgements with the no-rating judgements of the second coder. The composition of this cell suggests that the second coder may have relied on the no-rating code rather than on the nonapplicable code for many noninferential measures. While correct in essence, such judgements are not in keeping with coding instructions. The presence of the second well-populated divergent cell suggests that more fully specified coding definitions should be incorporated into the Coding Appendix. Like Task Characteristics, all information pertaining to Global Rating is contained on the actual coding form.

The frequency count of Global Rating yielded the following distribution. Of the 475 relevant coding occurrences, there are 410 (86.3%) no-rating judgements and 65 (13.7%) yes-rating ones. This scanty use of rating scales by family researchers is consistent with their general devaluation in psychological research. Some family process researchers have, however, recently begun to question the current status of rating scales. Riskin (1983), for one, has suggested that, along with providing an efficient measurement alternative, rating scales may be more appropriate for the coding of family interactions than some of the more molecular measurement devices.

C. Experimental Setting

Family process investigators have conducted research in a variety of experimental settings: the home, the hospital-based laboratory and the laboratory based in a research center. These settings differ in their similarity to the family's everyday environment and in their restrictiveness, and probably generate different expectations on the part of those being observed. Moreover, the effects of setting may be more pronounced for some clinical categories than for others. The fact that family researchers do not usually present reasons for their choice of setting, and at times fail to control for this factor across experimental groups, suggests that practical considerations override concerns with such possible differential effects. Although no attempt has been made to explore possible differences in the effect of setting on clinical and nonclinical families, some effort has been made to examine the effect of setting on nonclinical families (i.e., Bronfenbrenner, 1977; Krahn, 1980; O'Rourke, 1963). In keeping with these efforts, setting was incorporated into the present research as a study feature. Thus, the groundwork was prepared for examining during later stages of integration the feature's overall impact on family interaction as well as its differential impact on the interactions of various clinical categories (in terms of the statistical interaction between the features Setting and Diagnostic Category). In addition

to the three categories home, laboratory-based-in-a-research-center, and hospital laboratory, unknown and nonscorable categories were introduced for studies that fail to present relevant information or that that fail to control for setting across diagnostic groups.

The coding of Setting was troublesome only in one respect. Because the name of an institution is often the only information provided, it was difficult (without actually inquiring into the nature of each specified institution) to determine whether the experiment was conducted in a research-center laboratory or a hospital laboratory. This problem is clearly reflected in the appropriate judgement matrices. Of the 410 measure-specific and the 25 study-wide judgements, 195 and 10 respectively are divergent with respect to the two classes of laboratories. A decision was consequently made to collapse the two categories into a single 'laboratory' setting for the remaining analyses. This resulted in an improvement in the interrater agreement figures from 47.1% to 94.6% for the measure-specific analysis and from 56% to 96% for the study-wide analysis.

The frequency count of Setting yielded 27 instances (3.3%) of home setting, 624 (77.2%) instances of laboratory setting, 24 (3%) unknown codes and 134 (16.6%) instances of nonscorable or noncontrolled settings. Of the 675 scorable coding occurrences, 92.4% are accounted for by laboratory

setting judgements. This distribution points to a number of problems in the family interaction literature. First, the 16.6% nonscorable figure suggests there is room for the improved methodological control of setting. Second, in spite of the widely voiced concern about tapping into 'everyday' family life, researchers have been very reluctant to leave the conveniences of the laboratory and enter into the family home. Finally, both the lack of variability coupled with the insufficiently detailed reporting of Setting suggests that the feature may not be particularly useful for later stages of integration, except perhaps with reference to methodological control.

D. Naturalistic/Controlled Continuum

Many of the issues relating to task assignment, setting, and outcome measures that I have discussed can be conceptualized as aspects of one common issue, namely the relative merits of naturalistic versus controlled experimentation. While this broad issue is of concern in all branches of psychological research, it particularly comes to the fore in an area that attempts to describe something as complex as interactional data. Consequently, the field is peppered with promotions of various points on the continuum, with the majority of critics favouring some appropriate balance between the precision that is derived from controlled experimentation and the richness that is

derived from more naturalistic approaches (e.g., Lytton, 1971; Olson, 1970; Riskin & Faunce, 1972; Watzlawick & Weakland, 1977). These critics do not, however, provide any clear guidelines for attaining such a balance. Thus, the choice of a particular level of methodological structure must often be made with an uncomfortable awareness of its inadequacies (see Haley (1962) for the clearest articulation of this dilemma). Apart from classifying studies on the basis of task structure, inference level of outcome measure, and experimental setting, an effort was made to locate studies on the naturalistic/controlled continuum in terms of these three characteristics. While a multi-level variable would no doubt best suit this objective, concerns with statistical power and task manageability prompted me to use the following dichotomous coding scheme. Studies were classified as 'controlled' if they received at least two of the following ratings: High Structure, Noninferential, and Laboratory Setting (either based in hospital or research center). Studies were classified as 'naturalistic' if they received at least two of the following ratings: Medium/Low Structure, Inferential, and Home Setting. Given the coarseness of the coding scheme, it is worth pointing out that the labels 'controlled' and 'naturalistic' are more suggestive of general direction on the naturalistic/controlled continuum than of actual position.

An alternative perspective to the issue of degree of methodological control has been suggested by Hughes and Haynes (1978). These critics maintain that the issue is not that of choosing one end of the control continuum over the other, but of choosing a control strategy that is most effective for the research question of interest. If, for instance, parental compliance is being investigated, an unstructured free play situation might be less useful than a structured situation in which parents are instructed to encourage their child to engage in an unreinforcing task. If, on the other hand, the flexibility of interaction patterns is the research focus, a loosely structured situation that allows for the emergence of a variety of behaviors might be preferred over a rigourously structured one.

The issue of the goodness-of-fit between methodological control and research focus can be approached in a number of ways. First, (and perhaps most desirable) studies can be coded for fit on the basis of logical considerations, and the effect of this feature on study outcome then examined via correlational procedures. However, one could expect to encounter considerable difficulty in devising a reliable coding scheme for this feature. As an alternative strategy, some information about the form of control most appropriate for specific substantive areas (rather than for individual studies) can be empirically acquired via regression

procedures in either of two ways. First, separate regression equations can be constructed for each substantive area, with degree of methodological control (i.e., controlled vs. naturalistic) serving as a predictor in each equation. Second, interaction vectors can be constructed from coded information pertaining to substantive focus and methodological control, to be used as predictors in an overall regression equation. The incorporation of the feature Naturalistic/Controlled into the present research prepares the way for such exploration.

Insofar as reliability of Naturalistic/Controlled is jointly dependent on the reliability of its three constituent features, problems with coding, above and beyond those associated with the three features, were not anticipated. Some additional error was expected to enter in during the process of converting three codes into a single code. Reasonable 89.8% and 88% interrater agreement figures emerged from the measure-specific and study-wide calculations respectively, along with an r of, .79. Moreover, the distribution of the feature's categories is reasonably well-balanced. Of the 809 coding occurrences, there are 327 instances (40.4%) of controlled and 447 (55.3%) of naturalistic experimentation. Thus, on the basis of the reliability and frequency data, the feature Naturalistic/Controlled appears to be a suitable candidate for integrative purposes. Whether the combining of the

various aspects of degree of conrol has any predictive advantage is another issue, to be determined by appropriate meta-analytic procedures.

To summarize, frequency and reliability data by and large support the viability of the set of features relating to methodological tools for later stages of integration. Imbalances in the distributions of feature categories was, however, identified as a source of concern, specifically with reference to Task Structure, Process/Product, Global Rating, and Setting. It was suggested that such imbalances are more reflective of the nature of the research domain than of problems with the coding conventions. Only in the case of Setting was there serious questioning of the utility of the feature for further analyses. With regard to measurement reliability, the coding conventions yielded moderate to high levels, with interrater agreement figures ranging from 60.5% - 91.2% for measure-specific calculations and from 68% to 96% for study-wide calculations. Pearson _r_s range in value from .79 to .94. For all features, save Setting and Naturalistic/Controlled, recommendations for convention revisions were made, with the objective of reducing sources of nonreliability inherent in the coding scheme. A summarization of the characteristics of features relating to methodological tools is presented in Table 1.

Table 1

Features Relating to Methodological Tools

Feature	Problematic Aspect	Measure-Specific Reliability	Study-Wide Reliability	Pearson *r*	Coding Revisions Recommended	Retain/ Exclude
Task Structure	Overrepresentation by High Structure category.	91.2%	84%	.84	Yes	Retain
Emotional Tone	Overuse of nonscorable category. Reliability problems linked to nonscorable category.	60.5%	68%	.94	Yes	Retain
Task Character-istics	Overuse of nonscorable category.	74.4%	72%	N/A	Yes	Retain
Inference Level	----------	96.1%	N/A	.91	Yes	Retain
Process/ Product	Overrepresentation by Process category.	89.3%	N/A	N/A	Yes	Retain
Global Rating	Overrepresentation by No-rating category. Reliability concerns linked to inadequately detailed coding conventions.	64.4%	N/A	N/A	Yes	Retain
Setting	Overrepresentation by Laboratory Setting category.	94.6%	96%	N/A	No	Exclude
Naturalistic/ Controlled	----------	89.6%	88%	.79	No	Retain

N/A nonapplicable

Features Relating to Sample Composition

A. Diagnostic Category

Since investigators usually label their clinical samples as either schizophrenic or nonschizophrenic disturbed, one would expect that studies could readily be classified according to diagnostic category. A number of problems, however, preclude such easy organization of the literature. Leaving aside the issue of the appropriateness of the traditional psychiatric nosology as a basis for classifying families, a reviewer can expect to confront reports of sampling procedures and sample composition that are too casual for classification purposes (Riskin & Faunce, 1972). More of a hindrance is the fact that the composition of many of the samples utilized in family interaction research is too heterogeneous to allow anything but the application of the broadest of classification schemes. The issue of diagnostic heterogeneity will be discussed in some detail in the section on methodological rigour to follow. At this point the following broad diagnostic divisions were introduced as an initial step to classifying families according to the composition of their clinical samples: schizophrenia, nonpsychotic disturbance, and delinquency. Samples which include representatives from more than one of the three broad diagnostic divisions were coded as mixed diagnoses with the following exception; when information on sample proportions clearly indicates that one diagnostic category predominates (i.e., at least 90% representation),

the study was classified in terms of the predominant category. Each outcome was coded for this feature according to the diagnostic labels provided by the primary researcher in his/her description of clinical samples. Thus, the validity of the assessment procedures used to assign clinical subjects to diagnostic groups did not enter into consideration.

The coding of Diagnostic Category posed no particular problem, as evidenced by the 100% and 90% agreement rates that emerged respectively from the study-wide and measure-specific calculations. An examination of the measure-specific judgement matrix failed to reveal any particular pattern to the divergent judgements, leading one to attribute minor nonreliability to coder error and/or the inadequate reporting styles of primary researchers.

The frequency count of Diagnostic Category yielded 381 instances (47.1%) of schizophrenic samples, 300 instances (37.1%) of nonpsychotic disturbed, 88 instances (10.9%) of delinquent and 40 (4.9%) instances of mixed samples. Two aspects of this distribution need to be commented upon. First, the fact that less than 5% of clinical samples are of mixed composition indicates that family researchers have, on the whole, been at least minimally attuned to the possibility that differences in clinical expression can be accounted for in terms of differences in interactional patterns and communication styles. Second, while the welcomed broadening of the early systems focus to include

samples other than schizophrenics is clearly evident, it falls somewhat short with respect to delinquent families.

B. Composition of Family Unit Participating in Research

Although all family systems proponents maintain that the family is primary in the study and treatment of psychopathology, there is no strict agreement as to its composition. While it is common to equate 'family' with 'nuclear family', the definition has often been extended to include a variety of social units. Family therapists, in particular, have been flexible in their approach to family composition. For instance, some therapists have treated single adults and couples without children along with their families of origin (e.g., Framo, 1975; Olson, 1976; Whitaker, 1975). Others have invited large social networks including members of the extended family, friends and working colleagues into therapy, since they view all members of the network as participants in the dysfunctional interactional patterns in which the symptom-bearing individual is engaged (Attneave, 1976; Speck, 1967). Systems researchers have, in comparison, been conservative in their approach to family composition. While their objective has been to examine family units that are representative of family systems within which psychopathology operates, concern with the complexity of interactional networks has often placed limits on the size of the unit utilized in research.

The family unit that has received the greatest critical attention is the triad, generally consisting of two parents and a patient-child. This unit is thought to adequately represent the larger family system. In addition, its use insures that the complexity of interactional data is kept within reasonable limits, a concern that is especially relevant to those who analyze interactions sequentially (Krahn, 1980).

The family quadrad, generally consisting of two parents, a patient-child and a well sibling has also been studied, though less extensively than the triad. Apart from changing the quantitative aspect of interactions, the addition of one more member to the family triad allows for some qualitative comparisons (Riskin & Faunce, 1972). First, comparisons can be made between different pairs of triads within the same family. Researchers have particularly been interested in the extent to which parents respond differently to the patient-child and the well-sibling (e.g., Mishler & Waxler, 1966, 1968; Sharon, 1966). Second, comparisons can be made between triadic and quadradic interactions within the same family. It is possible that the overall tone of family interaction changes with the composition of the participating unit. Riskin & Faunce (1972) suggest that the addition of the well-sibling to the standard family triad could have an ameliorating effect on triadic interchange. It is equally possible that the presence of the well-sibling

exacerbates family disturbances. To my knowledge there are no empirical comparisons of triadic and quadradic interaction in the family literature. If such research was to yield differential effects, it could no longer be assumed that the standard triadic unit is representative of the larger family system. This is not to say that information concerning triadic interaction would not continue to be valuable in its own right. An additional unit to consider in family research is one that is larger than the quadrad. Primarily because of the practical complexities involved, this unit has received very limited attention by family researchers (Riskin & Faunce, 1972).

The family literature also refers to units smaller than the family triad: namely the mother-child dyad, the parental-dyad, and the individual family member. Given that many family process theorists have rejected the study of both the individual and the dyad outside the context of the broader family system, and given that it is, in part, precisely that family context that distinguishes the systems approach from other models of psychopathology, there is some question as to whether such studies qualify as systems research. In response to such concerns, Jacob (1975) has restricted his survey to studies that assess the interactions of both parents and at least one child. A number of counterarguments can, however, be brought to bear against such a decision. First, it seems somewhat arbitrary

to label dyadic interaction as nonsystems material and triadic interaction as systems material. Except in the case of the three-member family that has no ties with families of origin or with any other kin, the family triad is as much an incomplete subsystem as the family dyad. By the same token, spouses who have no children constitute as complete a family system as do spouses with children. Second, the assumption that the products of family interaction carry over to and are reflected in the interactions of family subsystems (including subsystems consisting of individual family members) is accepted by many scholars who identify themselves with systems thought. Acknowledging an interplay between individual and interactional levels of analysis, such scholars accept information on any subsystem of the family as valuable in the study of psychopathology. A number of critics can be identified with this perspective insofar as they incorporate into their reviews all available family units (e.g., Doane, 1978a, Fontana, 1966, Frank, 1965; Olson, 1970; Riskin & Faunce, 1972).[33]

---

[33] What may in fact be reflected in the debate over the appropriate size and composition of the family unit participating in research is the manner in which systems psychologists respond to discrepancies between the traditional intrapsychic perspective and the interactional perspective. Two broad categories of theorists can be identified in this regard. On the one hand there are those theorists who maintain that there are no inherent contradictions between intrapsychic and interactional perspectives, and that an understanding of the family is enhanced through the use of both models (e.g., Ackerman, Boszormenyi-Nagi, Framo, Lidz). On the other hand are those theorists who maintain that the contradictions are such that the interactional perspective can only be viewed in terms of a paradigmatic

Decisions relating to the appropriate composition of the family unit utilized in family interaction research can be arrived at in one of two ways: 1) on the basis of theoretical considerations and 2) on the basis of empirical examination. It is the latter strategy that was adopted in the present review. Thus, size and composition of the family unit participating in research did not serve as selection criteria. Rather all available units were incorporated into the review: namely the family quadrad, the family triad, the parental dyad, the mother-child dyad and the individual family member. Studies were then classified in terms of the composition of their research units, thereby allowing for the examination of the effect of Family Composition on study outcome during later integrative stages. This stategy, it is worth noting, is in keeping with the spirit of meta-analysis as outlined by Glass and his colleagues insofar as it protects against undue a priori elimination of studies.

One final note regarding family units smaller than the triad: Apart from concerns with insuring that research participants respond within a family context, there are no doubt historical reasons that can account for the degree of acceptance afforded to such units. For example, the mother-

---

shift in the Kuhnian sense of the word (e.g., Haley, Watzlawick, Weakland, Wynne). I suggest that adherents of the latter position would be more likely than those of the former to restrict their attention to studies that utilize units at least as large as the triad.

child dyad appears to be particularly poorly accepted as a target for systems analysis, possibly because research on that dyad has long been the domain of the child-development field. The parental dyad has, in comparison, received a notably wider acceptance. Such acceptance likely reflects the etiological primacy afforded to the parental relationship by numerous family interaction theorists (see Mishler & Waxler, 1965, Olson, 1970, Satir, 1977).

Problems that arose during the coding of Family Composition can be traced to the failure on my part to anticipate a number of idiosyncracies in the literature, and to incorporate into the coding guidelines the appropriate responses. The available definitions are not, for example, appropriate for those anomolous cases in which a family member other than the child is the identified patient, or in which the identity of the focal individual is either not reported or not consistent across families. In such cases, I chose the 'other' code by default. In spite of the incomplete coding conventions, reasonable 89.3% and 88% interrater agreement rates emerged from the measure-specific and study-wide computation respectively. No particular pattern regarding divergent judgements is evident.

The distribution of categories of Family Composition is in line with expectations. Of the 809 coding occurrences, there are 488 instances (60.3%) of family triads, 108 (13.3%) of tetrads, 80 (9.9%) of spousal dyads, and an absence of mother-child dyads. The 132 instances (16.3%) of

'other' judgements are difficult to interpret, given that I unfortunately applied the 'other' code both to the aforementioned anomolous family compositions as well as to compositions larger than the tetrad. The relatively heavy representation by this category suggests that improvements in the coding conventions are needed.

C. Family Unit on Which Measurement is Taken

Studies that utilize a particular family unit often measure the interactions of various subsets of that unit. For instance, experimental situations in which the family triad engages in an interactional task may generate outcome data on the interactions of the spousal dyad, the two parent-child dyads and/or the individual. As suggested in the previous section on family composition, there is some debate about the degree to which such submeasures accurately reflect systems principles. Particularly targeted for attack have been measures taken on the individual. Such measures have been described by critics as providing the basis for a quantitative rather than qualitative shift in perspective from nonsystems thought (see Jackson, 1977; Rabkin, 1965). Rather than excluding such measures, the feature Unit of Measurement was incorporated into the present research. Apart from allowing for the examination of its overall effect on family interaction, the inclusion of Unit of Measurement created the necessary framework for testing the interaction between the features Family

Composition and Unit of Measurement during later stages of integration. This interaction is important insofar as it can shed some light on the effect of family context on the behavior of various family subsystems. Such an analysis could, for instance, generate comparisons between dyadic exchanges that occur in the presence of a third family member and those that occur in isolation.

Though in appearance a straightforward feature, more than the usual number of problems were encountered during its coding. These can be traced to two major sources. First, coding guidelines are insufficiently detailed. In light of the feature's apparent straightforward quality, I had decided against the use of the Coding Appendix for details in addition to those contained in the coding form -- in retrospect,an unwise decision. Second, the feature categories are not designed to readily accommodate those atypical aggregate measures that manage to attend, to some degree, to interactional sequences. For example, for cases in which the researcher specifies the direction of an exchange, by differentiating between father-to-child and child-to-father interactions, for instance, coding alternatives are less than ideal. By choosing the father-child dyad as the appropriate unit of measurement for both interactions, the coder muddles the distinction between operations that jointly reflect the behavior of father and child and those that reflect the behavior of either father-with-reference-to-child or child-with-reference-to-father.

On the other hand, by choosing the individual-father code for the first interaction and the individual-child code for the second, the coder ignores the available interactional information. In a similar vein, coding guidelines are inadequate for those measures that have an interactional quality built into them independent of the family unit to which they are applied. For example, the measure 'successful intrusions' simultaneously reflects the attributes of the sender and receiver of a communication. In the absence of sufficiently detailed guidelines, the coder is left uncertain whether to select a Unit of Measurement category with respect to the receiver, the sender, or the interacting dyad. Thus, revisions in the coding conventions are clearly called for. These could either take form of an expanded set of categories or an expanded set of instructions, particularly in reference to anomolous measures and operations.

There is one additional problem associated with the feature Unit of Measurement that came to light during the course of coding. Resulting from the inappropriate translation of systems tenets into measurement operations, the problems impacts not so much on measurement reliability as on the evaluation of the effect of Unit of Measurement on study outcome during later stages of the meta-analysis. Specifically, a number of measures chosen by family researchers because of their presumed capacity to reflect

the family-as-a-whole, are as likely to be sensitive to a single family member (generally the identified patient) as to the entire family unit. In such cases, information regarding the family unit may, by and large, be equivalent to that provided by the focal individual. For example, measures of 'evenness of participation' are as likely to be of large magnitude when a single family member (e.g., the schizophrenic member) contributes very little to family exchange as when uneven participation is a more general family characteristic. In response to this concern, it might be worthwhile to complement the feature Unit of Measurement with an additional feature that differentiates between family measures that are as likely to reflect the behavior of the focal individual as of the entire unit and those that more confidently reflect the shared authorship of all family members.

The number of problems encountered during the course of coding Unit of Measurement is mirrored in the relatively low 66.3% interrater agreement. Examination of the feature's judgement matrix reveals no particular pattern to the divergent judgements,most of which are distributed throughout the matrix. The one exception is the presence of a single well-populated cell accounting for 10.9% of coding occurrences, that crosstabulates my individual-child judgements with the triad ratings of the second coder. The reason for this particular coding pattern is not, at this point, apparent to me.

With regard to the frequency distribution, family triad is the most represented unit of measurement accounting for 18.5% of the 809 outcome measures. Measures taken on the individual are also relatively frequent, with father, mother and child categories accounting for 15.8%, 16.6% and 16.4% respectively. Three other categories account for greater than 3% of the distribution. Measures taken on the family tetrad account for 4.2%, while those taken on the spousal dyad account for 11.2%. Finally, the 'other' category was relied upon for 8.2% of all coding occurrences. As with the feature Family Composition, the use of the 'other' code is difficult to interpret insofar as I applied it both to units of measurement greater than the tetrad and to anomolous measures and operations that I had difficulty shoehorning into the available codes. The remaining six categories each contribute less than 3% to the frequency distribution, with percentages ranging from .5 to 2.3.

D. Developmental Stage of the Family

Systems theorists have conceptualized family life as a series of stages (Hill, 1964, 1965), each of which can be characterized in terms of unique developmental milestones (e.g., marriage, birth of the first offspring, onset of adolescence, onset of "empty-nest" period) and in terms of unique demands associated with these milestones. Members of a family respond to such demands by continually modifying and adjusting the ways in which they interact with one

another. The degree and appropriateness of such adjustments are thought to determine to a large extent how well the family system meets the needs of its members (Haley, 1973). There is furthermore the understanding that certain stages in the family life cycle place particularly heavy demands on the family system and as such require that relatively extensive adjustments be made. Such stages may be particularly vulnerable to disturbances in the family insofar as they are associated with an increased likelihood that needs for adjustment go unmet. The period of offspring-adolescence has often been earmarked as one such stage (see Mishler & Waxler, 1965). A number of empirical studies have supported theoretical expectations regarding the effects of developmental stage on patterns of interaction (Ferreira, 1963a, 1963b, Ferreira & Winter, 1968a, Jacob, 1974, Murrell & Stachowiuk, 1967). In keeping with these empirical efforts, the feature Developmental Stage was incorporated into the present research. Certain coding difficulties were, however, anticipated. Specifically, I anticipated that the identification of developmental stage would often be problematic, particularly for families with more than one offspring. In such cases it is not clear whether developmental stage should be defined in terms of the age of the focal child or that of the sibling, or in terms of some averaging of these ages. If an averaging procedure is chosen, one must decide whether to restrict the calculations to offspring that participate in

research or to include in the calculation the ages of nonparticipating offspring as well. When the ages of family offspring span a wide range, any averaging procedure seems inappropriate. In light of such ambiguities, I decided to define the developmental stages of research families solely in terms of the age of the focal child. The following age categories were used as a basis for such classification: 2-4, 5-9, 10-13, 14-17 and over 17 years of age. The boundaries of these age categories are loosely based on education levels (e.g., preschool, primary school) and culturally-defined developmental distinctions (e.g., preteen, teenager, young adult). Two methodological conditions had to be met in order for a study to be classified according to developmental stage: 1) research samples had to be homogeneous with respect to this feature, with homogeneity defined either in absolute terms or in terms of at least 90% representation by one age level; 2) clinical and nonclinical groups had to be equivalent with respect to developmental stage.

