

10-1-2004

Social Science Knowledge in Family Law Cases: Judicial Gate-keeping in The *Daubert* Era

Sarah H. Ramsey

Robert F. Kelly

Follow this and additional works at: <http://repository.law.miami.edu/umlr>

Recommended Citation

Sarah H. Ramsey and Robert F. Kelly, *Social Science Knowledge in Family Law Cases: Judicial Gate-keeping in The Daubert Era*, 59 U. Miami L. Rev. 1 (2004)

Available at: <http://repository.law.miami.edu/umlr/vol59/iss1/2>

This Article is brought to you for free and open access by Institutional Repository. It has been accepted for inclusion in University of Miami Law Review by an authorized administrator of Institutional Repository. For more information, please contact library@law.miami.edu.

University of Miami Law Review

VOLUME 59

OCTOBER 2004

NUMBER 1

Social Science Knowledge in Family Law Cases: Judicial Gate-Keeping in the *Daubert* Era*

SARAH H. RAMSEY** & ROBERT F. KELLY***

ABSTRACT

This article addresses the pressing issue of the appropriate use of social science research by courts in family law decision-making. The jurisprudence of family law is increasingly influenced by social science research. Frequently, social science-based instrumental and empirical claims, rather than moral ones, are used to justify family law decisions. In addition, debate about the proper use of social science by courts has been spurred on by the United States Supreme Court's ruling in *Daubert v. Merrell Dow Pharmaceuticals* that makes courts gate-keepers responsible for assessing the validity of expert scientific testimony.

The article recommends an expansion of the gate-keeping model and argues that judges should scrutinize not only expert testimony but other sources of social science as well. Implementing this new model requires understanding the complex interrelationship among courts'

* The authors are grateful for the helpful suggestions made by colleagues on earlier drafts of this paper, for the research assistance of Heidi A. White, and for the editorial assistance of Christy Ramsdell. The authors also thank Judge R. James Williams, Supreme Court of Nova Scotia, Family Division, for inviting the authors to participate in the *Nova Scotia All Courts Education Seminar* where earlier versions of sections of this paper were presented. The authors appreciate the research support provided by the Syracuse University College of Law and Le Moyne College.

** Professor of Law and Laura J. and L. Douglas Meridith Professor for Teaching Excellence, Syracuse University College of Law; B.S., Duke University, 1965; J.D., University of North Carolina at Chapel Hill, 1973; LL.M., University of Michigan, 1982.

*** Professor of Sociology, Le Moyne College; B.A., St. Joseph's University, 1972; M.A. and Ph.D., Rutgers University, 1975 and 1978.

modes of access, evaluation, and use of social science. Hence, one goal of this article is to examine how courts use and access social science in family law decisions. Another goal is to provide research evaluation standards to assist users of that research. To accomplish these goals, the article develops definitions and a typology of uses of social science by courts and explores how courts access social science research. It also proposes standards for judges and others in assessing and using social science research and provides a set of key questions to be used in such assessments. The article concludes that decision-making in family law cases can and should benefit from social science research, but that judges must take into account the need to critically evaluate the contribution that such research can make.

TABLE OF CONTENTS

I. INTRODUCTION	3
II. DEFINITIONS AND A TYPOLOGY OF USES OF SOCIAL SCIENCE	5
A. <i>Definitions</i>	5
B. <i>Typology of Uses</i>	7
1. THE GENERAL-RULE LEVEL	7
2. THE INTERMEDIATE LEVEL: THE CONTEXT FOR THE CASE	8
3. THE CASE-SPECIFIC LEVEL	10
III. SOURCES OF SOCIAL SCIENCE	12
A. <i>Expert Testimony</i>	12
1. RELEVANCE AND ASSISTANCE	13
2. QUALIFYING AN EXPERT	13
3. RELIABILITY	17
4. TESTIMONY ON THE ULTIMATE QUESTION	22
5. STANDARD OF REVIEW	23
B. <i>Briefs</i>	23
1. LEVEL OF SPECIFICITY AND CHOICE OF TRIAL OR APPELLATE LEVEL	24
2. BIAS	25
C. <i>Judicial Notice</i>	28
IV. GUIDELINES FOR ASSESSING THE VALIDITY OF SOCIAL SCIENCE RESEARCH	34
A. <i>The Logic of Scientific Inference and Progress: Empirical Testing</i>	38
B. <i>Methodological Standards of Design and Analysis</i>	42
1. MEASUREMENT VALIDITY	42
a. The Need for Empirically Validated Measurement Systems	43
i. Direct Research Claims	46
ii. Science-Based Practice Claims	46
b. The Need for Multiple and Independent Measures of Variables	48
i. Direct Research Claims	49
ii. Science-Based Practice Claims	50
2. INTERNAL VALIDITY AND THE PROBLEM OF CAUSAL INFERENCE	52
a. The Randomized Experiment	53
b. Alternatives to the Randomized Experiment: Quasi-Experimentation	57
3. EXTERNAL VALIDITY	61
a. Sampling and Generalizing	61
b. The "Fit" Between Research and Application	65
4. ANALYTIC VALIDITY	67
a. Statistical Significance	68

b. Specification of Models and Relationships 69

C. *The Social Practice of Science: Community Socialization, Review, and Communication* 70

 1. COMMUNITY SOCIALIZATION 70

 2. COMMUNITY REVIEW 72

 3. COMMUNITY COMMUNICATION 76

V. CONCLUSION 79

I. INTRODUCTION

The jurisprudence of family law is increasingly influenced by social science research. This trend is evidenced in court decisions,¹ interdisciplinary conferences,² multidisciplinary professional organizations,³ and in major law reform projects such as the American Law Institute’s thirteen-year development of the *Principles of the Law of Family Dissolution*.⁴ Frequently, social science-based instrumental and empirical claims, rather than moral ones, are used to justify family law proposals and decisions.⁵

In addition, debate about the proper use of social science by courts in general has been spurred on by the United States Supreme Court’s ruling in *Daubert v. Merrell Dow Pharmaceuticals* that makes courts responsible for assessing the validity of expert scientific testimony.⁶ Under *Daubert*, courts must ensure “that an expert’s testimony both rests on a reliable foundation and is relevant to the task at hand,” consequently placing a greater burden on courts when experts are used.⁷

1. *E.g.*, Baures v. Lewis, 770 A.2d 214 (N.J. 2001).

2. *E.g.*, Am. Bar Ass’n Family Law Section & The Johnson Found., *High-Conflict Custody Cases: Reforming the System for Children – Conference Report and Action Plan*, 34 FAM. L.Q. 589 (2001).

3. *E.g.*, AFCC, Association of Family