It soon became apparent during the process of coding that this first condition was rarely met. Indeed, an examination of the frequency distribution reveals that samples that are homogeneous with respect to the designated age-levels account for only 8.1% of the coding occurrences. Moreover, the interrater agreement rates of 100% and 99.9% for study-wide and measure-specific calculations respectively indicate

that both coders were in agreement regarding this appraisal. Thus, it is obvious that concerns with the impact of the family's developmental stage on patterns of interaction expressed by systems theorists have not been integrated into methodological practice. Consequently, the viability of the feature Developmental Stage for a meta-analysis of the collected literature is highly questionable. Before excluding the feature from consideration, however, it might be worthwhile to first reexamine it in terms of a less rigorous criterion for homogeneity.

E. Sex of the Focal Child

The effect of the sex of the child on the parent-child relationship has long been examined on both theoretical and empirical levels in the child development field. Family systems theorists have, in contrast, been negligent in their attention to this factor. There have, however, been a number of empirical studies on the effect of the sex of the focal child on family process. These studies have yielded positive findings, though only in reference to certain diagnostic categories (Check, 1964, 1965; Ferreira, 1963; Ferreira & Winter, 1965; Hetherington, Stouwie, & Ridberg, 1971). In the present review, studies were coded on the sex of the focal child to provide the basis for the examination of the feature's overall effect on family interaction as well as of the interaction between sex and diagnostic status. Two methodological criteria had to be met in order

for studies to be incorporated into these procedures: 1) research samples had to be homogeneous with respect to the sex of the focal child, with homogeneity being defined in terms of at least 90% representation by one sex; 2) clinical and nonclinical groups had to equivalent with respect to this feature. The fact that only 48.7% of the available 809 coding units met the first of these criteria underscores the need for a greater methodological consideration of the impact of the sex of the focal child on patterns of interaction.

Of the 449 scorable coding occurrences, 68.8% are male and 18.9% female, with the remaining 12.5% falling either under the unknown or the nonapplicable classification. Thus, clinical families with male offspring are represented more than three times as frequently as those with female offspring. This can partially be accounted for in terms of the difference in the average age of onset of schizophrenia in male and female populations, and in terms of the effect of such a difference on sample availability. While males are likely to receive a diagnosis of schizophrenia during the period of early-to-mid adolescence, females are more likely to experience onset in early adulthood after they have left home (Seeman, 1983). Given that the likelihood of getting the entire family to participate in research is greater when all members live together than when members reside in separate households, and when the focal individual

is a minor rather than an adult, the imbalance in the sex distribution becomes understandable.

With regard to the reliability of Sex of the Focal Child, a 79.3% interrater agreement figure emerged from the data -- an unexpectedly modest figure in light of the straightforward nature of the feature. Examination of the appropriate judgement matrix reveals that the single well-populated divergent cell accounting for 15.6% of all coding occurrences crosstabulates my 'nonscorable' judgements with the 'unknown' judgements of the second coder. No explanation for this coding pattern is immediately apparent.

To summarize, of the five features relating to sample composition, three cause some concern for futher analyses; two because of overrepresentation by the nonscorable category, and one because of a relatively low level of measurement reliability. With respect to the former concern, nonscorable judgements on over 90% of coding occurrences pertaining to Development Stage and on over 50% of coding occurrences pertaining to Sex of the Focal Child, point to the failure on the part of family interaction researchers to adequately take into account the potential influence of the two demographic factors when determining the composition of their research samples. A suggestion was made to exclude Development Stage from the feature set on the grounds of insufficient variability of feature categories. Sex of the Focal Child was, on the other hand,

retained, because its scorable categories (i.e., Female and Male) provide sufficient data to allow for the examination of the impact of the feature on study outcome during later stages of the meta-analysis.

Measurement reliability is a concern with respect to the feature Unit of Measurement, where a relatively low interrater agreement figure of 66.3% emerged from the data. It was suggested that this reliability problem is as much attributable to the nature of the research domain as to the nature of the coding guidelines. A number of guideline revisions were, however, recommended, both to clarify the coding task and to introduce some relevant conceptual distinctions into the guidelines. Reliability of the remaining four features relating to sample composition is high, with interrater agreement figures ranging in value from 79.3% to 99.9% for the measure-specific calculations and from 88% to 100% for the study-wide calculations. A summarization of the characteristics of features relating to sample composition is contained in Table 2.

Features Relating to Methodological Rigour

In keeping with Glass's guidelines (see chapter on Integrative Techniques), methodological rigour was not used as a selection criterion in the present research.

Table 2

Features Relating to Sample Composition

Feature	Problematic Aspect	Measure-Specific Reliability	Study-Wide Reliability	Pearson *r*	Coding Revisions Recommended	Retain/ Exclude
Diagnostic Category	Underrepresentation by Delinquent category.	90%	100%	N/A	No	Retain
Family Composition	Overuse of other category.	89.3%	88%	N/A	Yes	Retain
Unit of Measurement	Reliability problem linked to nature of research domain as much as to inadequate coding guidelines.	66.3	N/A	N/A	Yes	Retain
Developmental Stage	Overrepresentation by nonscorable category.	99.9%	100%	N/A	No	Exclude
Sex of Focal Child	Overrepresentation by nonscorable category.	79.3%	N/A	N/A	No	Retain

N/A nonapplicable

Rather,studies were coded on how well they adhere to some elementary methodological standards, in order to prepare the groundwork for the empirical examination of the degree to which family process data is sensitive to methodological violations.

A. Demographic Comparability

The demographic comparability of clinical and nonclinical families has been focused on as a particularly important methodological feature in the field of family research (Rabkin, 1965; Fontana, 1966; Riskin & Faunce, 1972; Jacob, 1975). A wide range of demographic variables are thought to be related to patterns of interaction. (See Fox (1978) and Jacob (1975) for comprehensive lists.) Family process researchers have acknowledged the potential importance of such variables either by restricting their samples to one level of the variable of concern or by matching heterogeneous samples on demographic criteria. Attempts to empirically examine the impact of demographic variables have, however, been scarce, suggesting however, been scarce, suggesting obvious limits to this acknowledgement.

As an initial attempt to address the issue of demographic comparability, studies included in the present research were differentiated according to whether they ignore demographic considerations by relying on unmatched samples, or control for demographic indices through the use of appropriate

methodological or statistical strategies. Only the most commonly controlled for indices were incorporated into this assessment. These are as follows: a) socioeconomic status (SES), described in terms of the education and/or occupation of the parent(s) or in terms of some composite score based on these two features; b) family composition, described in terms of the birth order of the focal child, the number of siblings, and/or the size of the family (The latter two pieces of information are often equivalent insofar as most researchers have restricted themselves to the assessment of the intact nuclear family.); c) family composition described in terms of the sex of the focal child and d) developmental stage, most often described in terms of the age of the focal child and more occasionally in terms of the ages of parents and/or of all participating offspring.

The following rating scale was applied to each of the above demographic indices:

Score of 1: no information <u>or</u> vague statement suggesting likelihood of sample comparability on index <u>or</u> statement of statistical support for nonequivalence.

Score of 2: statement that matching procedures were used, without any verbal or statistical information in support of the effectiveness of the procedure.

Score of 3: verbal statement indicating that statistical results support the successful contruction of matched samples <u>or</u> statistical data in the form of <u>t</u> on <u>F</u> ratios or x^2 values that supports such construction <u>or</u> restricted use of only one value of the demographic variable <u>or</u> use of statistical procedures that control for unmatched variables (e.g., ANCOVA).

Thus, individual studies received ratings on each of the four demographic indices that ranged in value from 1 to 3. When information on more than one feature of a demographic index was reported (e.g., both education and occupational level of parent), only one such feature received a rating.[34] The resulting four ratings were combined into a composite score of Demographic Comparability ranging in value from 4 to 12, with the higher scores representing greater degrees of comparability.

A number of problems with the rating scheme for Demographic Comparability were apparent during its development. First, the equating of progressively more adequate reports of demographic information with progressively more adequate control for demographic features may be questioned. While I am reasonably comfortable with equating "no information on a demographic feature" with "failure to control for that feature" (a rating of 1), I am less comfortable with the distinction between verbal and numerical indicators of comparability (rating 2 vs. 3). In a sense, the rating scheme as much reflects a concern with comparability as actual comparability. The second problem

[34] The decision to attend to only one aspect of a demographic feature in cases of multiple aspects was made with the understanding that it would, at times, result in the coder ignoring explicit evidence of noncomparability regarding one aspect of a demographic feature, that is reported alongside evidence of comparability on a second aspect of that feature. As a consequence, high scores for Demographic Comparability were occasionally assigned to studies that fail to control for one or more potentially influential demographic factors.

concerns the use of a composite score. Specifically, the assumption of the equivalent importance of the four features that enter into it (reflected in equal weights) was not empirically substantiated. Demands for task manageability and for the economic use of study features during later stages of the meta-analysis prompted the construction of a composite score.

A number of additional problems only became apparent during the course of coding. First, the wording of the scoring criteria, particularly with respect to scores 1 and 2, lacks clarity. Specifically, it is difficult to distinguish between the criterion of a "vague statement suggesting the likelihood of sample comparability" (score of 1), and that of a "statement that matching procedures were used, in the absence of verbal or statistical information in support of the effectiveness of the procedures" (score of 2). Scoring criteria are, moreover, insufficiently detailed. In particular, instructions are not provided for those situations in which information pertaining to the specified demographic feature is presented without reference (either verbal or numerical) to sample comparability. For instance, information pertaining to the sex of the focal child, often contained in reports of sample composition rather than in reports of demographics, generally allows for the coding of Demographic Comparability. However, in the absence of explicit instructions regarding the use of such information, the coder is left uncertain whether to conduct

his/her own significance test on the raw data, or to accept the reporting of the information as evidence of a moderate level of concern with demographic comparability (score of 2), or to adhere to the letter of the coding guidelines with a resulting judgement of noncomparability (score of 1).

Coding guidelines are incomplete in yet another respect in that the distinction between the researcher's sensitivity to the issue of demographic control and the actual degree of sample comparability is not sufficiently articulated. This is particularly likely to be a source of coding indecision for studies in which the researcher demonstrates obvious sensitivity to demographics by evaluating in a post hoc fashion the impact of those demographic variables that went uncontrolled on study findings. Insofar as such an evaluation does not, in and of itself, either alter or compensate for the composition of unmatched research samples, outcomes from such studies are appropriately coded as noncomparable. Instructions to this effect should have been incorporated into the coding conventions.

Some of the problems with Demographic Comparability just discussed are reflected in the frequency and reliability data. These results will first be presented with respect to the individual feature components, to be followed by discussion of the composite feature.

With the exception of Family Composition, the distributions of the component features are similar in form. Specifically, in over 60% of the coding occurrences, with values ranging from 61.8% to 74.4%, samples are matched on the demographic feature of interest. Judgements of noncomparability account for the remaining portions of the distributions. Evidence of an intermediate level of comparability is negligable, with representation by scores of 2 ranging from .7% to 2.8%. Family Composition differs from the other distributions insofar as only 42.4% of its distribution consists of ratings of demographic control.

Two comments can be made regarding these distributions. First, contrary to the impression one gathers from critics who focus on the failure to attend to demographic comparability as a major source of outcome inconsistency, it appears that a good portion of family interaction research has incorporated such control into its methodological practice. Room for improved control is of course still indicated. Second, the negligible representation of scores of 2 renders the three-category feature dichotomous. This situation may reflect problems with the clarity of the coding conventions, or the tendency on the part of researchers to either fully ignore or fully attend to a particular demographic index when determining the composition of their research samples. On the basis of my impressions of the literature I would argue against the

latter account. Various levels of concern with demographic comparability that cannot fully be captured by a yes/no distinction do seem to be reflected in the literature. Nevertheless, I do not feel that a revision of the coding guidelines would greatly alter the coding pattern of Demographic Comparability, given the unclear manner in which researchers convey their concern. Consequently, I recommend that the three-level coding scheme be simplified into a dichotomous form. With regard to the reliability of Demographic Comparability, moderate to high levels emerged from the analyses. Specifically, the measure-specific calculations yielded Pearson rs of .63, .84, and .93 for the component indices 'developmental stage', 'socioeconomic status', and 'family composition' respectively, coupled with interrater agreement rates of 77.1%, 90% and 95.4%. A similar picture emerged from the study-wide calculations, with interrater agreements ranging from 84% to 92%. An examination of the appropriate judgement matrices suggests that problems with reporting styles and coding guidelines resulted not so much in idiosyncratic response patterns as in a mutual discounting of the most troublesome element of the coding scheme. More specifically, both coders made scanty use of scores of 2.[35]

[35] This coding pattern draws attention to a fact that is often overlooked by critics of measurement devises; namely that measurement reliability does not necessarily imply measurement accuracy. Aftanas (1983) has recently proposed an evaluative scheme that highlights this distinction. He suggests that criteria of 'repeatability' and 'veridicality' replace the less

As expected, sex-of-the-focal-child is the one index of Demographic Comparability that yielded relatively poor reliability results: specifically a Pearson r of .43d an interrater agreement figures of 68.5% and 68% from measure-specific and study-wide calculations respectively. This can readily be accounted for by the absence in the coding guidelines of specific instructions regarding the use of information reported in a context other than that pertaining to demographics.

With regard to the treatment of Demographic Comparability as a composite feature, the intercorrelations between its four indices were examined as an a priori test of its appropriateness. With the exception of one negligible relationship between the indices sex-of-the-focal-child and developmental stage (r=.01), low to moderate levels of interrelatedness emerged from the data, with the remaining five correlations ranging in value from .27 to .66. Two conclusions can be drawn from these figures: 1) that the use of the composite score in subsequent analyses is justifiable, given that each index contributes independent information to the composite; and 2) that family interaction researchers cannot be divided into those-who-are and those-who-are-not sensitive to the issue of demographic comparability. Rather, empirical studies vary in the choice and number of features that are controlled for across

specific criterion of reliability to insure that such a distinction be made.

research samples.

The frequency distribution of the composite score provides further information about demographic control. Of the 809 coding occurrences, only 57 judgements (7%) reflect the absence of control on all four of the selected demographic indices (i.e., composite score of 4). In contrast, 234 judgements (28.9%) reflect the presence of control on all four indices. Poor attention to demographic comparability (i.e., scores of 4 to 8) is reflected in 40.4% of all coding judgements, while reasonable-to-excellent attention (i.e., scores of 9 to 12) is reflected in the remaining 59.6%. Thus, the picture that emerges from the composite frequency distribution is consistent with the one that emerges from the distributions of the individual indices. Of course a test of the relative utility of the individual and composite features depends on the use of regression procedures that are introduced during later stages of integration.

B. Sample Specificity

Sample Specificity, (i.e., the degree to which a sample is homogeneous in composition) is another methodological feature of importance, though it has not been well attended to in methodological critiques. There are a number of obvious disadvantages associated with the use of samples that are heterogeneous with respect to influential factors.

First, such usage may result in a within group variance of large enough magnitude to block the emergence of any significant effect. Moreover, if such an effect does emerge, the levels of the sample characteristic to which it can be appropriately generalized remain unknown. Three aspects of sample specificity may be of particular relevance in the family process literature: specificity of diagnostic category, of developmental stage, and of sex-of-offspring.

a) Diagnostic Specificity. As already indicated in the section on Diagnostic Category. samples utilized in family research tend to be too heterogeneous to be classified in terms of anything but the broadest of diagnostic divisions. As a partial antidote to problems resulting from such imprecise coding and in order to more fully address the effects of diagnostic distinctions on family interaction, studies were classified not only in terms of broad diagnostic distinctions but also in terms of the diagnostic specificity of their clinical samples.

Before presenting coding guidelines for Diagnostic Specificity I would like to first suggest a possible reason for the continued use of heterogeneous samples in interactional research. Given the general rejection in the clinical field of unitary conceptualizations of both schizophrenia and nonschizophrenic disturbances, such usage seems puzzling. It derives, I suspect, from a belief in the broad applicability of the systems model and from a faulty

understanding of what that implies. Although the pioneers of family systems theory focused almost exclusively on the schizophrenic family, their descriptions of family process were thought to extend to all forms of disturbance. On both an implicit and explicit level, systems proponents held (and continue to hold) that the actual form of symptomatic behavior manifested by a family member is a product of adventitious circumstances and, as such, signifies very little about the overall patterns of disturbance within the family (see Minuchen et al., 1975).

Fortunately, some movement away from this perspective can be detected in the family interaction literature. As evident from the feature Diagnostic Category, the focus on the schizophrenic family has gradually been shifting within the last decade to include a variety of other disturbed populations. This shift reflects some acknowledgement of the possibility that different psychopathologies are associated with different patterns of family interaction. Second, there are a number of investigators who have managed to narrow their focus to relatively tightly defined populations. Steinglass (1976) and Marshall and Neill (1977), for instance, have respectively examined alcoholics and obese individuals and their families from a systems perspective, and have offered case-study material to substantiate their theoretical claims. Of the more research-oriented reports, studies by Mishler and Waxler

(1968) and Hetherington, Stouwie and Ridberg (1971) come to mind. Mishler and Waxler divided schizophrenic families into good and poor premorbid groups, a procedure that has since been followed by a number of other researchers, while Hetherington et. al. differentiated delinquent families into unsocialized-psychotic, neurotic-disturbed and socialized-subcultural categories. The fact that both groups of investigators provided empirical justification for their use of homogeneous subgroups underscores the importance of diagnostic specificity in family process research. Finally, critics have done their part by strongly advocating further differentiation of the major schizophrenic/nonschizophrenic division in family research (e.g., Jacob, 1975; Mishler & Waxler, 1965; Riskin & Faunce, 1972).

Favourable as these signs may be, the issue of diagnostic heterogeneity remains to be addressed in the body of family process research currently available for integration. Efforts were made in this regard in the present study by classifying studies in terms of their diagnostic specificity, thereby preparing the way for the examination of the effect of this feature on study outcome during later stages of integration.

One final point regarding Diagnostic Specificity: The composition of the nonclinical control is of course as much an issue as that of the clinical sample. Unfortunately, the

reporting of information pertaining to the sampling and composition of nonclinical groups is markedly casual in family process research. Nonclinical controls are defined almost exclusively in terms of the absence of diagnostic labels and/or the absence of exposure to mental health factilities rather than in terms of the presence of specific characteristics. Hence the problem of diagnostic specificity of control groups was not addressed in this research.

Studies were assigned one point for each of the following four criteria. Thus, each study received a rating of diagnostic specificity ranging in value from 0 to 4, with higher scores representing greater degrees of specificity.

1) Clinical samples are restricted to only one of the three major diagnostic divisions - schizophrenia, nonpsychotic disturbance and delinquency.

2) Clinical samples are further restricted to one reasonably well-defined subgroup of a major diagnostic division. Schizophrenic samples may be subdivided according to any of a number of bases such as premorbid adjustment, symptom chronicity, severity of presenting symptoms and nosological subtype (e.g., paranoid, hebephrenic, etc.). Delinquent samples may be subdivided according to the nature of the legal infraction (e.g., aggressive vs. victimless act) or according to nosological subtype (e.g., unsocialized-psychothic, nerotic-disturbed, and socialized-subcultural). Nonpsychotic disturbed samples may be subtyped into a variety of diagnostic groups, namely neurotics/true phobia, simple phobia, emotional-somatic disorder, and substance abuse. Examples of specific diagnoses included under these headings are presented in Appendix E.

3) Clinical families (or as is most often the case, "identified patients") have been subjected to the same or similar diagnostic procedures. This condition can be met in one of two ways: a) through the researcher's reliance on a single subject source - even

if that source is an entire institution, the likelihood of diagnostic consistency is higher than in cases in which the researcher has sampled from a variety of hospitals, clinics and/or private practices; or b) through the independent assessment of each participating family by a member of the research team.

4) Clinical families are homogeneous with respect to the locus of the presenting problem. That is, the symptom-bearer is either the child, or one of the spouses, or (occasionally) the marital relationship.

A number of problems with the rating scheme were evident prior to its application.. First, the scheme fails to take into account the proportions of the various categories of subjects that make up a research sample. In this sense it fails to differentiate between samples that consist predominantly of one clinical subgroup with only marginal representation from other subgroups, and samples in which a range of subgroups are equally represented. The former type of sample is of course more homogeneous than the latter. While some studies do contain information on specific sample proportions, the general exclusion of such information from research reports precluded incorporating it into the rating scheme. The second problem has already been discussed in reference to the rating scheme for Demographic Comparability, namely the untested assumption of the equivalent importance of the four features that make up the composite score. Again demands for task manageability and for the economic use of study features in meta-analytic procedures prompted the construction of a composite score. Third, with respect to the second criterion, there are no

grounds for assuming that the specified subtypes reflect equivalent degrees of homogeneity, either within or across diagnostic groups. Consequently, the criterion reflects a concern with rather than a precise level of diagnostic specificity.

Additional problems with the feature that became apparent during the course of coding can be traced to the following sources. First, the fact that diagnostic information is often presented in general unelaborated terms (e.g., adolescent adjustment reaction, reactive behavior disorder) makes it difficult to apply with a high degree of confidence the first two criteria of the coding scheme. The fact that these criteria are largely operationalized in terms compatible with traditional psychiatric nosology creates additional coding problems, specifically with reference to studies whose authors clearly reject the value of the nosological approach to diagnosis. This situation can best be illustrated through example. In a study by Bugental, Love, Kaswan and April (1971), the whole gamut of child-related problems are described with reference to the school environment in a manner consistent with the behaviorist approach to assessment. In accordance with the decision to base diagnostic distinctions on the labels provided by primary researchers, the coder is forced to ignore the sample's obvious heterogeneous quality and classify it in terms of 'students with behavioral problems', a subgroup of

the major diagnostic category 'nonpsychotic disturbed.' (see Appendix C). Even if coding guidelines were modified to take into account differences in researchers' approaches to psychiatric nosology, the general poverty of diagnostic details in the family process literature would preclude an accurate coding of Diagnostic Specificity. The reporting styles of primary researchers also cause problems with respect to the third coding criterion, more because of incomplete (hence ambiguous) information regarding the nature of diagnostic assessment than because of the absence of such information.

Apart from the reporting styles of family process researchers, coding difficulties can be linked to an omission in the coding conventions, specifically to the absence of instructions pertaining to cases in which elements of the provided diagnostic information are not consistent with regard to Diagnostic Specificity. For example, clinical subjects may be described in terms of a variety of nosological subtypes (indicative of nonspecificity) at the same time as being equated on symptom severity or chronicity (indicative of specificity). In such cases, I relied on a liberal scoring standard by attending to the most 'specific' element of the diagnostic information. Instructions pertaining to 'inconsistent' diagnostic information should be incorporated into the coding conventions.

Calls for a more dramatic revision of the coding scheme for Diagnostic Specificity come from the frequency and reliablity data. These results, in fact, jointly suggest that all but the second criterion of the feature be dropped from consideration. With respect to the first and fourth criteria, both coders agree (interrater agreement ratios of 99.8% and 100%) that there is insufficient variability in the data to warrant their inclusion, with specificity judgements accounting for over 95% of coding occurrences in both cases. With respect to the third criterion, its exclusion is being recommended on the grounds of inadequate reliability, attributable mainly to the reporting styles of the primary researchers rather than to inadequate coding conventions. Interrater agreement figures of 51.8% and 64% emerged from the measure-specific and study-wide calculations, respectively.

Data on the second criterion, on the other hand, supports its potential utility for the purposes of integration. The distribution of feature categories is adequately balanced, with specificity judgements accounting for 34.1% of coding occurrences and nonspecificity judgements accounting for the remaining 65.9%. Moreover, adequate interrater agreement ratios of 84.4% and 84% from the measure-specific and study-wide calculations respectively suggest that the coding conventions are serviceable in their present form.

b) Developmental-Stage Specificity. As already indicated in the section on Developmental Stage, there is some speculation that each developmental stage in the family life cycle may be associated with unique forms of interaction. If such is the case, families at different stages in the cycle ought to be assessed separately (see Hill, 1965). Confusing results might well emerge if families with preschoolers and families with adolescents were treated as equivalent in, for instance, a study on parental intrusiveness. In acknowledgment of this concern, studies included in the present thesis were coded for degree of developmental specificity of their samples. The number of the following age levels (defined in reference to the focal child) included in study samples provided the basis for such ratings: 2-4, 5-9, 10-13, 14-17 and over 17 years of age. Thus, for instance, a study that includes subjects ranging in age from 5 to 16 years received a rating of 3. In the case of a discrepancy between the developmental specificity of clinical and nonclinical samples, the study was assigned the rating of the least specific sample. As in the case of Diagnostic Specificity, the coding scheme for Developmenal Specificity fails to attend to specific proportions of age levels included in research samples. Again, reports of sample composition are generally not sufficiently detailed to allow for such distinctions.

The application of the coding scheme for Developmental Specificity posed no particular problem, as is evident in the adequate interrater agreement figures of 92% and 96% for measure-specific and study-wide calculations respectively. With respect to the frequency distribution of the feature's categories, it presents a slightly more positive picture than the one that emerged from Developmental Stage, where the scarcity of homogeneous samples resulted in a predominance of nonscorable judgements. While samples of homogeneous composition are indeed poorly represented, samples restricted to only two age levels account for 49.9% of the coding occurrences. This mildly, redeeming fact is, unfortunately undermined by the relatively common use of samples with representation from more than two age levels, with three-level samples accounting for 10.4% and four-level samples for 23.5% of the distribution.

The reliability and frequency results pertaining to Developmental Specificity by and large support the viability of the feature for integrative purposes. However, given the imbalances in the frequency distribution, a collapsing of the categories into a dichotomous scheme, based on a less-than and greater-than-two-age-levels distinction should be considered.

c) Sex Specificity. As a final aspect of Sample Specificity, studies were coded according to whether their samples are homogeneous with respect to the sex of the focal child. Homogeneity can be achieved in one of two ways: by

restricting samples to one sex or by analyzing data separately according to sex. Studies were evaluated in terms of the presence or absence of sex specificitiy rather than in terms of degree of specificity with one exception; when information on sample proportions clearly indicates that one sex predominates (i.e., at least 90% representation) the study was classified as homogeneous. In cases in which clinical and nonclinical samples differ on sex specificity, the study was coded in terms of the nonspecific sample.

Much of the information pertaining to Sex Specificity is identical to, or readily derivable from, information pertaining to Sex of the Focal Child. Coding problems for both features are evident in the presence of one well-populated divergent cell that accounts for 15.6% of coding occurrences. Again, no explanation for the cell's crosstabulation of my nonspecific judgements with the unknown judgements of the second coder is readily apparent. The reliability of Sex Specificity is nonetheless adequate, as reflected in the interrater agreement figures of 81.5% and 96% for measure-specific and study-wide computations respectively. Moreover, the frequency distribution of the feature is well-balanced, with nonspecific judgements accounting for 44.5% and specific judgements for 48.7% of coding occurrences. Unknown and nonapplicable judgements make up the remaining 6.8% of the distribution. Thus, Sex

Specificity appears to be a good candidate for meta-analytic purposes.

C. Observer Expectancies

Given the power of diagnostic labels, control for observer expectancies may be an important feature on which to differentiate family interaction studies. Consequently, studies were classified according to whether they incorporate such control through the use of blind rating procedures. Two conditions of blind rating are specified in the coding conventions: 1) all information regarding diagnostic and treatment status must be withheld from the individual who observes and codes the behavior of family members; and 2) for interaction material whose coding format contains information pertaining to setting (i.e., ongoing interaction and videotape vs. audiotape and written transcript), the experimental setting must be equivalent for all diagnostic groups.

The coding scheme for Observer Expectancies consists of three categories: yes-control, no-control, and nonapplicable, for use with noninferential measures. The decision to exclude noninferential measures from consideration during the coding of Observer Expectancies was based on the assumption that noninferential measurement is immune to distortive influence. In retrospect this was a poor decision, particularly in light of the fact that researchers who most strongly advocate the use of

noninferentials precisely on the grounds of such immunity (i.e., behaviorists), have recently subjected the assumption to careful scrutiny and found it wanting (see Kent & Foster, 1977). Moreover, during the course of coding it became apparent that family process researchers also reject the assumption, insofar as they are as likely to present information pertaining to Observer Expectancies with reference to noninferential measures as to inferential ones.

Perhaps because it was difficult to keep the exclusion of noninferential measures in mind while coding outcomes on Observer Expectancies, particularly in the presence of relevant information, the nonapplicable category proved to be a source of nonreliability, accounting for 45.6% of the divergent judgements. A second likely source of nonreliability was the lack of sufficient coding details, specifically with reference to those situations in which both a blind rater and an informed criterion rater are jointly responsible for the measurement of the interactional material. In such cases the divergent judgements of the two raters are generally first negotiated and then incorporated into the study's data base in the form of compromise-ratings. In the absence of explicit instructions pertaining to these situations, the coder is left uncertain whether to code Observer Expectancies with respect to the blind rater or the informed criterion rater, or to base that decision on the degree to which the two sets of judgements converge.

In spite of problems with the coding conventions, they allowed for a reasonable level of interrater reliability, with an interrater agreement ratio of 80.7% emerging from the measure-specific calculations and an agreement ratio of 80% emerging from the study-wide calculation. With regard to the frequency distribution of the categories of Observer Expectancies, of the 539 applicable coding occurrences 45.5% received a yes-control judgement and 54.2% a no-control judgement. Thus, Observer Expectancies seems to be a potentially useful feature for integration, particularly if its boundaries are extended to incorporate noninferential outcomes.

D. Subject Expectancies

The degree to which the expectancies of those being observed are controlled for is yet another potentially influential methodological feature. Subject expectancies can be influenced in a variety of ways. The manner in which subjects are solicited may have a profound effect on their response to the experimenal context, as may the purported rationale of the study and the nature of payment for research participation, be it monetary remuneration, access to therapy, or access to research findings. As already indicated, the setting in which the experiment is conducted also likely impacts on subjects' expectancies and consequently on the nature of their responses.

A point worth noting: whatever the measures taken to control for subject expectancies, some differential reactivity to observation must be expected. Clinical and nonclinical families come to a study with markedly different visions of themselves, both self and/or other-imposed, and these visions no doubt affect the manner in which they participate in research (Haley, 1963; Riskin & Faunce, 1972).

Studies were coded on Subject Expectancies according to the degree to which clinical and nonclinical groups are equated on the above four 'demand' indices, with a score of 1 assigned to each controlled index. Thus, each study received a rating on Subject Expectancies that ranged in value from 0 to 4, with higher scores representing greater degrees of control.

Impressions gathered during coding point to the elimination of Subject Expectancies from further consideration, largely because of the scarcity of relevant information coupled with the lack of clarity of the available information. It also became apparent during coding that the choice of Manner of Subject as a demand index was inappropriate, given that control on the index is generally not possible in clinical research. This is due to the fact that the researchers' contact with clinical subjects is largely dependent on the solicitation efforts of mental health workers, while contact with nonclinical

subjects is dependent on the solicitation efforts of those who work outside of the mental health profession (e.g., neighbours, school personnel, newspaper advertisements). There is only one set of circumstances in the family process literature in which clinical and nonclinical subjects are equated on Manner of Subject Solicitation; namely when the total research sample is selected from the general population and then divided into clinical and nonclincial subgroups on the basis of responses to a researcher-provided diagnostic measure (generally a self-report measure). There is a problem with this method of sample construction. Unless the researcher assumes that clinical and nonclinical families are equally represented in the general population, s/he must be content to research a 'clinical' sample that is likely not comparable to the typical clinical sample constructed on the basis of psychiatric criteria.

The data on Subject Expectancies is consistent with the above impressions. With regard to the manner of subject solicitation, both insufficient variability and low reliability constitute grounds for its exclusion. As expected, a hefty 84.1% of those coding occurrences on which information is available received a no-control judgement. This imbalance in the distribution could perhaps be worked with during later states of integration were it not for the unacceptable reliability rates of 60% and 53.7% for the study-wide and measure-specific computations respectively.

Moreover, the fact that the inadequate presentation of relevant information by primary researchers is likely the largest contributor to the index's nonreliability, argues against attempts to revise the coding conventions for the purposes of retaining the index. The index Setting also presents a highly imbalanced distribution, with 80.5% of coding occurrences consisting of yes-control judgements, 16.6% of no-control judgements and the remaining 3% of unknown judgements. In this case, however, interrater agreement rates of 83.9% and 88% suggest adequate reliability. However, the usefulness of Setting as an index of Subject Expectancies depends on its joint consideration with other relevant indices. By itself it adds no information that cannot readily be extracted from the feature Setting. With regard to the index Nature of Payment the unavailability of relevant information undermines its viability, with unknown ratings accounting for a hefty 68% of all coding occurrences. A similar though less extreme picture emerges with respect to the index Purported Rationale where 48.3% of coding occurrences received an 'unknown' judgement. Of the remaining occurrences, 40.7% were coded as yes-control and 59.3% as no-control. Reliability assessment of the index yielded moderate results: a 70.2% interrater agreement ratio from the measure-specific calculation and a 64% interrater agreement from the study-wide calculation.

To summarize, the feature Subject Expectancies does not appear to be suitable for the purposes of integration for a number of reasons: the scarcity of relevant information, the inadequate nature of available information resulting in poor measurement reliability, and the insufficient variability of index categories.

E. Reliability of Measurement

As in any research that involves the use of rating scales and coding schemes, the issue of reliability looms large in the area of family interaction. There are a variety of ways of defining reliability: it can refer to the stability of measurement over time, the internal consistency of the measurement instrument, the extent of agreement between two separate ratings by a single observer (i.e., intrarater reliability) and the extent of agreement between the ratings of two independent observers (i.e., interrater reliability). Family researchers who have been concerned with reliability have by and large restricted their attention to the lattermost form.

Some criticism has been extended against the rater-reliability approach to interactional data. Pointing to the relationship between high interrater ratings and the use of rigorous training procedures to teach observers how to use behavioral coding schemes, Haley (1964, 1972) has argued that the dependency on such training defeats the whole

purpose of using more than one observer. The training of observers to view the data in the same way is, according to Haley, tantamount to collapsing multiple observers into one observer. He furthermore has questioned whether the data generated from any behavioral coding scheme reflects the realities of the process under study or merely a shared perspective (i.e., "construct" vs. "consensus" validity). Hence, Haley's reliance on low inference measures that are less vulnerable to reliability concerns.

Haley's arguments are, I feel, important, because they underscore a common tendency to confuse reliability with validity; that is, to accept a demonstration of reliable group differences as evidence of group differences that are both theory-relevant and meaningful. In reference to family process research, it is worth pointing out that the detection of reliable differences between diagnostic groups does not inform us about the underlying nature of those differences or about the relevance of those differences to psychopathology. Riskin and Faunce (1972) make a similar point about process research, while Olson (1970) extends the argument to the family therapy outcome literature.

While aware of the drawbacks inherent in the rater-agreement approach, I accept its utility in family interaction research insofar as I accept the position that reliability must be viewed as a first and necessary (though by no means sufficient) requirement for valid information on

family process. Consequently, studies included in the present research were differentiated according to whether or not they contain reliability data. In addition, the nature of the reported reliability coefficients and their values were recorded, in order to prepare the groundwork for examining the effect of measurement reliability on the magnitude of study outcome during later stages of integration. During the process of developing the coding scheme, it became apparent that this objective would only be met with some difficulty. While a preliminary examination of the literature suggested that a good portion of family studies do report reliability figures, differences in the nature of the computations (e.g., Pearson _r_s, percentage of agreement, Winer MS within scores/MS between scores ratio), in the bases of the computations (i.e., total score vs. act-by-act) and in the terms in which results are reported (i.e., probability levels vs. magnitude of effect) interfere with meaningful comparisons.

As a compromise to this dilemma, I decided to restrict the scope of the examination of the effect of the feature Reliability to the coefficient(s) that had the largest representation in the literature, and at the same time attempt to translate as many as possible of the remaining coefficients into equivalent terms. In addition, I decided to ignore all other sources of discrepency among reliability figures.

The coding scheme for Reliability consists of three categories: coefficient-present, coefficient-absent, and nonapplicable, for use with noninferential measures. For the same reasons that were discussed with respect to Observer Expectancies, the decision to exclude noninferential measures from consideration during the coding of Reliability was, in retrospect, inappropriate. Not only do family process researchers report reliability information pertaining to noninferential measures, but the values of such coefficients more than occasionally exceed those pertaining to inferential measures -- an observation that struck both coders as a worthy topic for future consideration.

The frequency count of Reliability yielded the following information. Of the 473 applicable coding occurrences, there are 335 instances (68.7%) of coefficient-present and 148 (31.3%) of coefficient-absent. Thus, a concern with measurement reliability is reflected in a good portion of the family process literature. In studies that report reliability data, only two coefficients are represented to any substantial degree: percentage of agreement ratios and Pearson *r*s, with the former accounting for 61.2% of applicable coding occurrences and the latter for 30.3%. The percentage of agreement figures range in value from 66% to 97% with a mean of 75.5%. With the exception of one value of .40, the Pearson *r*s range from .64 to .98, with a mean of

.80. Thus, in studies that do report measurement reliability, adequate levels are by and large achieved.

With regard to the reliability of this study's coding scheme, high levels were anticipated for the feature Reliability, given that the coding task involves a straightforward recording of reported information without need of interpretation. Data on the feature bore out this expectation. Interrater agreements of 91% and 98.9% emerged with regard to the presence/absence of the reliability information and the nature of the information respectively.

F. Control for Overall Frequency

One final feature relating to methodological control entered into consideration in the present thesis. Study outcomes were evaluated according to whether or not they control for overall frequency of interactions. A simple tally of an aspect of interaction without any consideration of the extent to which it is represented in an entire interactional sequence, precludes the untangling of differences in the frequency of occurrence of the interactional feature of interest from differences in overall interactional frequency. (Of course, overall interactional frequency may be a legitimate substantive focus of a study.)

The coding scheme for Control for Overall Frequency consists of three categories: control-present, for cases in which the number of targeted interactional units is divided

by the total number of units; control-absent; and nonapplicable, for use in the following cases: 1) with global rating scales; 2) with noninferential measures that focus either on productivity or duration (e.g., turns to completion, decision time, eveness of rates of participation, numbers of initiated actions); and 3) in measurement situations in which the number of instances of a targeted behavior is fixed (e.g., number of wins in a game consisting of a fixed number of plays; number of times a person speaks first when the number of issues under discussion is fixed, as in the RDT).

In retrospect, it may not have been appropriate to distinguish between the control-present and nonapplicable categories. Although specific steps are not taken to control for overall frequency in the nonapplicable cases, the fact that overall frequency does not act as a confound in such cases may be sufficient to equate them with control-present cases.

The second problem with the coding scheme has already been discussed with reference to the feature Demographic Comparability; namely the failure to include instructions for cases in which a sensitivity to the issue of control is demonstrated through a post-hoc analysis of the effect of overall frequency on study findings. Again, while assessing the vulnerability of the targeted measures to the absence of control, such an evaluation does not alter the actual

measure. In spite of problems with the coding scheme, a reasonable 76.6% interrater aggreement ratio emerged from the data.

The frequency count of Control for Overall Frequency yielded the following distribution. Of the 809 coding occurrences, there are 287 (35.5%) control-present and 332 (41.0%) nonapplicable judgements, along with 190 (23.5%) control-absent judgements. If control-present and nonapplicable measures are combined into a single category, one can conclude that a large portion of the available literature, specifically 76.5%, has incorporated control for overall frequency into its methodological practice. Representation by the control-absent category is, at the same time, of large enough magnitude to warrant the retention of Control for Overall Frequency in the feature set.

To summarize, of the seventeen features relating to methodological control, seven appear to be inappropriate for the purposes of further analyses. Of these, four are indices of the composite feature Subject Expectancies and three are indices of the composite feature Diagnostic Specificity. Grounds for rejection include insufficient variability of feature categories, scarcity of relevant information, and inadequate reliability. The viability of the remaining ten features for the purposes of integration is, on the other hand, supported by the frequency and

reliability data. While imbalances in the distributions of feature categories are evident, the less well-represented categories in all cases provide enough data to justify an empirical examination of the relationship between feature and study outcome. Reliability of the feature subset, moreover, tends to be high, with measure-specific calculations ranging in value from 76.6% to 98%, and study-wide calculations from 84% to 100%. Only in the case of one index of Demographic Comparability (i.e., Sex-of-the-Focal-Child) is reliability poor, preimarily because of insufficiently detailed coding conventions. A summarization of the characteristics of features relating to methodological rigor is contained in Table 3.

Features Relating to Research Focus

Substantive domains: the reclassification of studies. In designing a study on family process, the researcher generally chooses a particular aspect(s) of interaction on which to focus his/her investigation. This choice is directed by theoretical expectations about dimensions of interaction considered to be most closely related to psychological well-being and psychopathology in the family. Many of these expectations were first introduced into the field by four influential groups of investigators led by Bateson, Lidz, Wynne, and Bowen. Although extensions and refinements have emerged with regularity in the literature,

Table 3

Features Relating to Methodological Region

Feature	Problematic Aspect	Measure-Specific Reliability	Study-Wide Reliability	Pearson *r*	Coding Revisions Recommended	Retain/ Exclude
Demographic Comparability	2-level scheme more appropriate then 3-level.				Yes	Retain
SES	----------	90%	84%	.84		
Family Composition	----------	95.4%	92%	.93		
Sex of Focal Child	Poor reliability attributed to coding guidelines.	68.5%	68%	.43		
Developmental Stage	----------	77.1%	84%	.63		
Diagnostic Specificity					No	Retain criterion 2 Exclude criteria 1,3,4
Criterion 1	Insufficient variability of categories.	99.8%	100%	N/A		
Criterion 2	----------	84.4%	84%	N/A		
Criterion 3	Poor reliability attributed to reporting styles of primary researchers.	51.8%	64%	N/A		
Criterion 4	Insufficient variability of categories.	99.8%	100%	N/A		
Developmental Specificity	Imbalances in distribution.	92%	96%	.93	Yes	Retain

N/A nonapplicable

Table 3 continued.

Features Relating to Methodological Region

Feature	Problematic Aspect	Measure-Specific Reliability	Study-Wide Reliability	Pearson *r*	Coding Revisions Recommended	Retain/ Exclude
Sex Specificity	----------	81.5%	96%	N/A	No	Retain
Subject Expectencies					No	Exclude
Manner of solicitation	Inappropriate choice of index.	53.7%	60%	N/A		
Purported rationale	Overrepresentation by unknown code.	70.2%	64%	N/A		
Nature of payment	Overrepresentation by unknown code.	79%	88%	N/A		
Setting	Imbalanced distribution.	83.9%	88%	N/A		
Reliability	Inappropriate exclusion of noninference measures.				Yes	Retain
Presence/ absence		91%	N/A	N/A		
Nature of coefficient		98.9%	N/A	N/A		
Control For Observer Expectancies	Distinction between Control-Present and N/A categories questionable.	76.6%	N/A	N/A	Yes	Retain

N/A nonapplicable

it is these initial expectations that have directed much of the research.

One obvious way of categorizing studies in a review is in terms of their chosen dimension(s) of interaction. Disaggregating studies according to this feature allows the review to go beyond the most general level of questioning (i.e., Can clinical and nonclinical families be differentiated in terms of their patterns of interaction?) and enter into more specific levels of inquiry such as: Do all dimensions of interaction that have received theoretical attention support empirical associations with psychopathology of equivalent magnitude, or does research point to the primacy of some dimensions over others? Do all diagnostic categories exhibit dysfunction within the same dimensions of interaction or can diagnostic categories be differentiated in terms of specific areas of dysfunction?

While it is a relatively straightforward task to identify dimensions of interaction that have been emphasized in theory, the task of identifying the substantive foci of individual research efforts presents some discouraging hurdles. First and foremost are the notable lack of correspondences between operations, labels, designs and constructs of interest. Apart from the unnecessarily liberal application of labels to relatively limited sets of operations and constructs, this lack of correspondence takes the following forms: 1) identical operational procedures

have been used to measure what seem to be distinct constructs; and 2) researchers have at times failed to incorporate into their work appropriate correspondences between research designs and theoretical constructs. Studies that ostensibly examine the ways in which power is distributed and coalitions formed within the family system are particularly vulnerable to this latter criticism. In light of the problems relating to terminology and operational specification, it seemed inappropriate in this review to use the designations of primary researchers as the basis from which to categorize studies according to their substantive focus. Consequently, I decided to rely on my own designations for this purpose, in accordance with Doane's (1978a) recommendations.

The task of reclassifying study constructs entails a number of requirements. The reviewer must be able to 1) recognize equivalent forms of operations independent of labels; 2) identify equivalent functions of operations through a clear assessment of research designs; and 3) attach theoretical labels to studies by evaluating the nature of the information produced, and by linking that information to constructs associated with theoretical work on family interaction. The success of my reclassification effort can be evaluated on logical grounds in terms of the extent to which it improves correspondences between labels, operations, and constructs. More important, however, is an

empirical test of its capacity to ferret out from the literature meaningful findings about dimensions of interaction. To establish optimal conditions for such an assessment, I decided to record the substantive designations of the primary researchers along with my own. This would allow me to compare the predictive utility of the two sets of designations during later stages of integration.

Problems with implementation. However sound the logic behind the reclassification strategy, I encountered enormous difficulty during my attempts to devise serviceable classification guidelines. What became apparent to me is that there are a number of impediments to substantive classification that can be linked not so much to issues of correspondence, as to the nature of the measures used by family researchers to index interactional constructs. To begin with there are those measures that are too diffuse to be readily tied to any theoretical notion of family life. A good example of such a measure is the nine-point scale designed by Shapiro and Wild (1976) to index the "consensus, clarity and complexity" of family Rorschachs. More of a problem because of their prevalence, are those measures that are insufficiently explicit in focus to clearly relect any one dimension of interaction. What the prevalence of such measures clearly points to is a willingness on the part of interactional researchers to sacrifice interactional pithiness in an effort to attain scientific rigor. Examples

of such troublesome measures are the widely used indices of verbal activity or word productivity, that quantify verbal output without any attention whatsoever to the information conveyed on either a content or a relational level. When computed on the family as a whole, such measures are meant to reflect the extent of a systems-wide communication breakdown (with an inverse relationship between verbal output and communication breakdown being postulated). I would argue that measures of verbal output used in this manner may relay very little information about the family but for its garrulousness, a characteristic that has no obvious link to systems tenets. Measure of verbal activity have also involved comparisons between family subsystems. Used thusly, they are meant to reflect the relative dominance of family members. Again, whether quantity of speech, independent of its informational function, reflects relative dominance or merely relative garrulousness is open to question. Concerns with relegating too much of the data base to a nonscorable status on the grounds of conceptual ambiguity and imprecision necessitated the construction of coding guidelines less rigorous than I had hoped. In other words, quidelines were designed to link as many of the inexplicit measures as possible to specific interaction dimensions while restricting the use of the nonscorable category.

An additional impediment to reclassification is created by a variety of measures that seem to index more than one substantive domain. While such measures may reflect problems on the level of theory and/or operations, they may also point to conceptual overlap among domains. Examples of multi-domain measures are those that assess the degree to which family members acknowledge each other's communications or, conversely, the degree to which members fail to acknowledge each other's communications through the use of evasions, deflections, tangentializations and the like. Two theoretically important substantive areas may be equally implied by such measures, namely communication disturbance and family cohesion. A multiple classification option was incorporated into the reclassifcation scheme to be used with measures that relect conceptually overlapping interactional dimensions. Guidelines for multiple classifications are contained in Appendix C.

Given that the above coding impediments can only be "strategized away" to a limited degree, the reviewer of the family interaction literature must be content to apply an imprecise classification of substantive domains to an imprecise set of operations and designs. From this somewhat compromised position, descriptions of a number of dimensions of interaction, along with guidelines for the allocation of studies under substantive headings, will now be presented. These dimensions were chosen first because they have

received a good deal of theoretical attention and second, because a preliminary examination of the literature suggested that a sufficient number of studies that reflect these dimensions were available for coding. The descriptions to follow are broad and general in tone, and focus on ideas that have achieved wide acceptance among family interactionists. While specific theorists are on occasion cited, there are no references made to constructs that are uniquely associated with any one theorist. It is at this point worth noting that the family interaction field is not characterized by a high level of conflict among the ranks (at least in regards to theory as opposed to methodology). In spite of divergent opinions concerning the interplay between intrapsychic and interactional perspectives, scholars who work within the systems paradigm tend on the whole to acknowledge the validity of each other's formulations. Moreover, they have tended over time to expand their own formulations to include the interactional foci of fellow theorists. This "strain towards comprehensiveness", which was first noted by Mishler and Waxler (1965) and more recently by Okun and Rappaport (1980,p. 73), affords a certain cohesion to the area that was not evident during its early development. Thus, theoretical distinctions in the family systems field can better be understood in terms of emphasis than in terms of kind.

A. Substantive Domains

As previously pointed out, the early family investigators wedded systems principles both to psychiatric principles and to a range of largely interdisciplinary formulations concerning human relationships: specifically role theory, communication theory, and information theory. Particular blends of these formulations suggest unique vantage points from which to conceptualize the family. Haley (1967) has proposed three such vantage points; the family can be examined in terms of its structure and the roles of its members, in terms of the processes between members, and as a productive unit responsible for specific tasks. Okun and Rappaport (1980) have examined family interaction along similar lines insofar as they have grouped family theories into two major categories that coincide with the first two vantage points proposed by Haley (referred to respectively as Structural and Communication theories).

Family structure: role distribution. In addition to relying on psychiatric principles, the early investigators of family structure borrowed heavily from sociological role theory. In particular, the Parsonian model of normatively correct role-behaviour was explicitely acknowledged (see Lidz, 1963) and the psychological health of the family was, and continues to be, assessed according to its terms. The Parsonian model defines normality largely in reference to gender-appropriate behaviour. The picture of the normal

family that emerges is one in which the father assumes family leadership and much of the responsibility for the family's instrumental needs, whereas the mother assumes a secondary position and much of the responsibility for the family's affective needs. In addition to maintaining these gender-linked roles, spouses establish a strong parental coalition as the basis from which to transmit modes of behaviour and thought appropriate to the age and gender of their offspring. Thus, the soundness of family structure, in Parsonian terms, rests on the establishment of appropriate boundaries between the sexes and between generations. Deviations from this prescribed normative structure are regarded as evidence of abnormality. In particular, a structure marked by sex-role reversal has been singled out as a strong indicator of psychopathology. This structure is characterized by an intrusive and dominant mother and a passive and ineffectual father.

Though it continues to be an influential force in theory and research, the traditional Parsonian model of family role distribution has been criticized on a number of grounds; first, for its failure to take into account inter-familial and inter-cultural variations (Mishler & Waxler, 1965), and second, for its emphasis on role distinctions rather than role complementarity. In regards to this latter criticism, both Guttman (1977) and Jackson (1977) have argued that the extensive collaboration between spouses may be more

important than who does what in insuring that family needs are consistently met, a view that has received some popular support in this age of Women's and Men's Liberation.[36]

Family Structure: emotional boundaries. While the importance of gender-related boundaries has been thus questioned, the notion of boundaries remains central to all family theorists that emphasize structure (Okun & Rappaport, 1980). Much attention has, for instance, been directed at the emotional boundaries that are established between the family system and its individual members. Implied in many of the constructs utilized by structural theorists (e.g., Bowen's fusion/differentiation of self, Minuchin's emeshment/disengagement, Reiss's consensus sensitive/ environment sensitive families) is the notion that the psychological well-being of the family can be defined in terms of the capacity of its members to experience (and to permit others to experience) both feelings of autonomy and of group cohesion. In contrast, individual /system boundaries associated with dysfunctional families have been characterized either by inappropriate emotional fusion between family members which limits the experience and

[36] Rather than establishing separate allocation quidelines for Role Distribution, I incorporated them into the quidelines for the domain Boundaries of Authority. While distinct theoretical origins prompted me to separate the two domains for the purposes of discussion, a shared central concern with distribution of power resulted in sets of measures that differ only with respect to the focus of the comparison: husband-wife for role distribution and parent-child for boundaries of authority.

expression of autonomy, or by inappropriate emotional distance between members which limits the experience and expression of group cohesion.[37]

> Allocation guidelines for Family Cohesion: The following categories of measures are included under this substantive heading: 1) measures that reflect the degree to which family members communicate with one another (e.g., time spent in and out of mutual button pushing in coalition games, total interaction time, number of initiated actions); 2) measures that reflect the extent to which values, opinions are shared among family members (e.g., spontaneous agreement scores on RDT, explicit agreements and disagreements); 3) measures that reflect the responsiveness of family members to one another, expressed either through the incorporation of individuals' choices into joint family choices (e.g., choice fulfillment on RDT) or through signs of attentiveness during communication (e.g., direct responses to questions, amplification of others' comments, visual contact during direct address, nonverbal digressions such as whistling and newspaper reading during family exchange); and 4) measures that reflect the degree to which family members explicitly declare their own opinions and feelings and allow and encourage other members to likewise express themselves (e.g., explicit exchange of opinions, requests for such declarations from others, ratio of person-oriented communication concerning the wishes, opinions and beliefs of other family members to total number of person-oriented communications).

Family structure: boundaries of authority. If only because of the spousal subsystem's chronological supremacy in the family life cycle, most family theorists would agree that in order for functional individual/system boundaries to

[37] The labels of substantive domains that appear in the allocation guidelines were chosen on the basis of their accessibility to coders, who may or may not be familiar with family systems constructs. These labels do not, in all cases, match those used in the presentation of theory.

be established for all family members, such boundaries must first be successfully negotiated between spouses. Failure to do so can result in any of the following dysfunctions: 1) marital conflict; 2) symptom formation in one spouse; and 3) symptom formation in one or more of the children. While there is no epidemiological data of which I'm aware that suggests different rates of occurrences for these dysfunctions, much of the theoretical work and most of the empirical work on family process focuses on families in which the child is the symptom bearer. Views of child symptomatology that have been advanced by interactional theorists have been highly consistent. For such theorists, the notion of boundaries is again considered of pivotal importance, in this case the boundaries of authority or power that exist between generations. The type of boundary associated with healthy functioning is described as hierarchical in form. Such a structure is thought to be anchored in a strong parental coalition which provides the basis for a judicious allotment of power to the family offspring during the various stages of their development. The structure associated with dysfunctional families is, in contrast, described either as egalitarian or authoritarian. The egalitarian structure is characterized by an inconsistent or relatively equal distribution of power among family members regardless of age, with a resulting blurring of inter-generational distinctions, while the authoritarian structure is marked by a complete usurpation of control by

the parents with a resulting rigidification of generational distinctions. Both dysfunctional power structures are thought to be anchored in a troubled marital relationship that is unable to contain within its boundaries the dissatisfaction of the spouses. In the face of chronic marital discord and escalating conflict, coupled with an inability to communicate dissatisfaction directly, husband and wife begin to involve their child as a go-between.

Three characteristic patterns of involvement have been described in the literature, the first two associated with egalitarian power structures and the latter with authoritarian power structures (see Minuchin et. al., 1975). The first pattern is marked by a split spousal dyad and a child who is openly pressed to serve as a spokesman for both parents. The second pattern is also marked by a split spousal dyad, but the child in this case forms a stable cross-generational alliance with one parent against the other. The extent to which the excluded parent tries to upset this alliance varies. In some instances, the parent may simply disengage from family life. The third pattern is distinguished by an ostensibly united spousal dyad that fails to acknowledge any underlying marital strife. Husband and wife instead channel all communication into parental concern for and frustration with their unmanageable or "mentally-ill" child and become involved in various forms of scapegoating. In the first two patterns the child is

allotted excessive power insofar as s/he is invited to take the place of the disenfranchised spouse. In the latter pattern the child is allotted insufficient power insofar as s/he assumes a handicapped or "bad" role.

Allocation guidelines for Boundaries of Authority: Two categories of measures can be identified that fit under this substantive heading. The first consists of attempts to detect the presence of cross-generational alliances by means of measures that reflect the relative strengths of parent-parent and parent-child bonds (e.g., differential rates of communication between husband and wife and parent and child, differential rates of agreement and disagreement, differential rates of support, differential rates of who-follows-whom in speech sequence). Measures that compare diagnostic categories in terms of lines of alliances must base such comparisons on within-group differences between parent-child and husband-wife interactions.

The second category consists of those measures that attempt to assess the relative dominance of family members. This can be expressed in any of the following ways: 1) relative capacity to influence others or to be influenced by others (e.g., changes in opinion to match that of another family member, extent to which individual preferences are incorporated into group preferences); 2) relative capacity to direct the flow and content of family exchange (e.g., successful interruptions, defined in terms of entrance of third party into a two-party interaction that leads to a change in topic, being successfully interrupted, speaks first, speaks last); 3) relative use of controlling and being controlled forms of communication (e.g., ratings of explicit attempts to direct or limit the behavior of another person, behaviorally-focused topic change concerning deportment demands); 4) relative performance on a competitive game; and 5) relative amount of participation in family exchange, either of an active or passive nature (e.g., talking time, statement length, number of statements received).

Measures that examine relative dominance/passivity must base any comparisons that are made between diagnostic categories on within-

group comparisons, the nature of which varies from measure to measure. There are first those measures that focus on husband-wife comparisons, often in an attempt to tap the theoretical construct of role-reversal. In addition, there are those measures that focus on parent-child comparisons, the concern in such cases being the appropriateness of the positions assumed by family members in the family power structure. The results of such studies are often presented in terms of dominance hierarchies which serve as the basis for cross-diagnostic comparisons. Finally, there are those measures that focus on individual-group comparisons by assessing the individual's communication relative to that of the family unit (i.e., through the calculation of the proportion of the family's use of a dominance-related measure attributed to the individual).

Family structure: boundaries with external social systems. Apart from the boundaries that exist within the family system, there are those that exist between the family system and external social systems. While these latter boundaries have received rather limited theoretical attention and virtually no empirical attention,[38] an observation that frequently appears in the literature is that high levels of family dysfunction are associated with relatively impermeable between-systems boundaries. Reasons for restricted interaction with the external social environment are thought to be two-fold. On the one hand, it fulfills a defensive function insofar as it permits the family to avoid information that might challenge its status

---

[38] In a 1966 paper titled New Stage in Family Settings, Lyman Wynne called for more conceptualization and hypotheses-formation in relation to the problem of boundaries between nuclear family units and external social structures. This call has not, as yet, been answered.

quo. On the other hand, restricted interaction may be imposed on the highly dysfunctional family to the extent that its members (at least its symptom-bearers) are impaired in self-maintenance and social skills.

A number of ideological mechanisms seem to be utilized by family members that support such an isolationist policy. For instance, members often share the myth of a dangerous and hostile extra-familial environment, coupled with the myth of a harmonious, satisfying and self-sufficient family environment. When one family member comes to be labelled as a problem by extra-familial social structures (e.g., school system, juvenile court, etc.), these myths may undergo slight modification. Family life may, in such instances, be characterized as normal and untroubled <u>save</u> for the presenting symptom. Alternatively, extra-familial social structures may be held totally responsible for the problem, thereby sparing the family unit any discredit (see Ferreira's (1977) short paper on family myths). Given the absence of empirical attention to the boundaries between the family and external social systems, no attempt was made to devise coding guidelines for this domain.

In sum, theorists who have focused on the internal structure of family systems have earmarked gender, age and degree of individual autonomy as the important parametres of such structure. Those theorists who have attended to between-systems structures, have identified the degree of

interaction between the family and external social systems as an important indicator of family functioning.

Family process. The process view of families was derived from the merging of communication theory and systems theory principles. Scholars who work from this vantage point are primarily concerned with the bits and pieces of communication that make up an interaction rather than with the structural context within which communication occurs (Okun & Rappaport, 1980, p. 73). Moreover, they share the premise that communication between family members functions primarily to establish the structural context of family life, by containing within it the rules of that context. This premise is articulated in terms of a distinction between two aspects of communication: 1) the content/report aspect through which the speaker conveys facts, thoughts, beliefs, feelings and the like and 2) the relationship/command aspect through which the speaker simultaneously conveys information about, and brings into play the rules of the family system. Communication or process theorists maintain that it is the latter function, generally referred to as meta-communicational, that holds the key to the understanding of family life. Thus, a primary research objective is to ferret out and articulate the interactional rules contained within the meta-communication of family members.

Communication of emotional rules. As in the case of structural theorists, process theorists do differ in regards to the substantive domains within which they promote this search for rules. A number of theorists, for instance, have focused on ways in which the emotional system of the family is expressed through communication (e.g., Wynne, Rychoff, Day & Hirsch, 1958; Satir, 1972; Minuchin, 1975). These theorists have differentiated families according to how they modulate affective expression. Healthy functioning is defined by them in terms of a family's capacity to express a wide range of emotions without experiencing excessive anxiety and without feeling that the durability of the family unit is threatened. In contrast, poorly functioning families are characterized by a highly restrictive range of emotional expression and by the experience of marked anxiety when the limits of that range are approached. The quality of a restricted emotional climate may vary from family to family. While some families permit only nurturing and pacifying expression, others cannot deal with tenderness of any sort and use anger and blame as the sole currencies of exchange. Such differences are not, by and large, considered to be particularly important or profound by family theorists. What is considered important is that both modes of affect are inflexible and moreover essentially false. In families in which negative affect is prohibited, much of the emotional energy of their members is directed at perpetuating a myth of family harmony. This myth is

repeatedly invoked as a means of softening or suppressing signs of impending distress. Instances of genuine positive affect such as the exchange of support, validation, and the like are in short supply. In families in which positive affect is prohibited, chronic discord, persistent efforts to undermine the worth of fellow members, and recurrent threats of the dissolution of the family unit similarly serve as collective defenses. A reliance of such modes of exchange leaves little room for family members to express dissatisfaction directly or to engage in genuine confrontation.

The restrictive affective patterns that have just been described have been conceptualized both in terms of systems-wide constructs and in terms of emotional styles adopted by family members. Of the former category, the constructs pseudomutuality and pseudohostility, first introduced by Wynne and his colleagues (1958), are the most well known in the field. Of the latter category, Satir's (1972) four-class typology of dysfunctional affective styles is consistent with many of the clinical descriptions contained in the literature. The styles described by Satir are restrictive in one of two ways; 1) they do not incorporate the personal into communication to a sufficient degree (i.e., "reasonable analyzing" and "distracting" styles); and 2) they do incorporate the personal into communication but mainly in reference to the other rather than the self (i.e., "blaming" and "placating" styles.

Allocation guidelines for Affective Expression: Three categories of measures can be identified that fit under this substantive heading: 1) measures that reflect the amount or intensity or range of affective expression without regard to style or valence (e.g., ratings of strength of feelings or alternatively of lack of feeling, inter-personal expressiveness, range of evaluative communications, relative amount of expressive as opposed to instrumental exchange); 2) measures that reflect positive affective expression (e.g., ratings of support, reciprocal reinforcement, tension release, laughter, jokes); and 3) measures that reflect negative affective expression (e.g., ratings of conflict, ratings of depressive mood, sarcastic disagreements, criticisms, reactions of shock to proposed solutions, ratings of inability or unwillingness to engage in supportive communication).

Communication of rules relating to authority. Apart from focusing on affect, communication theorists have also been interested in the ways in which family members express power relationships through communication. Two approaches to the issue can be identified. The first approach involves evaluating family relationships according to the frequency of "controlling" and "being controlled" communications. In contrast, the second approach does not define power in terms of the actual exercise of control, but rather in terms of a family member's success in defining the power structure (regardless of his/her position that structure) and in terms of a member's acceptance or rejection of the definitions imposed on him/her by other family members. In regards to this latter approach, theorists (e.g., Haley, 1963; Jackson, 1959; Watzlawick, 1964) have distinguished between two kinds of communication; 1) symmetrical communication in which two

interactants exchange the same kinds of behaviour (e.g., instruction/counterinstruction, assertion/assertion), thereby demanding that equality mark the exchange; and 2) complementary communication in which two interactants exchange different types of behavior (e.g., question/answer, assertion/agreement), thereby establishing a "one-up" and "one-down" position between them. Only the first of the approaches has received sufficient empirical attention to warrant its inclusion in a classification scheme.

> Allocation guidelines for Domineering Expression: Measures that assess the frequency of "controlling" and "being controlled" forms of communication are included under this heading (e.g., behaviorally focused topic change concerning deportment demands; dictatorial decisions, defined as instances in which one family member's decision is chosen to represent the family group in the face of an opposing majority decision; ratings of explicit attempts to direct or limit the behavior of another person; intrusiveness, defined as instances in which one person speaks when another has been explicitly invited to speak; proportion of person-oriented communications concerning the motives, feelings, wishes and experience of other rather than of self).

Communication clarity. There is a final dimension of communication that, if itself dysfunctional in form, may compound the effects of communication that is dysfunctional along other dimensions, namely its clarity. Disturbances in communication clarity have been described in a variety of ways; in terms, for instance, of poorly integrated messages, frequent shifts in attention, absence of a shared focus of attention, frequent use of cryptic referents, frequent use

of disqualificatons, and a general lack of clarity and precision or, alternatively, a rigid insistence or clarity and precision.

A form of communication disturbance that warrants particular attention is one in which there is a habitual lack of congruence between different elements or levels of messages. Such incongruence can take a number of forms and can involve either single or multiple channels of communication. For instance, a message may contain conflicting literal and metaphoric elements. Or a message conveyed by verbal content may be contradicted by its postural, tonal or gestural cues. Moreover, each level of an incongruent message contains within it a negative injunction. Given the contradictory nature of the message levels, an appropriate response to one negative injunction insures an inappropriate response to a second injunction. This form of communication has been labelled the "double bind" (see Bateson, Jackson, Haley & Weakland, 1956).

Incongruence in communication does not in and of itself generate disturbance; it can in fact mark some highly satisfying forms of exchange such as humor, play and ritual. It's impact only becomes damaging when a number of conditions operate concurrently. To begin with, it must occur within the context of a vital ongoing relationship in which the receiver of contradictory communication is dependent on, or at least perceives him/herself to be

dependent on, the sender (e.g., the child-parent relationship). Second, the receiver of double binds is prohibited from pointing to his/her dilemma and from attempting to clarify a message by commenting on its contradictory nature. Any such attempts are met with disqualifications that take a variety of forms such as evasions, sleights-of-hand, literalizations, status disqualifications and redundant questions (see Sluzki, Beavin, Tornopolsky and Veron, 1967). The acceptance of such disqualifications serves to consolidate the double bind. Finally, the receiver of the double bind does not have the option of ignoring the message by failing to respond or by extracting him/herself from the relationship in which it occurs. Consequently s/he must respond in what is necessarily an inappropriate way. In pathogenic circumstances, the consequences of such inappropriate responses are highly punishing.

Two broad forms of behavioral disturbances result from repeated exposure to such forms of communication. First, the receiver stops trying to decipher messages directed at him/her. Second, s/he opts for the only available means of escaping punishment, by communicating in a manner that is not readily deciphered by others; that is, by responding in kind. Thus the distinction between sender and receiver of double binds becomes blurred as the two members of an interacting dyad come to assume both roles.

Apart from wreaking psychological havoc on the individual, double-bind communication undermines the capacity of the entire family to function as an effective unit. Since members both receive and send out ambiguous messages, there is minimal exchange of useful information. Consequently, the group is unable to engage in problem solving, decision making, and other forms of productive activity in a very efficient or effective manner.

The theory of the double bind has attracted a great deal of attention and has generated a sizable body of literature. In the words of one of its original authors, it has precipitated "if not a scientific earthquake, at least a fair amount of commotion (Weakland, 1974). In spite of this kind of influence, the most recent review of the double bind (Olson, 1972) as well as the present integrative effort suggest that very little of the literature is directed at an empirical substantitation of the theory. A likely reason for this is that the double bind theory may not be a "theory" at all, or if so, it is a theory that in Bateson's own words is "so slippery that perhaps no imaginable set of empirical facts could contradict it (Bateson, 1966). In this sense, its influence rests not so much on any immediate explanatory power, as on the fact that it offers an alternative language or possibly an alternative epistimology, along with an alternative set of premises about human behavior (Bateson, 1966, Weakland, 1974). The

most important of these premises, insofar as it locates the study of psychopathology squarely within the interactional realm, is that communication and psychopathological behavior are intimately linked. As to the actual incorporation of the double bind into research efforts, far more common than attempts at theory-testing are attempts to operationalize features of communication that are consistent with double bind premises (e.g., various forms of communication disqualification).

> Allocation guidelines for Communication Clarity: Four categories of measures can be identified that fit under this substantive heading: 1) measures that assess the extent to which the communication of family members is unintelligible because of its bizarre or "schizophrenese-like" quality; 2) measures that assess the extent to which information is clearly and explicitly exchanged among family members (e.g., explicit agreements and disagreements, explicit declarations of ideas and opinions, identification of unclear or bizarre communication in others, requests for clarification); 3) measures that assess not so much the clarity of distinct bits of communication as the extent to which those bits are cohesively linked (e.g., shifts in themes, parallel messages, appropriate and inappropriate topic change); and 4) measures that assess the extent to which family members disqualify either their own communication (e.g., self-contradictory statements, lack of affective congruence between communication channels) or that of other family members (e.g., verbal tangentializations or evasions, change of subject that are labelled as answers, redundant questions implying doubt or disagreement without openly stating it).

In sum, theorists who have studied the family in reference to the communication that occurs between members have focused on the ways in which emotional structures and

power structures are expressed through communication, as well as on the clarity and accuracy of such expression.

The family as a task-oriented unit. This perspective, which has borrowed from both information and general system theory, is primarily concerned with differentiating families according to how effective they are in accomplishing those tasks that insure that the needs of their members are consistently met. Two views of family effectiveness can be identified, one clinical and the other informational. The clinical view is concerned with the capacity of the spousal dyad to meet demands related to love-making, parenting, social life, housekeeping, money-making and the like, and the capacity of the offspring to acquire social and instrumental skills appropriate to their stages of development. Family effectiveness from this perspective is judged by how harmoniously the family functions as a unit and, more often, by the absence of the expression of psychopathology in any of its members. The informational view, which is more "microanalytic" in focus, is concerned with the capacity of family members to transmit bits of task-related information among themselves without excess noise and confusion. Of the two views it is the latter that is more closely reflected in research efforts. Brought into the experimental setting, families are presented with an interactional task and their effectiveness in fulfilling its requirements is then judged in terms of energy expended

(e.g., time to completion, number of turns to completion) and in terms of the quality and quantity of the product of the interaction (e.g., number of correct responses, number of requested task components included in the product, quantity of output). Thus, once brought into the experimental context, the notion of family effectiveness is narrowed down to one of efficiency and productivity. This may present some problems insofar as, at some levels, experimental efficiency and clinical effectivensss may not go hand in hand. Put into other words, apects of a family's communication, essential to effective functioning, may not be particularly efficient.

> Allocation guidelines for The Family-as-a-Task-Oriented-Unit: Four categories of measures fit under this substantive heading: 1) measures that index the energy expended during an experimental task or, phrased somewhat differently, measures that index family efficiency at meeting task requirements (e.g., decision time on joint RDT, turns to completion on laboratory games); 2) measures that index the quantity of a task product independent of quality considerations (e.g., number of adjectives incorporated into a story, word production on joint family TAT); 3) measures that index the quality of a task product (e.g., clarity and cohesiveness of family TAT, number of correct responses in referential communication task, extent to which family members are able to manipulate chosen outcome in coalition game) and 4) measures that index families' ability to work within a task set to a sufficient degree to meet basic task requirements (e.g., opening gamit scale that measures extent to which family members accept task set, number of times family fails to agree on solution or come to a decision, number of five requested story components included in family TAT).

Flexibility/rigidity of interaction patterns. Apart from the three perspectives just described, there is one other aspect of family interaction that can be identified in the literature, namely its flexibility. It is important to point out that the dysfunctional patterns of interaction that have been discussed in reference to a variety of substantive areas are not the exclusive property of poorly functioning families. Effectively functioning families no doubt demonstrate any or all of these same patterns from time to time. They can, however, be distinguished by their ability to shift into more adaptive modes of exchange when circumstances demand it. Poorly functioning families, in contrast, persist in maintaining maladaptive patterns regardless of the pressing nature of their members' needs and in spite of the pervasive distress that characterizes family life. A body of family rules that are either overly blurred or overly rigid operates to protect this inflexibility. An insufficiently articulated or blurred body of rules fails to provide a basis around which to negotiate change, while an overly rigid body of rules is by definition resistant to modification. The distribution of roles in the dysfunctional family is similarly described as either overly rigid or overly blurred. Again, this prevents family members from assuming patterns of relating that adjust in accordance with changing circumstances. Doane's classification scheme (1978, pp. 362-363) was relied upon to allocate studies under the substantive heading of Flexibility/Rigidity.

> Allocation guidelines for Flexibility/Rigidity: Measures that demand no inferences about content and that are to a large extent concerned with the distribution of family exchange are placed in this category (e.g., evenness of distribution of who-follows-whom, degree to which distribution of who-speaks-to-whom is balanced, predictability of speakers in sequence, correlation between amount of silence across experimental tasks, rates of balanced triadic interaction relative to dyadic interaction, degree to which rates of frequency of speech among members are balance, frequencies of speech disruptions).

During the process of coding, it became apparent that the classification guidelines for the feature Substantive Domain can not be applied without considerable difficulty. To begin with, operational specifications are often presented in what seem to be incomplete terms. It is, in fact, difficult at times to determine whether the specifications provided by the primary researcher are those that were actually used to measure interactional material or whether they are versions that have been edited for journal presentation. When standard coding instruments are involved (e.g., the IPA), it is of course easy to access original specifications, regardless of the degree of detail provided in any paarticular study. When, on the other hand, idiosyncratic operations are involved, the problem of incomplete or ambiguous specifications is less easily addressed. In light of the fact that many of the troublesome specifications are contained in studies published ten to fifteen years ago, it seemed unrealistic to attempt to obtain original specifications.

Confronted with a large number of observations that are insufficiently detailed, the coder may find him/herself relying to some extent on a strategy that has been rejected for the purpose of a review of the family interaction literature; namely, to use the labels of operations provided by the primary researchers as substantive cues. During the course of coding I found myself referring to the labels in the case of particularly difficult substantive decisions, in an effort to avoid overuse of the 'nonscorable' code.

Apart from the form of operational specifications, coding difficulties can also be linked to the nature of the measures. Specifically, more often than anticipated, measures seem to reflect as much one substantive domain as another. The inclusion of the multiple-classification option in the coding guidelines for Substantive Domain did little to alleviate the problem insofar as it was not relied on to any substantial degree. It is uncertain whether such limited usage can be attributed to a lack of clarity in the guidelines, or to the fact that the option can not be expected to operate effectively within the context of a ten-category scheme that already taxes the decision-making capacities of the coder.

There is one further way of accounting for the problems that the coder can expect to experience during his/her efforts to code the feature Substantive Domain, namely that the task itself may be misquided. The task was incorporated

into the present thesis in response to Doane's (1978a) argument that the conceptual thrust of family interaction studies can only be revealed to the reviewer through his/her own efforts at tightening up the correspondences among substantive labels, constructs and operations. The appropriateness of the task, therefore, rests on the correctness of the assumption that the major problem regarding the classification of the interactional literature into substantive areas is one of correspondences. If, however, the problem goes beyond correspondences to a dirth of appropriate links between operations and constructs, the task of establishing such links through the use of the coding conventions for Substantive Domain is rendered not only difficult but ultimately unprofitable. Put into other words, in order that the task of classifying Substantive Domain be seen as feasible, the coder must have some confidence that the family interaction literature meets some standard of conceptual adequacy.

My own confidence in the conceptual status of the literature was clearly threatened during the course of coding Substantive Domain, not only because of the difficulties I encountered in making substantive decisions, but also because of the problems I experienced when attempting to record the substantive designations of the primary researchers, in an effort to build into the procedures some protection against the subjectivity of

Doane's reclassification strategy. Interestingly, both coders abandoned this task about half-way through the process of coding, after independently deciding that it could not be accomplished with any degree of reliability. Attempts to read between the lines of researcher's weak efforts to link their operations to a theoretical context were largely unsuccessful. It is often the case, moreover, that researchers simply fail to specify any such context.

In light of the many coding difficulties, I did not expect high levels of interrater agreement to emerge from the reliability assessment of Substantive Domain. Moreover, I had some concern regarding the meaningfulness of such an assessment, insofar as I expected some degree of agreement to arise out of a joint reliance on the labels of operations rather than out of convergent appraisals of conceptual foci. Nevertheless, it was important to determine which, if any, of the substantive categories can be classified with a reasonable degree of reliability.

The reliability assessment of the feature Substantive Domain can be distinguished from the assessments of the other study features in a number of ways. To begin with, because of the difficult and time-consuming nature of the classification task, it was restricted to 278 rather than to 410 judgements. Second, the inclusion of the multiple-classification option resulted in an unequal number of judgements per coder. The following rule was consequently

introduced to equalize judgements for the purposes of the assessment: When only one of the two coders applied a multiple-classification to a particular outcome, the case was treated as a single classification. More specifically, a match between the single judgement of one coder and any one element of the multiple judgement of the second coder was entered into the interrater agreement ratio as an agreement. The absence of a match, on the other hand, was entered in as a disagreement. Thus, a liberal standard of partial rather than absolute agreement pertained in the assessment. The inclusion of the multiple-classification option had another effect in that it interfered with the capacity of the assessment to track down the exact nature of divergent judgements. Specifically, in those cases in which the single substantive judgement of one coder fails to match any element of the multiple judgement of the second coder, there is no basis for selecting one particular element of the multiple judgement over another as the source of the descrepancy. Similarly, in cases in which no match exists among the elements of two sets of multiple judgements, there is no basis for selecting out pairs of divergent judgements. Fifty-three percent of all divergent judgements regarding Substantive Domain are 'nonspecifyable'.

A final way in which the reliability assessment of Substantive Domain can be distinguished from other such assessments is that its results are more logically

interpreted in reference to its ten distinct substantive codes than in reference to a single dimension. Agreement ratios were computed for each substantive domain by dividing the number of times both coders applied the domain to an identical outcome by the total number of times coders made use of the domain. Of the ten domains only five generated interrater aggreement ratios that exceed 65%, namely Boundaries of Authority, Affective Expression, Communication Clarity, Quantity of Communication, and Communication Continuity. Agreement levels for these five categories range in value from 80% to 100%. In contrast, Domineering Expression, Family Cohesion, Flexibility/Rigidity, Family-as-a-Task-Oriented-Unit, and Nonscorable generated values ranging from 16.7% to 64.7%.

It is interesting to speculate on the nature of such differences. That high levels of agreement emerged for the two codes that are defined in terms of morphological distinctions is not surprizing, insofar as the relevant measures did not have to be linked to abstract theoretical constructs either during the construction or application of the coding conventions. The high level of agreement for Communication Clarity is also understandable, in that it is consistent with the relatively favorable appraisal that the domain has received by the more pessimistic as well as the more optimistic among the ranks of family process critics (see Doane, 1978a; Goldstein & Rodnick, 1975; Jacob, 1975).

The high level of agreement for Boundaries of Authority, on the other hand, is surprizing, insofar as the substantive domain has been singled out as one in which family interactionists have had particular problems in achieving conceptual clarity (Liem, 1980; Turk & Bell, 1972). As to the remaining reliability differences, they my reflect differences in may grasp of areas of theory or, differences in the conceptual clarity and/or methodological adequacy of the various substantive areas.

There are a number of ways of responding to the reliability data pertaining to Substantive Domain. One could decide to eliminate the feature from further consideration on the grounds of reliability concerns. In light of the centrality of the feature, such a decision would, in my opinion, be tantamount to rejecting the suitability of the family interaction literature for integration on conceptual grounds. Alternatively, a decision could be made to restrict the integration to those substantive areas that can be coded with a reasonable degree of reliability. Such a decision could of course result in a serious reduction of the data base. Examination of the frequency distribution of the categories of Substantive Domain provides some indication of the extent of such a reduction. Of the 1004 coding occurrences, there are 252 (25.1%) instances of Family Cohesion, followed by 167 (16.6%) instances of Affective Expression and 125 (12.5%) of

Communication Clarity. The remaining eight substantive categories are represented more or less equivalently, with contributions ranging from 4.1% to 7.5%. This distribution suggests that large data losses would likely result from the elimination of the less reliable substantive categories from the integration. Finally, a decision can be made to rework the less serviceable of the coding conventions with the objective of improving measurement reliability. On the basis of my impressions of the family interaction literature, I wqould not expect such efforts to be very fruitful.

B. Mode of Communication

One aspect of interaction that greatly adds to the complexity of the family systems field is that it occurs across a variety of verbal and nonverbal channels. Apart from a concern with thoroughness, there are a number of reasons for examining as wide a range of channels as possible in family research. First, and most obvious, is the fact that any restriction on the range of channels results at best in a reduction in the amount of available information and at worst in a distorted pattern. Second, the observation that information conveyed on one channel of communication need not correspond to information conveyed on a second channel has been made repeatedly, particularly by students of communication disturbances. Full exploration of the expression of such incongruence demands that comparisons be made across as wide a range of channels as possible.

In spite of such considerations, methodological critics have supported the tendency of family process researchers to examine patterns of interactions largely from a verbal perspective. Such support has at its basis the belief that nonverbal behavior is inherently too complex and unreliable for scoring purposes (see Haley, 1967; Riskin & Faunce, 1972). Nonverbal behavior may indeed be difficult to score because of its subtle and fleeting nature. However, the use of videotapes during data collection mitigates these concerns to a large extent insofar as it allows for multiple viewing of the data. As to the issue of complexity, it of course increases when nonverbal as well as verbal behaviours are taken into account -- which is not to say that the independent assessment of either behavioral category is inherently more complex than the other. Of course, when both categories are included in a single study, the question of congruence between channels of communication enters in as an additional vantage point from which to examine the interactional material.

There are a number of available observational scales that include both verbal and nonverbal behavioral categories, most of which have emerged out of the child development field: the Parent Child Interaction Code (Lytton, 1973), the Human Interaction Code (White & Watts, 1978) and the Behavior Coding System (Patterson, Ray, Shaw and Cobb, 1969). The reader is referred to Fox (1978) for a

discussion of the respective advantages and disadvantages of these scales. Unfortunately, family systems researchers have not made use of these multi-channel scales. The one multi-channel scale that they have made use of emerged out of the field of sociology, namely the Bales Interactional Process Analysis (IPA). Apart from the fact that the IPA has been the target of extensive criticism (see Winter & Ferreira, 1967), it is often used by family researchers in conjunction with recording formats that strip its observational definitions of their multi-channel properties. By, for example, transfering all interactional information on to a typed transcript, the researcher often presents the coder with only the verbal aspects of the IPA definitions.

The feature Mode of Communication was incorporated into the present research effort to prepare the groundwork for comparing the informational value of the various channels of communication, as well as for examining the differential effect of multi-channel and single-channel data during later stages of integration. The coding scheme for Mode of Communication consists of the following categories: 1) verbal; 2) verbal/tonal; 3) postual; 4) verbal/tonal/postural; 5) nonapplicable, for measures that focus on the duration of an interaction (e.g., time to completion) and for those 'artificial' measures associated with highly structured measurement situations (e.g., button pushes); and 6) nonscorable, for composite measures that are

not uniform with respect to Mode of Communication. Rather lengthy instructions were inserted in the Coding Appendix to facilitate the coding task. They are as follows:

> Studies are classified in terms of the mode of communication used by the researcher as a source of interactional information. Two features of studies can direct this classification: 1) the definition of the outcome measure and 2) the device used to record interactional material (e.g., typescript, audiotape, videotape). In cases in which the definition of the outcome measure is restricted to a consideration of one mode of communication, information about the recording device is immaterial. For instance, a measure of "interruptions" defined in terms of instances of simultaneous speech is restricted to a consideration of the verbal mode of communication, regardless of the choice of recording device. On the other hand, in cases in which the definition of the outcome measure or coding category incorporates aspects relating to more than one mode of communication, the recording device must be relied upon to provide information about the researcher's choice of mode. For instance, a researcher who uses those IPA categories that incorporate postural, tonal and verbal aspects of communication, may either attend to all three aspects by using videotape during data collection, or may restrict attention to a specific aspect of the definition, for instance the verbal aspect, by recording interactional material on typescript.
>
> The following guidelines can be applied to the coding of the study feature Mode of Communication: First, examine the definition of the outcome measure to determine whether or not it incorporates more than one mode of communication. If only one mode is under consideration, code the measure in terms of that mode. If, however, more than one mode enters into the definition, refer to the recording device (typescript, audiotape or videotape) to determine which aspect of the definition is actually being incorporated into the data.

During the course of coding Mode of Communication, two problems with the guidelines became apparent. First, they

fail to incorporate a verbal/postural category, to be used specifically with reference to who-speaks-to-whom kinds of measures. Second, the assumption that the typed transcript automatically implies the verbal mode of communication is not always correct. There are a number of cases in which the researcher transfers interactional information from videotape to typed transcript while retaining postural and/or tonal information in the form of verbal description. While such second-hand information may not be equivalent to the postural and tonal information contained on the actual videotape, the coding scheme should allow for a record of its inclusion on the typed transcript.

In spite of problems with the coding guidelines, the feature Mode of Communication was coded with a reasonable level of measurement reliability, with an interrater agreement figure of 77.6% emerging from the data. The only divergent cell of the judgement matrix that could possibly be described as well-populated (accounting for 9.4% of coding occurrences), crosstabulates my verbal/tonal/postural judgements with the verbal judgements of the second coder. This coding pattern can likely be attributed to the fact that I took into account the presence of postural and tonal information on typed transcripts, while the second coder more likely responded according to the letter of the coding guidelines.

As to the frequency count of the categories of Mode of Communication, as expected the verbal category proved to be the most heavily represented, accounting for 37.5% of the 809 coding occurrences. The verbal/tonal and verbal/tonal/postural categories are also adequately represented, accounting respectively for 28.7% and 19.8% of the distribution, while the nonapplicable category accounts for 11.4%. Since the remaining three categories combined account for less then 3%, they should be collapsed into a single 'other' category. What this distribution indicates is that, while failing to explore the differential effects of modes of communication on interactional data, family researchers havae not been a restrictive in their use of the modes as critics have suggested.

C. Subgroup Comparison

Two final features relating to research focus were incorporated into the present thesis, both of which are concerned with the bases against which diagnostic comparisons are made. The first of these, referred to as Subgroup Comparison, applies only to those measures that are taken on subgroups of the family unit participating in research. Such measures can be differentiated into two categories: 1) those that compare diagnostic categories in terms of calculations derived from comparisons between family subgroups and 2) those that compare diagnostic categories in terms of measures taken on particular

subgroups independent of their relationship to other subgroups. An example of the former category would be a measure that first compares mother-child interactions with father-child interactions within each diagnostic category and then compares diagnostic categories in terms of the resulting ratios. An example of the second category would be a measure that compares diagnostic categories in terms of the behavior of a specific family member (e.g., normal father vs. clinical father). This distinction may be important insofar as it may to some degree reflect the extent to which the notion of the interrelatedness of the family system is incorporated into a research design.

The frequency distribution of the feature Subgroup Comparison suggests that it be excluded from the feature set on the grounds of insufficient variability. Of the 547 applicable coding occurrences, only 47 (8.6%) involve subgroup comparisons.

As to the feature's reliability, coders were unable to differentiate measures with any reasonable degree of consistency according to whether or not they involve subgroups of the family unit. An interrater agreement ratio of only 61% emerged from the data in spite of the straightforward nature of the study feature. This strongly suggests that the coding quidelines, presented on the coding form without explanation or elaboration, are too spartan for coding purposes. Of the 214 coding occurrences that were

consistently judged as relevant to the feature Subgroup Comparison, coders had no trouble agreeing on the nature of the subgroup comparison, as is evident in the healthy 92.5% agreement figure.

D. Response Category Comparison

The second comparison feature also groups measures into two categories: 1) those that involve comparisons between more than one response category (e.g., ratio of positive social-emotional to negative social-emotional communications); and 2) those that involve a single or composite response category independent of its relationship to other categories. This feature was incorporated into the research on the grounds of its intuitive appeal. Underlying the use of measures that incorporate response-category comparisons is an inchoate understanding of the desirability of examining constellations or profiles of interactional attributes rather than single attributes extracted from their interactional context.

The coding of Response Category Comparison presented no problem as is evident from the interrater agreement figure of 91.8%. The frequency count of the feature yielded the following information. Of the 809 coding occurrences, there are 139 (17.2%) comparison-present judgements, 622 (76.9%) comparison-absent judgements, and 48 (5.9%) nonapplicable judgements. Thus, family researchers have tended on the

whole to focus on one interactional attribute at a time during statistical analysis of family process data.

To summarize, of the four features relating to research focus, both Mode of Communication and Response Category Comparison were judged to be acceptable candidates for the purposes of further integrative procedures. Subgroup Comparison, on the other hand, was rejected on the grounds of insufficient variability. The feasibility of classifying the interaction literature into substantive domains is uncertain, given that the associated problems present no clear-cut solution. A summarization of features relating to research focus is contained in Table 4.

Assessment of Redundancy

In the present research, 35 features were selected for this potential utility as predictors for a meta-analysis of the family interaction literature. Of these, 25 were judged to be suitable for such procedures with respect to the criteria of information availability, feature category variability, and measurement reliability. As a final assessment of the feature set an effort was made to determine the degree to which the features are interdependent, with the objective of weeding out from the set highly redundant elements.

Table 4

Features Relating to Research Focus

Feature	Problematic Aspect	Measure-Specific Reliability	Coding Revisions Recommended	Retain/ Exlude
Substantive Domain	Reliability concerns attributed as much to problems inherent in the literature as to inadequate coding conventions.		Uncertain	Uncertain
Boundaries of Authority		81%		
Domineering Expression		39%		
Family Cohesion		63.6%		
Affective Expression		98.8%		
Communication Clarity		80%		
Flexibility/ Rigidity		16.7%		
Family-as-a-Task Oriented-Unit		64.7%		
Quantity of Communication		100%		
Communication Continuity		83.3%		
Nonscorable		40%		
Mode of Communication	----------	77.6%	Yes	Retain
Subgroup Comparison	Overrepresentation by No-Comparison category.	92.5%	No	Exclude
Response Category Comparison	Overrepresentation by No-Comparison category.	91.8%	No	Retain

An intercorrelation matrix was constructed for the purposes of this assessment into which was entered study features that 1) satisfy the aforementioned suitability standards, and 2) assume either dichotomous or continuous numerical formats. With respect to this later criteria, an attempt was made to increase the number of entries into the matrix by excluding the nonscorable and nonapplicable coding categories from calculations involving features that are otherwise defined in dichotomous terms (e.g., Sex-of-the-Focal-Child, Control (No Control) for observer Expectancies, Presence (Absence) of Reliability Coefficient). While this manipulation likely affected the accuracy of the data, I felt that the increase in the amount of relevant information justified its use.

Examination of the correlational matrix reveals nine relationships of moderate size, with absolute values ranging from .51 to .67. Pairs of features relating to methodological rigor are involved in seven of these relationships. What these moderate correlations suggest is that, to some degree, concern with experimental control dictates methodological choices across a variety of study features. A lack of such concern is of course reflected across the same set of features. In two of the nine cases, the feature Diagnostic Category is shown to be moderately related to indices of methodological rigor, namely Observer Expectancies and Presence of Reliability Coefficient. The

observation that methodological adequacy differs as a function of the diagnostic status of experimental families has already been made by Jacob (1975) who accounts for the relationship in terms of changes in methodological practice. Specifically, the more recent family studies are more likely than the earlier ones to utilize nonpsychotic samples rather than schizophrenic samples, and to incorporate protection against methodological violations into their designs.

In addition to the nine moderate relationships one other relationship emerged out of the intercorrelation procedures that is, in my opinion, of large enough magnitude to justify the exclusion on the grounds of redundancy, namely the -.92 correlation between the composite feature Naturalistic/Controlled and one of its component features, Inference Level. Given that all of the information contained in the composite feature is represented elsewhere, a decision was made to exclude it from the feature set.

DISCUSSION

The present thesis incorporated a number of tasks, each of which was linked to a specific objective of research integration. The following discussion will be organized around these tasks and objectives.

Feature selection for the purposes of descriptive summary. A major task of the thesis was to select a set of features characteristic of the family process literature that would allow, first, for the catagorization and summarization of methodological and procedural aspects of interest and, second, for the establishment of a set of predictor variables suitable for use during later stages of integration. The importance of the first of these functions lay in its capacity to reveal the methodological preferences of family process researchers, as well as to pinpoint aspects of research that have been both over and underused. Such information was readily extracted from the frequency distributions of feature categories. On the basis of these distributions, the family process literature can be characterized in the following terms: 1) an overreliance on moderately structured tasks that assume the form either of variations of the RDT or the Consensual Projective; 2) an overreliance on the laboratory as a setting within which to

conduct interactional research; 3) a strong preference for molecular, single-attribute aggregate measures, coupled with infrequent use of molar rating scales, of measures that incorporate sequential information, and of measures that involve comparisons of behavioural categories and of family subgroups; 4) a heavy reliance on process measures as the primary source of interactional information to the neglect of product measures; and 5) a strong preference for the family triad as the participating unit of research, coupled with some attention to both the spousal dyad and the family tetrad. The family process literature can in addition be characterized in terms of its attention to a number of modes of communication (though not to comparisons between modes) as well as to a number of diagnostic categories.

What is suggested by the frequency data is a literature decidedly uniform in its choice of methodological and procedural practices, though less so in choice of research foci. Such uniformity is suprising, in light of the fact that critics of the literature have consistently described it in terms of its nonuniformity. Such discrepant impressions may simply reflect differences in foci, with critics having emphasized relatively subtle procedural and methodological details, and I, the broad boundaries of the research field (i.e., the choice of interactional task and of measurement instrument).

Regardless of the degree of uniformity, what is obvious is that family process researchers have been decidedly rigid in their use of methodology and paradigm. Such rigidity can likely be attributed to a number of factors. An emphasis on 'scientificism' that has held sway throughout the history of family interaction research may account for the overuse of molecular measures as well as for the reluctance to move outside of the controlled laboratory setting. As already indicated in the discussion on Family Composition, inadequate exploration of the full range of family subsystems is suggestive of a misunderstanding of interactional principles. Finally, the failure of the domain to evolve a varied and flexible research practice may reflect a protective response against the full weight of the demands imposed by the nature of the research. Confronted with the complexity of the subject matter, the highly abstract level of the theoretical context, and the absence of an appropriate methodological tradition, the individual engaged in interactional research may (understandably) shun the risks of experimentation in favor of 'known' procedures. That no evidence has yet emerged supporting the utility of any particular research practice underscores the inappropriateness of this response. Moreover, it is particularly ill-suited to the existing developmental stage of family interaction research.

Feature selection for the establishment of a set of predictor variables. In order to evaluate the appropriateness of the selected study features for use as predictor variables, three criteria were brought into effect. First, a sufficient amount of relevant information had to be available from the collected studies. To a large degree, protection against undue missing information was built into the present research effort. Coding guidelines were at times manipulated to allow the coder to estimate feature values from ambiguous and/or incomplete reports. For example, the conventions for the feature Developmental Stage allow the coder to estimate age ranges from central tendency data. An additional protection against missing information derived from the method of selecting both features and feature categories; namely one that was based as much on what was known to be represented in the literature as on methodolocial and theoretical concerns. Whether because of the availability of relevant information and/or such protective measures, only two of the 35 features initially selected for consideration were excluded on the grounds of information scarcity. Both are indices of the composite feature Subject Expectancies.

Sufficient variablility was the second criterion against which features were judged as suitable for the purposes of integration. While imbalances in frequency distributions are more the rule than the exception, only four of the 35

features meet the exclusion standard of at least 90% representation by a single feature category; namely, Setting, Developmental Stage, and two of the indices the composite feature of Diagnostic Specificity.

Feature redundancy, expressed in terms of a high level of interrelatedness between features, constituted the third criterion for exclusion. While sizable correlations of the magnitude of .5 to .6 are not uncommon, there is only one correlation that I consider of large enough magnitude to warrant exclusion on the grounds of feature redundancy; namely the -.92 correlation between the features Naturalistic/Controlled and Inference Level. On the basis of this figure, a decision was made to eliminate the composite feature Naturalistic/Controlled from the set of potential predictor variables. In terms of the three designated criteria, the selection of study features for the purposes of integration proved to be essentially sound, resulting in the retention of 28 of the 35 features initially brought under consideration.

Serviceablility of the coding conventions. A second major task of the present research was to create a coding system that would allow for the operationalization of study features, as well as for their translation into numerical terms. The serviceability of the system was assessed with reference to measurement reliability by examining the degree to which the judgements of two independent coders converged.

Measurement reliability in the present research was jointly dependent on three factors, the first of which was the clarity and comprehensiveness of the information provided by the primary researcher. In cases in which inadequate information could be attributed to the sloppy reporting style of the researcher, the inclusion of any relevant details in a study often allowed the coder to estimate feature values. More problematic were the cases in which inadequate information was suggestive of conceptual confusion.

Measurement reliability was also dependent on the clarity and comprehensiveness of the coding conventions, determined not only by my skill in developing unambiguous operational definitions but also by the ease with which the selected study features lent themselves to numerical translations. In some cases the clarity of the feature's conceptual underpinnings influenced my capacity to incorporate nonambiguous distinctions into the conventions. Whenever nonreliability was attributed to my lack of skill, suggestions for coding revisions were offered, with the objective of facilitating the coding task and of improving reliability levels. Of course I had no such power of remediation over problems inherent in the literature.

Finally, measurement reliability was dependent on the degree to which the coding conventions were appropriately applied to the literature. With regard to this factor, it

was obvious during training sessions that measures taken to insure the consistent application of the coding conventions were less than adequate. To begin with, the unavailability of established criterion protocols resulted in training sessions that were not sufficiently systematic. The absense of such protocols in addition brought into question the meaningfulness of the reliability assessment, insofar as no standards existed against which to evaluate the adequacy of either coders' judgements. The meaningfulness of the assessment was also somewhat compromised by the fact that the second coder was involved in the construction of the coding scheme, and that a 'naive' coder was at no point introduced into the procedures. Finally, the fact that no retraining sessions were introduced at any point to protect against observer drift may have reduced measurement reliability.

While aware of all of these limitations, I had anticipated that the reliability assessment would be helpful in pinpointing weaknesses both in the literature and in the coding conventions, as well as in providing a reasonable estimation of the serviceability of the conventions in their present form. These expectations, in my opinion, were adequately met. As to the reliability of the coding scheme, moderate to high levels emerged from the assessment for most of the study features. Only two of the 35 features were excluded on the grounds of inadequate reliability.

Evaluation of the suitability of the family process literature for quantitative integration. The final major objective of the present thesis was to evaluate the preparedness of the family process literature for integration, and more specifically, the appropriateness of meta-analysis as the integrative strategy of choice. Serving as guidelines for this evaluation were the criteria for meta-analysis suggested by Cook and Leviton (1980) and Strube and Hartmann (1983); namely, that of sufficient complexity, and of methodological, statistical, and conceptual adequacy.

Given that methodological and procedural rather than conceptual considerations were emphasized in the selected set of study features, there was limited data available with which to evaluate the conceptual adequacy of the family interaction literature. Consequently, impressions that emerged during the coding process played an important role in the evaluation. Furthermore, given that no attempt was made to disaggregate the literature on the basis of statistical considerations, the evaluation of the literature's statistical status was based entirely on impressions. Thus, both quantitative and qualitative components entered into an evaluative conclusion; namely one that questioned the appropriateness of futher meta-analytic procedures on the grounds of both statistical and conceputal concerns. Each of the four specified criteria will now be commented upon in turn.

Complexity of the literature. In spite of an overall restricted use of research strategy, it is my opinion that sufficient variability in methodological detail exists in the family literature to warrant the use of complex qualitative approaches to integration. That family interaction research is marked by variation in methodological features is to be expected. To begin with, the fact that the research is of a clinical naature insures that certain methodological choices (e.g., subject characteristics, control for expectancy effects, setting) are as much dictated by prevailing practical exigencies as by research-based considerations. To the extent that these exigencies vary from study to study, so do the affected methodological features. Moreover, in light of the breadth of the subject matter, investigators can only incorporate a small number of the relevant factors into their research designs. Given the absence of facilities that coordinate family interaction research, such choices are bound to vary considerably from study to study. As to the restricted use of research strategy and paradigm, it impacts more on the generalizability of findings that emerge from any integrative effort than on the complexity of the integrative task.

Methodological adequacy. The frequency distributions of feature categories provided the information needed to evaluate the methodological adequacy of the literature.

Methodological inadequacy, defined in terms of violations that pervade the data base, are readily detected in the form of highly invariable distributions that consist predominantly of coding judgements indicative of methodological inadequacy. Of the 17 features chosen to index methological rigour, only four yielded highly invariable distributions. Of these, only one feature was consistently judged as methodologically inadequate: namely, Manner of Solicitation, an index of the composite feature Subject Expectancies. Thus, on the basis of the frequency information, there is nothing to suggest methodological weakness pervasive enough to disqualify the family process literature for integration. It is important to note that this conclusion does not in any way contradict the frequency data that clearly points to a need for improved methodological control on many of the selected dimensions.

The degree to which family interaction researchers should focus on methodological rigor is a source of debate at the present time. That high levels of rigor can strip research of its validity by focusing on relationships that bear little resemblance to the subject matter of interest has alsways been a source of concern among students of family process. At the same time, however, the appeal of scientific rigor (and hence respectability) has exerted a strong force among individuals who have been anxious to substantiate their deeply held convictions. The fact that

the possibility of rejecting the scientific study of family interaction on the grounds of it being unamenable to empirical techniques was raised soon after the emergence of the research domain (see Frank, 1965; Lennard, et. al., 1965) may have tipped the balance in favor of "scientifism".

The recent questioning of the appropriateness of the prevailing balance is expressed largely in terms of its unsuitability to the current stage of development of the research area. More specifically, there is some concern that the emphasis on rigor has interfered both with the development of conceptual clarity and with the creation of research tools appropriate for generating and measuring interactional material (see Liem, 1980; Olson, 1983; Riskin, 1983 Turk, & Bell, 1972). In response to such concern, there is emerging evidence of attempts to redirect family interaction research into paths more in keeping with an exploratory stage of a research history than with a hypothesis-testing stage. At a recent conference on family interaction held at the University of Manitoba (May, 1983) both Riskin and Olson recommended the use of the following 'regressive strategies: 1) the use of the case-study approach to research rather than the standard large N design; 2) a reliance on relatively nonrigorous measurement tools such as the "quick and dirty" rating scale and the self-report inventory; 3) a rejection of "safe" established interactional tasks and coding schemes, in favor of rigorous

experimentation with alternatives created specifically to meet the needs of idiosyncratic research objectives; and 4) an examination of "normal" family processes independent of any comparisons with clinical processes. Reflected in all of these recommendations is a primary emphasis on methodological development coupled with a willingness to compromise on methodological rigor.

Statistical adequacy. A number of impressions emerged during the course of coding the family interaction literature that jointly point to pervasive statistical weakness in the data base. Such weakness can first of all be attributed to the practices of the primary researchers. Particularly vulnerable to attack are the numerous studies that apply inferential techniques to data derived from multiple-measure coding schemes, without introducing any protection against experiment-wise error. Given that statistically significant inferential statistics generally provide the data for ES estimations, one can expect such studies to yield ESs that are highly inflated in magnitude.

While a trend towards more rigorous statistical practice was apparent to me during my survey of the literature, the manner in which the more competent statisticians report their findings does not always allow the meta-analyst to take advantage of the statistical advances in the field. Particularly troublesome are those studies that report inferential statistics derived from multifactorial designs

without providing either the group means or the multiple comparison figures needed for ES calculations.

Unworkable reporting styles are not only restricted to multifactorial designs or, for that matter, to any particular set of family process studies. Rather, studies typically contain both usable and nonusable findings. Particularly irksome are those studies that yield only a small portion of their outcomes to a meta-analytic data base. Though perhaps somewhat extreme, a study by Haley (1967) that yields only 3 of its 75 findings cannot be viewed as atypical. The fact that a large number of the published findings of family process research are excluded from a meta-analytic data base on statistical grounds insures an inaccurate integration.

Another related feature of data reporting that generates statistical concern is strong tendency on the part of family process researchers to restrict their treatment of nonsignificant findings to verbal statements. A meta-analyst can respond to this situation in one of two ways, both of which introduce bias into the integrative procedures. S/he can exclude nonsignificant findings from the data base and, in so doing, insure average ES calculations of inflated magnitude. Alternatively, s/he can assign to each nonsignificant finding an ES value of '0' and, in so doing, insure average ES value of conservative magnitude -- highly conservative, I would guess, given the

prevalence of nonsignificant findings in the collected literature.

There is one final aspect of the family process data base that must be considered in an evaluation of its statistical adequacy; namely, the degree to which its data points are independent of one another. As indicated in the chapter on Integrative Techniques (see p. 73), a meta-analyst working with a psychological literature can routinely expect some degree of nonindependence to plague his/her efforts. While no guidelines have emerged regarding the degree to which nonindependence can be tolerated, it is reasonable to assume that, at a certain point, the interpretability of any inferential statistic performed during integration becomes seriously compromised. While the clinical nature of family process research led me to anticipate considerable overuse of both samples and measures, the degree to which nonindependence in fact characterizes the data base possibly exceeds any workable limit. Focusing only on the overuse of measures as a source of nonindependence, the 45 collected studies can be characterized in the following terms: The number of outcomes per study ranges in value from 1 to 98, with 31 of the 46 studies contributing at least 10 outcomes each to the data base. The average number of outcomes per study is 18. It is worth pointing out that the degree of nonindependence in Smith and Glass's meta-analysis of the therapy outcome literature that promted Landman and Dawes

(1982) to reintegrate the literature, was of the magnitude of 2.2 outcomes per study.

To conclude, my impressions of the statistical status of the family process data base lead me to question the appropriatenss of meta-analysis as the integrative strategy of choice for the family interaction literature. Such questioning primarily arises from my perception of a marked mismatch between the characteristics of the data base (i.e., unavailability of precise information, high degree of nonindependence) and the characteristics of meta-analytic techniques (i.e. dependence on specific statistical information, vulnerability to violation of assumption of independence). While there is little that any integrative strategy could do to override the statistical invalidity generated by the practices of the primary researcher short of gaining access to much of the original data, an integrative strategy less exact than meta-analysis would be better able to circumvent some of the other statistical concerns that have been discussed. Such a strategy would, of course, have to improve upon those relied upon in past reviews to justify another 'go' at the literature, by incorporating as many as possible of the relatively rigorous procedures available to traditional reviews that are outlined by Cook and Leviton. Of course in choosing a more traditional integrative strategy for the family interaction literature, the reviewer must be willing to balance out

gains in the amount and representativeness of the useable data, with a restricted capacity regarding the number of study features that can be brought into consideration.

Conceptual adequacy. While my familiarity with critiques of the literature had prepared me to expect some level of conceptual weakness to mark the family literature, the extent of the weakness only became apprarent during my own efforts to develop and apply the coding guidelines for the study feature that bears most directly on conceptual concerns, that is Substantive Domain. As previously indicated, Doane's recommended reclassification strategy did less to facilitate this effort than I had anticipated. Limitations to her stategy may reside in the inaccuracy of its underlying assumption; specifically that correspondences between instructional operations and constructs need only be teased out through the appropriate reorganization of the literature. Concerns with the conceptual status of the family literature were further strengthed during my attempt to apply the coding guidelines for the other two features that bear on conceptual adequacy (i.e., Unit of Measurement and Subgroup Comparisons), as well as during my efforts to record the conceptual designation of the primary researchers.

In Spite of the above indications of conceptual weakness, a number of the categories of Substantive Domain were coded with high levels of reliability regarding the coding of any one such domain is not to be confused with the domain's

capacity to differentiate clinical and nonclinical families. Nonetheless, the fact that I was able to devise coding conventions that can reliably order portions of the family interaction literature on the basis of conceptual considerations suggests that some degree of conceptual clarity is operating. Whether that degree is sufficient to warrant futher integrative procedures must be left to the judgement of the individual reviewer.

In conclusion, much of the impetus for the present integrative effort arose out of a hope of being able to support my ideological committment to family systems theory with empirical evidence. Given my understanding of the level of methodological development in the field, I did not expect definitive patterns linking clinical status to family interaction to emerge from the data. What I did expect was to be able to provide confident direction to future research by initiating meta-analytic procedures that would allow me to empirically isolate those aspects of past research that have been particularly productive. I had furthermore hoped that such direction would generate renewed interest in a putatively moribund research domain, and in so doing, allow for the possiblity of narrowing the gap between systems research and clinical practice.

In light of the breadth of the integrative objective, the present research effort was restricted to a number of preliminary tasks designed to prepare the groundwork for futher stages of integration. By and large this objective

was successfully met; the research yielded both useful descriptive information and a set of serviceable coding conventions. However, concerns with statitical and conceptual adequacy generated some questions regarding the choice of integrative strategy most appropriate for the family interaction literature.

REFERENCES

Ackerman, N.W. (1958). Psychodynamics of family life, diagnosis and treatment in family relationships. New York: Basic.

Alexander, J. F. (1973). Defensive and supportive communication in normal and deviant families. Journal of Consulting and Clinical Psychology, 40, 223-231.

Attneave, C.L. (1976). Social networks as the unit of intervention. In P.J. Guerin Jr. (Ed.), Family therapy: Theory and practise. New York: John Wiley & Sons.

Bateson, G. (1959). Cultural problems posed by a study of schizophrenia process. In A. Averback (Ed.), Schizophrenia: An integrated approach. New York: Ronald Press.

Bateson, G. (1966). Slippery theories. International Journal of Psychiatry, 2, 415-418.

Bateson, G., Jackson, D.D., Haley, J. & Weakland, J. (1956). Towards a theory of schizophrenia. Behavioral Science, 1, 251-264.

Beihl, H.O. (1979). The family as patient: Theoretical, practical, ethical and political implications. Unpublished paper, University of Manitoba,

Bell, R.O. (1964). Structuring parent-child interaction situations for direct observation. Child Development, 35, 1009-1020.

Bertalanffy, L. von. (1956). General system theory. General systems, 1, 1-10.

Bronfenbrenner, V. (1977). Towards an experimental ecology of human development. American Psychologist, 32, 513-531.

Bugental, D. E., Love, L. R., & Kaswan, J. W. (1972). Videotaped family interaction: Differences reflecting presence and type of child disturbance. Journal of Abnormal Psychology, 79, 285-290.

Burgum, M. (1942). The father gets worse: A child guidance problem. American Journal of Orthopsychiatry, 12, 474-486.

Chase, L.J. & Chase, R.B. (1976). A statistical power analysis of applied psycholgical research. Journal of Applied Psychology, 61, 234-237.

Cheek, F.E. (1964). A serendipitous finding: Sex roles and schizophrenia. Journal of Abnormal Psychology, 69, 392-400.

Cheek, F.E. (1965). The father of the schizophrenic. Archives of General Psychiatry, 13, 336-345.

Cohen, J. (1977). Statistical power analysis for the behavioral sciences. New York: Academic Press.

Cook, T.D. & Leviton, L.G. (1980). Reviewing the literature: A comparison of traditional methods with meta-analysis. Journal of Personality, 48, (4), 449-472.

Cooper, H.M. (1979). Statistically combining independent studies: A meta-analysis of sex differences in conformity research. Journal of Personality and Social Psychology, 37, 131-145.

Doane, J.A. (1978a). Family interaction and communication deviance in disturbed and normal families: A review of research. Family Process, 17, 357-376.

Doane, J.A. (1978b). Questions of strategy: Rejoiner to Jacob and Grounds. Family Process, 17, 389-394.

Eysenck, H.J. (1978). An exercise in mega-silliness. American Psychologist, 33, 517.

Farina, A. (1960). Patterns of role dominance and conflict in parents of schizophrenic patients. Journal of Abnormal and Social Psychology, 61, 31-38.

Farina, A., & Holzberg, J.D. (1968). Interaction patterns of parents and hospitalized sons diagnosed as schizophrenic or nonschizophrenic. Journal of Abnormal Psychology, 73, 114-118.

Ferber, A., Mendelsohn, M., Napier, A. (1973). The book of family therapy. Boston: Houghton Mifflin.

Feldman, L.B. (1976). Goals of family therapy. Journal of Marriage and Family Counseling, 2, 103-113.

Ferreira, A.J. (1963a). Rejection and expectancy of rejection in families. Family Process, 2, 235-244.

Ferreira, A.J. (1963b). Decision making in normal and pathological families. Archives of General Psychiatry, 8, 68-73.

Ferreira, A.J. (1977). Family myths. In P. Watzlawick, J.H. Weakland (Eds.), The interactional view. New York: W.W. Norton Company, Inc.

Ferreira, A.J., & Winter, W.D. (1965). Family interaction and decision making. Archives of General Psychiatry, 13, 214-223.

Ferreira, A.J. & Winter, W.D. (1968a). Information exchange and silence in normal and abnormal families. Family Process, 7, 251-276.

Ferreira, A.J. & Winter, W.D. (1968b). Decision making in normal and abnormal two child families. Family Process, 7, 17-36.

Ferreira, A.J., Winter, W.D., & Poindexter, E.J. (1966). Some interactional variables in normal and abnormal families. Family Process, 5, 65-70.

Foley, V.D. (1978). An introduction to family therapy. New York. Grune & Stratton.

Fontana, A.F. (1966). Family etiology of schizophrenia: Is a scientific methodology possible? Psychological Bulletin, 66, 214-224.

Fox, G.L.K. (1978). Home and laboratory observations of parent-child interactions: A methodological comparison. Master's Thesis, University of Manitoba.

Framo, J.L. (1972). Symptoms from a family transactional viewpoint. In C.J. Sager & H.S. Kaplan (Eds.), Progress in group and family therapy. New York: Brunner/Mazel.

Framo, J.L. (1975). Personal reflections of a family therapist. Journal of Marriage and Family Counseling, 1, 15-28.

Framo, J.L. (1976). Family of origin as a therapeutic resource for adults in marital and family therapy: You can and should go home again. Family Process, 15, 193-210.

Frank, G.H. (1965). The role of the family in the development of psychopathology. Psychological Bulletin, 64, 191-205.

Freud, S. (1917). Introductory lectures on psychoanalysis, part III, lecture 28: analytic therapy. Standard Edition, 16, 448-463.

Gage, N.L. (1976). The scientific basis of the art of teaching. New York: Teachers College Press.

Gallo, P.S. (1978). Meta-analysis - a mixed meta-phor? American Psychologist, 33, (May), 515-517.

Glass, G. (1976). Primary, secondary and meta-analyses of research. Educational Researcher, 5, (10), 3-8.

Glass, G. (1977). Integrating findings: The meta-analysis of research. In D.S. Shulman (Ed.), Review of research in education, 5, 351-379.

Glass, G. (1978). Policy for the unpredictable (uncertainty research and policy). Educational Researcher. 8, (9) 12-14.

Glass, G. (1980). Summarizing effect sizes. New Directions for Methodology of Social and Behavioral Science, 5, 13-31.

Glass, G. & Smith, M. (1978). Reply to Eysenck. American Psychologist, 33, 517-519.

Glass, G. & Smith, M. (1980). Ask not for whom the bell tolls. American Psychologist, 35, (Feb.), 223.

Goldfried, M. (1977). Behavioral assessment in perspective. In I.D. Cone & R.P. Hawkins (Eds.), Behavioral assessment: New directions in clinical psychology. New York: Brunner/Mazel

Goldstein, M. & Doane, J.N. (1982). Family factors in the onset, course, and treatment of schizophrenic spectrum disorders. The Journal of Nervous and Mental Disease, 170, 692-700.

Goldstein, M., Judd, L., Rodrick, E., Alkice, A., & Gould, E. (1968). A method for studying social influence and coping patterns within families of disturbed adolescents. Journal of Nervous and Mental disease, 147, 233-251.

Goldstein, M., Rodnick, E. (1975). The family's contribution to the etiology of schizophrenia: Current status. Schizophrenia Bulletin, 14, 48-63.

Goodrich, D.W. & Boomer, D.S. (1963). Experimental assessment of modes of conflict resolution. Family Process, 2, 15-24.

Group for the Advancement of Psychiatry (GAP). (1970). The field of family therapy. New York: Science House.

Gurman, A.S. (1975). Some therapeutic implications of marital therapy research. In A. Gurman & D. Rice (Eds.), Couples in conflict. New York: Jason Aronson.

Guttman, H.A. (1977). A guide to family function and structure. Unpublished paper. Jewish General Hospital, Montreal.

Haley, J. (1963). Family experiments: A new type of experimentation. Family Process, 1, 265-293.

Haley, J. (1963). Strategies of psychotherapy. New York: Grune & Stratton.

Haley, J. (1964). Research on family patterns: An instrument measurement. Family Process, 3, 41-65.

Haley, J. (1967). Speech sequence of normal and abnormal families with two children present. Family Process, 6, 81-97.

Haley, J. (1967). Experiments with abnormal families. Archives of General Psychiatry, 17, 53-65.

Haley, J. (1972). Critical overview of present status of family interaction research. In J.L. Framo (Ed.), Family interaction A dialogue between family researchers and family therapists. New York: Springer Publishing Company.

Haley, J. (1973). Uncommon therapy. New York: Ballantine.

Haley, J. (1976). Problem solving therapy. San Fransisco; Jossey-Bass.

Hansen, C. (1968). An extended home visit with conjoint family therapy. Family Process, 7, 67-87.

Hedges, L.V. (1981). Distribution theory for Glass's estimator of effect size and related estimates. Journal of Educational Statistics, 6, 107-128.

Hetherington, E.M., Stouwie, R.J., & Ridberg, E. (1971). Patterns of family interaction and child rearing attitudes related to the dimensions of juvenile delinquency. Journal of Abnormal Psychology, 18, 160-176.

Hill, R. (1964). Methodological issues in family development research. Family Process, 3, 186-206.

Hill, R. (1965). Decision-making and the family life cycle. In E. Shanas & C.F. Streib (Eds.), Social structure and the family generational relations. Englewood Cliffs, N.J.: Prentice-Hall.

Hughes, H.M. & Haynes, S.N. (1978). Structural laboratory observation in the behavioral assessment of parent-child interaction: A methodological critique. Behavior Therapy, 2, 428-447.

Jackson, D.D. (1959). Family interaction, family homeostasis and some implications for conjoint family psychotherapy. In J. Massernan (Ed.), Individual and family dynamics. New York: Grune & Stratton.

Jackson, D.D. (1965). The study of the family. Family Process, 4, 1-20.

Jackson, D. (1977). Family rules: Marital quid pro quo. In P. Watzlawick & J.H. Weakland (Eds.), The interactional view. New York: W.W. Norton & Company.

Jacob, T. (1974). Patterns of family conflict and dominance as a function of child age and social class. Developmental Psychology, 10, 1-12.

Jacob, T. (1975). Family interaction in disturbed and normal families: A methodological and substantive review. Psychological Bulletin, 82, 33-65.

Jacob, T. & Davis, J. (1973). Family interaction as a function of experimental task. Family Process, 12, 415-427.

Jacob, T. & Grounds, L. (1973). Confusions and Conclusions: A response to Doane. Family Process, 17, 337-391.

Jones, F. (1973). Current methodologies for studying the development of schizophrenia: A critical review. Journal of Nervous and Mental Disease, 157, 154-178.

Kalat, J. (1980). A misuse of statistics: Reply to Rimland. American Psychologist, 35, (Feb.), 223-224.

Kavale, K. & Glass, G. (1981). Meta-analysis and the integration of research in special education. Journal of Learning Disabilities, 14, (9), 531-538.

Kent, R.N. & Foster, J.L. (1979). Direct observational procedures: Methodological issues in naturelistic settings: In A.R. Ciminero, K.S. Calhoun & H.E. Adams (Eds.), Handbook of Behavioral Assessment. New York: JOhn Wiley & Sons.

Klopper, E.J., Tittler, B.I., Friedman, S., & Hughes, S.T. (1978). A multi-method investigation of two family constructs. Family Process, 17, (3), 83-93.

Krahn, G.L. (1980). The complimentarity of continuous and discrete state analysis of family interaction. Unpublished manuscript, University of Manitoba.

Kulik, J.A., Kulik, C.C. & Cohen, P.A. (1979). Research on Audio-Tutorial instruction: A meta-analysis of comparitive studies. Research in Higher Education, 11, 321-341.

Landman, J. & Dawes, R. (1982). Psychotherapy outcome. American Psychologist, 37, 504-516.

Lennard, H.L. (1961). Analysis of family conflict. In N. Ackerman, F. Beatman, & S.N. Sherman (Eds.), Exploring the base for family therapy, New York: Family Service Association of America.

Lennard, H.L., Beaulieu, M.R., & Embrey, N.G. (1965). Interaction in families with a schizophrenic child. Archives of General Psychiatry, 12, 166-183.

Liem, J. (1980). Family sudies of schizophrenia: An update and commentary. Schizophrenia Bulletin, 16, 429-453.

Lidz, T. (1963). The family and human adaptation. New York: International University Press.

Light, R. (1978). Capitalizing on variation: How conflicting research findings can be helpful for policy. Educational Psychology, 8, (9), 7-11.

Light, R. & Smith, P. (1971). Accumulating evidence: Procedures for resolving contradictions among different research studies. Harvard Educational Review, 41, 429-471.

Lytton, H. (1971). Observation studies of parent-child interaction: A methodological review. Child Development, 42, 651-684.

Lytton, H. (1973). Three approaches to the study of parent-child interaction: Ethological interview and experimental. Journal of Child Psychology and Psychiatry, 14, 1-17.

Mansfield, R.S. & Busse, T.V. (1978). Meta-analysis of research: A rejoinder to Glass. Educational Researcher, 6, 3.

Marshall, J.R. & Neill, J. (1977). The removal of a psychosomatic symptom: Effects on the marriage, Family Process, 16, 273-280.

McGaw, B. & Glass, G.V. (1980). Choice of the metric for effect size in meta-analysis. American Educational Research Journal, 17, (3), 325-337.

Minuchin, S. (1974). Families and family therapy. Cambridge, Mass.: Harvard University Press.

Minuchin, S., Baker, L., Rosman, L., Lietman, R., Milman, L., Todd, T.C. (1975). A conceptual model of psychosomatic illness in children. Archives of General Psychiatry, 321, 1031-1038.

Mishler, E.G. & Waxler, N.E. (1965). Family interaction process and schizophrenia: A review of current theories. Merrill-Palmer Quarterly of Behavior and Developement, 2, 269-315.

Mishler, E.G. & Waxler, N.E. (1966). Family interaction and schizophrenia. Archives of General Psychiatry, 15, 64-74.

Mishler, E.G. & Waxler, N.E. (1968). Interaction families: An experimental study of family processes and schizophrenia. experimental study of family processes and schizophrenia. New York: John Wiley and Sons Inc.

Mishler, E.G. & Waxler, N.E. (1975). The sequential patterning of interaction in normal and scphizohrenic families. Family Process, 14, 17-50.

Murrell, S.A. (1971). Family interaction variables and adjustment of nonclinic boys. Child Development, 42, 1485-1494.

Murrell, S.A. & Stachowiuk, J.G. (1967). Consistency, rigidity and power in the interactional patterns of clinic and nonclinic families. Journal of Abnormal Psychology, 72, 265-272.

Okun, B.K. & Rappaport, L.J. (1980). Working with families: An introduction to family therapy. Massachusetts: Duxbury Press.

Olson, D.H. (1970). Marital and family therapy: Integrative review and critique. Journal of Marriage and the Family, 32, 501-538.

Olson, D.H. (1983). Overview of marriage and family research. Presentation at Family Interaction Research Workshop. University of Manitoba, May.

O'Rourke, J.F. (1963). Field and laboratory: The decision making behavior of family groups in two experimental conditions. Sociometry, 26, 422-435.

Patterson, G.R., Ray, R.S., Shaw, D.D. & Cobb, J.A. (1969). Manual for coding of family interactions. Available from ASIS National Auxilleary Publications Service, c/o CCM Information Sciences, Inc., 909 third Avenue, New York, N.Y. 10022. Document # 01234.

Pillemer, D.B. & Light, R.J. (1981). Benefiting from variation in study outcomes.

Rabkin, L.Y. (1965). The patient's family: Research methods. Family Process, 4, 105-132.

Reiss, D. (1967). Individual thinking and family interaction: 11. A study of pattern recognition and hypothesis testing in families of normals, character disorders, and schizophrenics. Journal of Psychiatric Research, 5, 193-211.

Reiss, D. (1976). The family and schizophrenia. American Journal of Psychiatry, 133, 181-185.

Rimland, B. (1979). Death knell for psychotherapy? American Psychologist 34, (Feb.).

Riskin, J. (1983). Longitudinal research on nonclinical families. Presentation at Family Interaction Research workshop. University of manitoba, May.

Riskin, J. & Faunce, G. (1972). An evaluative review of family interaction research, Family Process, 11, 365-456.

Rosenthal, R. (1978). Combining results of independent studies. Psychological Bulletin, 85, 185-193.

Sanua, V.D. (1961). Socio-cultural factors in families of schizophrenics. Psychiatry, 24, 246-265.

Satir, V. (1972). Peoplemaking. Palo Alto, Science and Behavior Books,

Seeman, M.N. (1983). Schizophrenia and family studies. The Psychiatric Journal of the University of Ottawa, 8, 38-43.

Shapiro, L.N. & Wild, C.M. (1976). The product of the concensus rorschach in families of male schizophrenics. Family Process, 15, 211-224.

Sharan (Singer), Shlomo. (1966). Family interaction with schizophrenics and their siblings. Journal of Abnormal Psychology, 71, (5), 345-353.

Sluzki, C.E., Beavin, J., Tornopolsky, A. & Vernon, E. (1967). Transactional disqualification: Research on the double bind. Archives of General Psychiatry, 16, 494-504.

Smart, R.A. (1964). The importance of negative results in psychological research. The Canadian Psychologist, 5, 225-232.

Smith, M. (1980). Integrating studies of psychotherapy outcomes. New Directions for Methodology of Social and Behavioral Science, 5, 47-61.

Smith, M. & Glass, G. (1977). Meta-analysis of psychotherapy outcomes. American Psychologist, 32, 752-760.

Smith, M., Glass, G. & Miller, J. (1980). The benefits of psychotherapy. Baltimore: The John Hopkins University Press.

Smith, R.C. (1971). Verbal discussion versus note passing tasks in the study of family role structure. Journal of Nervous and Mental Disease, 152, 173-183.

Solvberg, H.A. & Blakar, R.M. (1975). Communication efficiency in couples with and without a schizophrenic offspring. Family Process, 14, 515-534.

Speck, R.V. (1967). Psychotherapy of the social network of a schizophrenic family. Family Process, 6, 208-214.

Stabenau, J.R., Tupin, J., Werner, M., Pollen, W. (1965). A comparitive study of families of schizophrenics, delinquents and normals. Psychiatry, 28, 45-59.

Steinglass, P. (1976). Experimenting with family treatment approaches to alcoholism, 1950-1975: A review. Family Process, 15, 97-123.

Strodbeck, F.L. (1951). Husband-wife interaction over revealed differences. American Sociology Review, 16, 468-473.

Strube, M. & Hartmann. (1983). Meta-analysis: Techniques, applications, and functions. Journal of Consulting and Clinical Psychology, 51, 14-27.

Turk, J.L. & Bell, N.W. (1972). Measuring power in families. Journal of Marriage and the Family, 34, 215-222.

Wahler, R.G. (1976). Deviant child behavior within the family: Developmental speculations and behavior change strategies. In H. Leitenberg (Ed.), Handbook of behavior modification and behavior therapy. New Jersey: Prentice-Hall.

Walrond-Skinner, S. (1976). Family therapy: The treatment of natural systems. London: Routledge & Regan Paul.

Watzlawick, P. (1964). An anthology of human communication. Palo Alto: Science & Behavior Books.

Watzlawick, P. (1966). A structured family interview. Family Process, 5, 256-271.

Watzlawick, P. (1977). Introduction. In P. Watzlawick & J.H. Weakland (Eds.), The interactional view. New York: W.W. Norton & Company.

Watzlawick, P. & Weakland, J.H. (1977). The interactional View. New York: W.W. Norton & Company.

Weakland, J.H. (1974). "The double-bind theory" by self-reflective hindsight. Family Process, 13, 269-277.

Wells, R.A. & Dezen, A.E. (1978). The results of family therapy revisited: The non-behavioral methods. Family Process, 17, 251-274.

Wells, R.A., Dilkes, T.C., & Trivelli, N. (1972). The results of family therapy: A critical review of the literature. Family Process, 11, 189-207.

Whitaker, C.A. (1975). A family therapist looks at marital therapy. In A.S. Gurman & D. Rice (Eds.), Couples in conflict. New York: Jason Aronson.

White. B. & Watts, J. (1973). Experience and environment: Major influences on the development of the young child. Englewood Cliffs, N.J.: Prentice-Hall, 1973.

Wiggins, J.S. (1973). Personality and Prediction: Principles of Personality Assessment. Massachusetts, Addison-Wesley.

Willson, V. (1980). The variance accounted for in meta-analysis of psychotherapy outcomes. American Psychologist, 35, 467.

Winter, W. D., & Ferreira, A. J. (1967). Interaction process analysis of family decision-making. Family Process, 6, 155-172.

Winter, W., Ferreira, A., & Olson, J. (1966). Hostility themes in the family TAT. Journal of Projective Techniques and Personality Assessment, 30, 270-274.

Wynne, L.C. (1966). New stage in family setting. International Journal of Psychiatry, 2, 418-422.

Wynne, L. (1981). Current concepts about schizophrenics and family relationships. The Journal of Nervous and Mental Disease, 169, 82-89.

Wynne, L., Ryckoff, I., Day, J., & Hirsch, S. (1958). Pseudomutuality in the family relations of schizophrenics. Psychiatry, 21, 205-220.

Zuckerman, E. & Jacobs, T. (1979). Task effects in family interaction. Family Process, 18, 47-53.

Appendix A

CITATIONS OF STUDIES AVAILABLE FOR META ANALYSIS OF FAMILY INTERACTION LITERATURE

Citations of Studies Available for Meta Analysis of Family Interaction Literature

Alexander, J. F. (1973). Defensive and supportive communication in normal and deviant families. Journal of Consulting and Clinical Psychology, 40, 223-231.

*Alkire, A. A. (1969). Social power and communication within families of disturbed and nondisturbed preadolescents. Journal of Personality and Social Psychology, 13, 335-349.

Becker, J., & Iwakami, E. (1969). Conflict and dominance within families of disturbed children. Journal of Abnormal Psychology, 74, 330-335.

*Behrens, M. L., & Goldfarb, W. (1958). A study of patterns of interaction of families of schizophrenic children in residential treatment. American Journal of Orthopsychiatry, 28, 300-310.

*Behrens, M. I., Rosenthal, A. J., & Chodoff, P. (1968). Communication in lower class families of schizophrenics: Observations and findings. Archives of General Psychiatry, 18, 689-696.

*Bugental, D. E., Love, L. R., & Kaswan, J. W. (1972). Videotaped family interaction: Differences reflecting presence and type of child disturbance. Journal of Abnormal Psychology, 79, 285-290.

*Bugental, D. E., Love, L. R., Kaswan, J. W., & April, C. (1971). Verbal-nonverbal conflict in parental messages to normal and disturbed children. Journal of Abnormal Psychology, 77, 6-10.

*Caputo, D. V. (1963). The parents of the schizophrenic. Family Process, 63, 339-356.

Cheek, F. E. (1964). The "Schizophrenic Mother" in word and deed. Family Process, 3, 155-177.

*Cheek, F. (1964). A serendiptous finding: Sex roles and schizophrenia. Journal of Abnormal Psychology, 69, 392-400.

Cheek, F. (1965). The father of the schizophrenic. Archives of General Psychiatry, 13, 336-345.

Farina, A. (1960). Patterns of role dominance and conflict in parents of schizophrenic patients. Journal of Abnormal and Social Psychology, 61, 31-38.

Faunce, E. E., & Riskin, J. (1970). Family interaction scales 11: Data analysis and findings. Archives of General Psychiatry, 22, 513-526.

Ferreira, A. J. (1963). Decision making in normal and pathologic families. Archives of General Psychiatry, 8, 68-73.

*Ferreira, A. J., & Winter, W. D. (1965). Family interaction and decision-making. Archives of General Psychiatry, 13, 214-223.

*Ferreira, A. J., & Winter, W. D. (1968). Decision-making in normal and abnormal two-child families. Family Process, 7, 17-36.

*Ferreira, A. J., & Winter, W. D. (1968). Information exchange and silence in normal and abnormal families. Family Process, 7, 251-276.

*Ferreira, A. J., Winter, W. D., & Poindexter, E. J. (1966). Some interactional variables in normal and abnormal families. Family Process, 5, 60-75.

Fisher, S., Boyd, I., Walker, D., & Sheer, D. (1959). Parents of schizophrenics, neurotics, and normals. Archives of General Psychiatry, 1, 149-166.

*Friedman, C. J., & Friedman, A. S. (1970). Characteristics of schrizophenic families during a joint story-telling task. Family Process, 9, 333-353.

Haley, J. (1963). Family experiments: A new type of experimentation. Family Process, 1, 265-293.

Haley, J. (1964). Research on family patterns: An instrument measurement. Family Process, 3, 41-65.

Haley, J. (1968). Testing parental instructions to schizophrenic and normal children. Journal of Abnormal Psychology, 73, 559-565.

Haley, J. (1967). Experiment with abnormal families: Testing done in a restricted setting. Archives of General Psychiatry, 17, 53-63.

*Leighton, L. A., Stollak, G. E., & Ferguson, L. R. (1971). Patterns of communication in normal and clinic families. Journal of Consulting and Clinical Psychology, 36, 252-256.

*Lennard, H. L., Beaulieu, M. R., & Embrey, N. G. (1965). Interaction in families with a schizophrenic child. Archives of General Psychiatry, 12, 166-183.

Lessin, S., & Jacob, T. (1979). Verbal-nonverbal congruence in normal and delinquent families. Journal of Clinical Psychology, 35, 391-395.

*Mead, D. E., & Campbell, S. S. (1972). Decision-making and interaction by families with and without a drug-abusing child. Family Process, 11, 487-498.

*Murrell, S. A. (1971). Family interaction variables and adjustment of nonclinic boys. Child Development, 42, 1485-1494.

*Murrell, S. A., & Stachowiak, J. (1967). Consistency, rigidity, and power in the interaction patterns of clinic and non-clinic families. Journal of Abnormal Psychology, 72, 265-272.

O'Connor, W., & Stachowiak, J. (1971). Patterns of interaction in families with low-adjusted, high-adjusted, and mentally retarded members. Family Process, 10, 229-241.

*Reiss, D. (1967). Individual thinking and family interaction: 11. A study of pattern recognition and hypothesis testing in families of normals, character disorders, and schizophrenics. Journal of Psychiatric Research, 5, 193-211.

Reiss, D. (1968). Individual thinking and family interaction: 111. An experimental study of categorization performance in families of normals, those with character disorders and schizophrenics. The Journal of Nervous and Mental Disease, 146, 384-403.

Reiss, D. (1969). Individual thinking and family interaction: 1V. A study of information exchange in families of normals, those with character disorders, and schizophrenics. The Journal and Nervous and Mental Disease, 149, 473-490.

Reiss, D. (1971). Varieties of consensual experience: 111. Contrasts between families of normals, delinquents, and schizophrenics. The Journal of Nervous and Mental Disease, 152, 73-95.

*Schuham, A. I. (1970). Power relations in emotionally disturbed and normal family triads. Journal of Abnormal Psychology, 75, 30-37.

*Schuham, A. I. (1972). Activity, talking time, and spontaneous agreement in disturbed and normal family interaction. Journal of Abnormal Psychology, 79, 68-75.

Schulman, R. E., Shoemaker, D. J., & Moelis, I. (1962). Laboratory measurement of parental behavior. Journal of Consulting Psychology, 26, 109-114.

Shapiro, L. N., & Wild, C. M. (1976). The product of the Consensus Rorschach in families of male schizophrenics. Family Process, 15, 211-224.

*Snyder, J. J. (1977). Reinforcement analysis of interaction in problem and nonproblem families. Journal of Abnormal Psychology, 86, 528-535.

*Sojit, C. M. (1971). The double bind hypothesis and the parents of schizophrenics. Family Process, 10, 53-74.

*Wild, C. M, Shapiro, L. N., & Goldenberg, L. (1975). Transactional communication disturbances in families of male schizophrenics. Family Process, 14, 131-160.

Winter, W. D., & Ferreira, A. J. (1967). Interaction process analysis of family decision-making. Family Process, 6, 155-172.

Winter, W. D., & Ferreira, A. J. (1969). Talking time as an index of intrafamilial similarity in normal and abnormal families. Journal of Abnormal Psychology, 74, 574-575.

*Winter, W. D., Ferreira, A. J., & Olson, J. L. (1965). Story sequence analysis of family TATs. Journal of Projective Techniques and Personality Assessment, 29, 392-397.

*Winter, W. D., Ferreira, A. J., & Olson, J. L. (1966). Hostility themes in the family TAT. Journal of Projective Techniques and Personality Assessment, 30, 270-274.

* Studies used in reliability assessment.

Appendix B

CODING FORM FOR FAMILY INTERACTION META-ANALYSIS

Coding Form for Family Interaction Meta-Analysis

General Information

Date Coded _____

Study Identification No. _____

Article Title _____

Author(s) _____

Journal _____

Year of Publication _____

Total No. of Comparisons _____

No. of this Comparison _____

Name of this Comparison _____

No. of Outcome Measures Per Comparison _____

No. of this Outcome Measure _____

Name of this Outcome Measure _____

Total N of Study (only indicate for 1st outcome of 1st comparison) _____

_____

Coder _____

FEATURES RELATING TO METHODOLOGICAL TOOLS

(see coding supplement pp. 1-4 for additional details)

A. Tasks Used to Generate Interactional Data

_____a) task structure: 1) high; 2) medium; 3) low

_____b) emotional tone of task: 1) highly emotional; 2) moderately emotional; 3) neutral

_____c) characteristics of task stimuli: 1) verbal stimuli used to generate verbally expressed differences; 2) projective stimuli used to generate verbally expressed differences; 3) general verbal exercises; 4) game situation

_____

_____

If more than one measure is recorded on this form, list features that are coded separately _____

_____

B. Outcome Measures

_____a) inference level: 1) inferential; 2) noninferential; 3) nonscorable for composite measures that are not uniform with respect to this feature

_____b) process/product distinction: 1) process of interaction; 2) product of interaction; 3) nonscorable for composite measures that are not uniform with respect to this feature

_____c) If inferential measure is used, is it in the form of a global rating scale? 1) yes; 2) no; 3) nonapplicable for noninferential measures

_____

_____

_____

_____C. Experimental setting: 1) home; 2) laboratory based in research setting; 3) hospital-based laboratory; 4) nonscorable - clinical and nonclinical samples are tested in different settings; 5) unknown.

_____D. Naturalistic/Controlled Continuum

1) controlled (if at least 2 of the following apply)
   - high structure
   - noninferential
   - laboratory setting, either hospital or laboratory based
2) naturalistic (if at least 2 of the following apply)
   - medium/low structure
   - inferential
   - home setting
3) unknown (if inference and structure coded under different control categories and setting is unknown).
4) nonscorable (if inference and structure coded under different control categories and setting is nonscorable).

_____

_____

FEATURES RELATING TO SAMPLE COMPOSITIONS

_____A. Diagnostic Category of Clinical Sample (at least 90% representation): 1) schizophrenic; 2) nonpsychotic disturbances; 3) delinquency; 4) mixed diagnosis

_____

_____

_____

_____B. Composition of Family Unit Participating in Research: 1) mo-child dyad; 2) spousal dyad; 3) family triad consisting of parents, and focal child; 4) family tetrad consisting of parents, focal child and well-sibling; 5) other.

_____C. Family Unit on Which Measurement is Taken: 1) individual-mo; 2) individual-fa; 3) individual-child; 4) mo-child dyad; 5) fa-child dyad; 6) spousal dyad; 7) family triad; 8) family tetrad; 9) spousal comparison; 10) mo-child comparison; 11) fa-child comparison; 12) mo-family comparison; 13) fa-family comparison; 14) child-family comparison; 15) other _____

_____

_____

Can study be scored on developmental stage (defined in terms of the age of the focal child): _____Yes, if affirmative response to all three of the following questions. ____No, if any one response is negative. ____Nonapplicable when child not IP.

1. Is clinical sample homogenous with respect to the age of the focal child (at least 90% representation by one of the following age categories: 1) 2-4 yrs.; 2) 5-9 yrs.; 3) 10-13 yrs.; 4) 14-17 yrs.; 5) >17 yrs.)? Yes____ No____ Unknown____

2. Is nonclinical sample homogenous with respect to the age of the focal child? Yes____ No____ Unknown____

3. If yes to questions 1 and 2: Are clinical and nonclinical samples equivalent with respect to this study feature? Yes____ No____

_____D. Developmental Stage of Family (defined in terms of age of focal child): 1) 2-4 yrs.; 2) 5-9 yrs.; 3) 10-13 yrs.; 4) 14-17 yrs.; 5) > 17 yrs.; 6) unknown; 7) nonscorable; 8) nonapplicable

Can study be scored on sex of the focal child? ____Yes, if affirmative response to all three of the following questions. ____No, if any one response is negative. ____Nonapplicable when child not IP.

1. Is clinical sample homogenous with respect to the sex of the focal child (at least 90% representation by one sex): Yes____ No____ Unknown____

2. Is nonclinical sample homogenous with respect to the sex of the focal child? Yes____ No____ Unknown____

3. If yes to questions 1 and 2: Are clinical and nonclinical samples equivalent with respect to this study feature? Yes____ No____

_____E. Sex of Focal Child: 1) female; 2) male; 3) unknown; 4) nonscorable; 5) nonapplicable

_____

_____

_____

FEATURES RELATING TO METHODOLOGICAL RIGOUR

A. Demographic Comparability

Score of 1: no information or vague statement suggesting likelihood of sample comparability on index or statement of statistical support for nonequivalence.

Score of 2: statement that matching procedures were used, without any verbal or statistical information in support of the effectiveness of the procedure.

Score of 3: verbal statement indicating that statistical results support the successful construction of matched samples or statistical data that supports such construction or restricted use of only one value of the demographic variable or use of statistical procedures that control for unmatched variables (e.g., ANCOVA)

Assign to each of the following demographic variables one of the above scores:

1) socioeconomic status..... _____
(defined in terms of education or occupation of parent(s) or in terms of some composite score based on these 2 features).

2) family composition..... _____
(defined in terms of birth order of focal child, number of siblings, or size of family).

3) sex of focal child ..... _____

4) developmental stage of family ..... _____
(defined in terms of age of focal child or age of parent(s) or ages of all participating offspring).

Total Score _____

B. Sample Specificity

a) diagnostic specificity (see coding supplement for details regarding question 2, p. 4).

1) Is clinical sample restricted to only one of the three major diagnostic divisions: schizophrenic, nonpsychotic disturbance or delinquency (at least 90% representation)?.....
Yes (score 1)____ No____ Unknown____

2) Is clinical sample further restricted to a more precisely defined subset of the major diagnostic division (at least 90% representation)?
Yes (score 1)____ No____ Unknown____

Total Score _____

3) Did the clinical families undergo similar/identical diagnostic procedures (single subject source or independent assessment by members of research team)?.....
Yes (score 1)____ No____ Unknown____

4. Are clinical families homogenous with respect to the locus of the presenting problem (symptom bearer is either the child or one of the spouses or the marital relationship)?.....
Yes (score 1)____ No____ Unknown____

Total Score _____

_____

_____

_____

b) developmental specificity (defined in terms of age of focal child)

1) Which of the following age levels are represented in the clinical sample? When age range is the only information reported, score for all age levels contained within its boundaries. When information consists of lower limit plus mean or medium value, begin count at lower limit, then score until mean or median category and add 1 to final score.

2-4 years..... _____Score 1
5-9 years..... _____Score 1
10-13 years..... _____Score 1
14-17 years..... _____Score 1
>17 years..... _____Score 1

Total Score _____ Unknown _____

2) Which of the following age levels are represented in the nonclinical samples? When age range is the only information reported, score for all age levels contained within its boundaries. When information consists of lower limit plus mean or median value, begin count at lower limit, then score until mean or median category and add 11 to final score.

2-4 years..... _____Score 1
5-9 years..... _____Score 1
10-13 years..... _____Score 1
14-17 years..... _____Score 1
>17 years..... _____Score 1

Total Score _____ Unknown _____

When discrepancies exist between the developmental specificity of clinical and nonclinical samples, score in terms of the least specific sample. Total Score_____ Unknown_____ Nonapplicable when child not IP_____.

c) sex specificity (defined in terms of sex of focal child)

1) Does one sex predominate in clinical sample (at least 90% representation)?: Yes____ No____ Unknown____

2) Does one sex predominate in nonclinical sample (at least 90% representation)?: Yes____ No____ Unknown____

In cases in which clinical and nonclinical samples differ on sex specificity, score study in terms of nonspecific sample.

_____ sex specificity: 1) yes; 2) no; 3) unknown; 4) nonapplicable when child not IP.

C. Control for Observer Expectancies

1) Is all information regarding diagnostic and treatment status withheld from observer and coder?: Yes____ No____

2) Are settings equivalent for diagnostic groups?: Yes____ No____ Unknown____

_____ Control for Observer Expectancies: 1) Yes, if affirmative response to both of the preceding questions; 2) No, if negative response to either of the preceding questions; 3) Unknown, if answer to setting question is unknown where the information question is responded to affirmatively; 4) nonapplicable for noninferential measures.

D. Control for Subject Expectancies

4 demand indices: 1) method of subject solicitation
2) purported rationale of study
3) nature of payment for participation
4) experimental setting

1) How were clinical subjects solicited?

_____ in response to an advertisement or a request from an institution (e.g., school) that is not engaged in mental health work.

_____ in response to a request from a mental health worker with whom the subject has entered into therapy on a self-referred basis

_____ in response to a "request" from a mental health worker to whom the subject has been referred for treatment that research participation be incorporated into therapy

_____ unknown

_____ other _____

_____

How were nonclinical subjects solicited?

_____ in response to an advertisement or a request from an institution (e.g., school) that is not engaged in mental health work

_____ unknown

_____ other _____

_____

Were clinical and nonclinical subjects solicited in an equivalent manner? Yes (score 1)____ No____ Unknown____

2) What research rationale did clinical subjects receive?

_____

_____

What research rationale did nonclinical subjects receive?

_____

_____

Did clinical and nonclinical subjects receive equivalent rationale?
Yes (score 1)____ No____ Unknown____

3) How were clinical subjects paid for research participation?
monetary remuneration____, access to research findings____,
access to therapy____, no payment____, unknown____,
other_____

How were nonclinical subjects paid for research participation?
monetary remuneration____, access to research findings____,
no payment____, unknown____, other_____

Did clinical and nonclinical subjects receive equivalent payment?
Yes (score 1)____ No____ Unknown____

4) Are clinical and nonclinical subjects tested within the same experimental setting? Yes (score 1)____ No____ Unknown____

Total Score_____ % Unknown_____

E. Reliability of Measurement: Does study report a reliability statistic?
Yes____ No____ Nonapplicable (for noninferential measures)____

1. If yes, statistic refers ____specifically to outcome measure of immediate interest; ____to an entire set of outcome measures to which the measure of interest belongs.

2. If study reports a Pearson r, indicate its value _____

3. If study reports a probability value associated with r in the absence of the coefficient, translate into conservative estimate of r _____.

4. If study reports alternate reliability coefficient, indicate nature and value _____.

F. Control for Overall Frequency: 1) yes, 2) no, 3) nonapplicable (see coding supplement p. 4 for criteria).

FEATURES RELATING TO RESEARCH FOCUS

_____ G. Mode of Communication: 1) verbal; 2) verbal/tonal; 3) postural; 4) verbal/tonal/postural; 5) nonapplicable (see coding supplement pp. 4-5 for criteria); 6) nonscorable for composite measures that are not uniform with respect to this feature.

_____ H. Substantive domain (see coding supplement pp. 7-17 for allocation guidelines): 1) boundaries of authority; 2) domineering expression; 3) family cohesion; 4) affective expression; 5) communication clarity; 6) flexibility/rigidity; 7) family-as-a-task-oriented-unit; 8) quantity of communication; 9) communication continuity; 10) nonscorable.

Reason for use of "nonscorable" code _____

_____

Does researcher explicitly link dependent measure to a theoretical construct? Yes____ No____

If yes, what is the construct? _____

_____

Under what substantive domain does the construct belong? Specify number: _____

Is substantive domain identified by researcher consistent with that identified by coder? Yes____ No____

I. Subgroup Comparison: Does measure involve subgroup (including individual) of the family unit that participates in research? (see Family Unit on Which Measurement is Taken, p. II): Yes___ No___

_____ If yes, 1) measure involves comparisons between family subgroups; 2) measure is taken on a single subgroups independent of its relationship to other subgroups.

_____ J. Response Category Comparison: 1) Measure involves comparison of more than one response category (e.g., ratio of positive social-emotional communications to negative social-emotional communications); 2) measure involves a single or composite response category independent of its relationship to other categories.

Additional comments: _____

_____

_____

_____

Appendix C

SUPPLEMENT TO CODING FORM FOR FAMILY INTERACTION META-ANALYSIS

Appendix C

SUPPLEMENT TO CODING FORM FOR FAMILY INTERACTION META-ANALYSIS

Tasks Used to Generate Interactional Data

a) Task Structure - see page 1 of coding form.

High Structure. Tasks are coded as high structure when a variety of restrictions placed on family participants result in a type of interaction quite unlike that which occurs in the family's everyday life, both in terms of its form and content. Restrictions may be placed on the physical positioning or movement of family members by, for instance, prescribing seating arrangements. Or the number of channels of communication generally availalbe to human interactants may be limited. Visual communication, for instance, may be blocked by the appropriate placement of visual barriers or by having family members communicate from separate rooms by microphone and receiver. Alternatively, verbal communication may be prohibited, or restricted to only one aspect of speech such as questions or yes/no responses. Strict control over family interactions may also take the form of directives concerning the number of family members able to participate at any one point during the experimental proceedings or the order in which they participate.

Medium Structure. Tasks that are coded as medium structure are characterized by a specific set of instructions that direct the content of family exchanges and by a specific set of stimuli that provide a basis around which to focus exchanges. Such tasks do not, however, apply any restrictions on the form of family exchange.

Low Structure. Tasks are coded as low structure when they leave research subjects free to direct both the form and content of family interactions. If experimental stimuli are introduced (e.g., toys, juice and cookies, blackboard), they serve largely to "naturalize" the experimental context rather than to provide a basis around which to direct exchanges.

b) Emotional Tone - see page 1 of coding form.

Highly Emotional. Tasks are coded as highly emotional if they engage research families in interactions around issues that are profoundly important in family life: issues concerning parental control, children's compliance, spousal roles and the like. Such tasks can also be identified in terms of their emphasis on interpersonal and affective issues relating to family life as opposed to instrumental issues. Moreover, the relevance of these issues to the research families is explicitely acknowledged in task instructions.

Moderately Emotional. Tasks are coded as moderately emotional if, like highly emotional tasks, they engage research families in interactions around interpersonal and affective concerns. Moderately emotional tasks can be distinguished from highly emotional tasks insofar as their instructions allow respondents to distance themselves from the issues in hand. That is, the personal relevance of such issues to the research families is neither directly acknowledged in task instructions nor directly reflected in task stimuli. Examples of moderately emotional tasks are ones that use projective stimuli depicting family experiences, and tasks that require subjects to comment on norms and values relating to family life rather than on their actual experiences.

Emotionally Neutral. Tasks are coded as emotionally neutral if they focus on experiences and issues that are not generally thought of as central to family life (e.g., questionnaire items that request subjects' opinions on car colors, dinner menus and the like, games involving identification of visual designs).

Outcome Measures

a) Inference Level - see page 2 of coding form.

Noninferential. Outcome measures are classified as noninferential if they require the observer to simply conduct a frequency count of an objective event without making any decisions as to its nature (e.g., interruptions, statements received, talking time, number of correspondences between individual members' questionnaire responses and group responses, numbers of adjectives incorporated into a story).

Inferential. Outcome measures are classified as inferential if they require the observer not only to determine the occurrence of an event but also to evaluate its nature subjectively (e.g., agreements, yielding, empathy, degree of vagueness and confusion by joint family TAT).

b) Process/Product Distinction - see page 2 of coding form. Outcome measures are classified as "process" if they evaluate the nature of the interaction that occurs while the family is engaged in an experimental task. Such measures can be contrasted with those that evaluate the end product of a family interaction. This distinction may be important insofar as the two types of measures may yield very different sorts of information. Though it can be taken for granted that the process of an interaction has an effect on the product, the form of the effect may not be readily predictable. For example, a family interaction that is not characterized by frequent disagreements need not result in a group-product whose form directly mirrors those disagreements. The product of such an interaction may be one that can best be described as a "nonproduct" due to the failure on the part of family members to reach any agreement, or one that can be described as "dictatorial" insofar as it reflects the opinions and decisions of only one member of the family, or one that can be described as "democratic" in that it closely reflects the opinions and decisions of the majority of family members, etc. Note that while all interactional tasks yield process-related measures (which may or may not be translated into an outcome measure(s) by the primary researcher), some do not yield product-related information, in particular those that incorporate very open-ended task requirements (e.g., general verbal exercises that require the family to simply discuss a topic provided by the researcher).

When coding the product/process feature it may be helpful to verbalize the following question: Is it some aspect of the interaction or some aspect of the product that is being reflected in the measure?

Examples of process measures: agreements, disagreements, yielding, acknowledgement, interruptions, statements received, talking time, button pushes.

Examples of product measures: number of correspondences between individual members' questionnaire responses and group responses; number of adjectives incorporated into a story, degree of vagueness and confusion of family TAT.

Sample Specificity

a) Diagnostic Specificity - see page 4 of coding form re question 2.

Clinical samples are further restricted to one reasonably well-defined subgroup of a major diagnostic division. Schizophrenic samples may be subdivided according to any of a number of bases such as premorbid adjustment, symptom chronicity, severity of presenting symptoms and nosological subtype (e.g., paranoid, hebephrenic, etc.). Delinquent samples may be subdivided according to the nature of the legal infraction (e.g., aggressive vs. victimless act) or according to nosological subtype (e.g., socialized psychopathic, neurotic-disturbed, and socialized subcultural). Nonpsychotic disturbed samples may be subtyped into a variety of diagnostic groups, namely neurotics/true phobia, simple phobia, emotional-somatic disorder and substance abuse. Examples of specific diagnoses included under these headings are presented in the following chart.

Examples of Nonsychotic Distrubed Diagnoses Organized into Specified Subgroups*

Neurotic - True (Complex) Phobic

Neurotic
Complex Phobic
Poorly adjusted
Underachievers
Low Achievers
Students with Behavioral Problems
Hyperactive Children

Simple (Monosymptomatic) Phobics

Dog Phobia
Acrophobia

Emotional - Somatic Disordered

Asthmatic
Insomnia
Anorexia
Overweight
Enuresis

Habitue

Alchoholics
Drug Addicts

Depressives

*These subgroup headings and most of the diagnoses were selected from the diagnostic coding scheme used by Smith, Glass and Miller (1980, pp. 195-196) in their integration of the therapy outcome literature.

Allocation guidelines for Boundaries of Authority: Two categories of measures can be identified that fit under this substantive heading. The first consists of attempts to detect the presence of cross-generational alliances by means of measures that reflect the relative strengths of parent-parent and parent-child bonds (e.g., differential rates of communication between husband and wife and parent and child, differential rates of agreement and disagreement, differential rates of support, differential rates of who-follows-whom in speech sequence). Measures that compare diagnostic categories in terms of lines of alliances must base such comparisons on within-group differences between parent-child and husband-wife interactions.

The second category consists of those measures that attempt to assess the relative dominance of family members. This can be expressed in any of the following ways: 1) relative capacity to influence others or to be influenced by others (e.g., changes in opinion to match that of another family member, extent to which individual preferences are incorporated into group preferences); 2) relative capacity to direct the flow and content of family exchange (e.g., successful interruptions, defined in terms of entrance of third party into a two-party interaction that leads to a change in topic, being successfully interrupted, speaks first, speaks last); 3) relative use of controlling and being controlled forms of communication (e.g., ratings of explicit attempts to direct or limit the behavior of another person, behaviorally-focused topic change concerning deportment demands); 4) relative performance on a competitive game; and 5) relative amount of participation in family exchange, either of an active or passive nature (e.g., talking time, statement length, number of statements received).

Measures that examine relative dominance/passivity must base any comparisons that are made between diagnostic categories on within-group comparisons, the nature of which varies from measure to measure. There are first those measures that focus on husband-wife comparisons, often in an attempt to tap the theoretical construct of role-reversal. In addition, there are those measures that focus on parent-child comparisons, the concern in such cases being the appropriateness of the positions assumed by family members in the family power structure. The results of such studies are often presented in terms of dominance hierarchies which serve as the basis for cross-diagnostic comparisons. Finally, there are those measures that focus on individual-group comparisons by assessing the individual's communication relative to that of the family unit (i.e., through the calculation of the proportion of the family's use of a dominance-related measure attributed to the individual).

Allocation guidelines for Domineering Expression: Measures that assess the frequency of "controlling" and "being controlled" forms of communication are included under this heading (e.g., behaviorally focused topic change concerning deportment demands; dictatorial decisions, defined as instances in which one family member's decision is chosen to represent the family group in the face of an opposing majority decision; ratings of explicit attempts to direct or limit the behavior of another person; intrusiveness, defined as instances in which one person speaks when another has been explicitly invited to speak; proportion of person-oriented communications concerning the motives, feelings, wishes and experience of other rather than of self).

Allocation guidelines for Family Cohesion: The following categories of measures are included under this substantive heading: 1) measures that reflect the degree to which family members communicate with one another (e.g., time spent in and out of mutual button pushing in coalition games, total interaction time, number of initiated actions); 2) measures that reflect the extent to which values, ideas, and opinions are shared among family members, (e.g., spontaneous (e.g., spontaneous agreement scores on RDT, explicit agreements and disagreements); 3) measures that reflect the responsiveness of family members to one another, expressed either through the incorporation of individuals' choices into joint family choices (e.g., choice fulfillment on RDT) or through signs of attentiveness during communication (e.g., direct responses to questions, amplification of others' comments, visual contact during direct address, nonverbal digressions such as whistling and newspaper reading during family exchange); and 4) measures that reflect the degree to which family members explicitly declare their own opinions and feelings and allow and encourage other members to likewise express themselves (e.g., explicit exchange of opinions, requests for such declarations from others, ratio of person-oriented communication concerning the wishes, opinions and beliefs of other family members to total number of person-oriented communications).

Allocation guidelines for Affective Expression: Three categories of measures can be identified that fit under this substantive heading: 1) measures that reflect the amount or intensity or range of affective expression without regard to style or valence (e.g., ratings of strength of feelings or alternatively of lack of feeling, inter-personal expressiveness, range of evaluative communications, relative amount of expressive as opposed to instrumental exchange); 2) measures that reflect positive affective expression (e.g., ratings of support, reciprocal reinforcement, tension release, laughter, jokes); and 3) measures that reflect negative affective expression (e.g., ratings of conflict, ratings of depressive mood, sarcastic disagreements,

criticisms, reactions of shock to proposed solutions, ratings of inability or unwillingness to engage in supportive communication.

Allocation guidelines for Communication Clarity: Four categories of measures can be identified that fit under this substantive heading: 1) measures that assess the extent to which the communication of family members is unintelligible because of its bizarre or "schizophrenese-like" quality; 2) measures that assess the extent to which information is clearly and explicitly exchanged among family members (e.g., explicit agreements and disagreements, explicit declarations of ideas and opinions, identification of unclear or bizarre communication in others, requests for clarification); 3) measures that assess not so much the clarity of distinct bits of communication as the extent to which those bits are cohesively linked (e.g., shifts in themes, parallel messages, appropriate and inappropriate topic change); and 4) measures that assess the extent to which family members disqualify either their own communication (e.g., self-contradictory statements, lack of affective congruence between communication channels) or that of other family members (e.g., verbal tangentializations or evasions, change of subject that are labelled as answers, redundant questions implying doubt or disagreement without openly stating it).

Allocation guidelines for Flexibility/Rigidity: Measures that demand no inferences about content and that are to a large extent concerned with the distribution of family exchange are placed in this category (e.g., evenness of distribution of who-follows-whom, degree to which distribution of who-speaks-to-whom is balanced, predictability of speakers in sequence, correlation between amount of silence across experimental tasks, rates of balanced triadic interaction relative to dyadic interaction, degree to which rates of frequency of speech among members are balance, frequencies of speech disruptions).

Allocation guidelines for The Family-as-a-Task-Oriented-Unit: Four categories of measures fit under this substantive heading: 1) measures that index the energy expended during an experimental task or, phrased somewhat differently, measures that index family efficiency at meeting task requirements (e.g., decision time on joint RDT, turns to completion on laboratory games); 2) measures that index the quantity of a task product independent of quality considerations (e.g., number of adjectives incorporated into a story, word production on joint family TAT); 3) measures that index the quality of a task product (e.g., clarity and cohesiveness of family TAT, number of correct responses in referential communication task, extent to which family members are able to manipulate chosen outcome in coalition game) and 4) measures that index families' ability to work

within a task set to a sufficient degree to meet basic task requirements (e.g., opening gamit scale that measures extent to which family members accept task set, number of times family fails to agree on solution or come to a decision, number of five requested story components included in family TAT).

Allocation guidelines for morphological rather than theoretical similarity: Two categories of measures are constructed on the basis of similarity of form rather than on the basis of common theoretical underpinnings. The first can be referred to as "quantity of communication" and consists of the following sorts of measures: total words spoken, total number of initiated actions as defined by Bales IPA categories, total number of speeches, amount of time spent in audible speech and amount of time spent in silence. The second category can be referred to as "communication continuity" and consists of measures such as interruptions, simultaneous speech and pauses. It is worth noting that measures classified under morphologically-based headings are generally classified under other substantive headings as well. For instance, studies that compare family members within each diagnostic category in terms of their communication output are classified under "boundaries of authority" as well as under "quantity of communication".

Allocation guidelines for "nonscorables": The code nonscorable is appropriate in the following cases: 1) when the operational specification provided by the researcher is insufficiently detailed; 2) with "diffuse" measures that consist of a number of loosely connected elements, the composite of which can best be subsumed under a very general label such as "global dysfunction" rather than under one of the more specific substantive headings contained in the coding scheme; 3) with anomalous measures that cannot be readily shoehorned into any one of the specified substantive codes; the coding scheme for substantive categories was not designed to be exhaustive; and 4) when the allocation guidelines are insufficiently clear and/or detailed. Space is provided on the coding form for the coder to specify why the nonscorable classification is being used.

Guidelines for multiple classifications: There are two circumstances in which a measure can be classified under more than one substantive heading: 1) when the nature of the measure is such that it reflects more than one substantive domain, suggesting that the domains may conceptually overlap. For example, measures of explicit declarations of opinions, beliefs and wishes reflect both the clarity of family communication and the degree to which self-expression is tolerated/encouraged in the family. Consequently, such measures should be coded under both "communication clarity" and "family cohesion"; and 2) when a

composite measure is derived by combining a number of previously scored response categories, all of which do not fit under a common substantive heading. When, however, a composite measure seems to consist of diverse behavioral categories that, when combined, form a measure whose sum is conceptually distinct from any one of its parts, it sould be treated as a single-category measure and coded according to its overall conceptual thrust.

Appendix D

JUDGEMENT MATRIX FOR FEATURE EMOTIONAL TONE

Appendix D

Judgement Matrix for Feature Emotional Tone

	Count Row % Column % Total %	Second Coder Highly Emotional	Moderately Emotional	Neutral	Nonscorable	Row Total
First Coder	Highly Emotional	53 58.2 86.9 12.9	2 2.2 1.1 0.5	1 1.1 1.0 0.2	35 38.5 62.5 8.5	91 22.2
	Moderately Emotional	1 0.7 1.6 0.2	123 83.1 64.7 30.0	3 2.0 2.9 0.7	21 14.2 37.5 5.1	148 36.1
	Neutral	1 1.3 1.6 0.2	2 2.7 1.1 0.5	72 96.0 69.9 17.6	0 0 0 0	75 18.3
	Nonscorable	6 6.3 9.8 1.5	63 65.6 33.2 15.4	27 28.1 26.2 6.6	0 0 0 0	96 23.4
	Column Total	61 14.9	190 46.3	103 25.1	56 13.7	410 100

Appendix E

LIST OF MEASURE-SPECIFIC AND STUDY-WIDE FEATURES

Appendix E

List of Measure-Specific and Study-Wide Features

Measure-Specific	Study-Wide
Inference Level	Task Structure
Process/Product	Emotional Tone
Global Rating	Task Characteristics
Unit of Measurement	Setting
Sex of Focal Child	Naturalistic/Controlled
Reliability	Diagnostic Category
Control for Observer Expectancies	Family Composition
Substantive Domain	Developmental Stage
Mode of Communication	Demographic Comparability
Subgroup Comparison	Diagnostic Specificity
Response Category Comparison	Developmental Specificity
	Sex Specificity
	Subject Expectancies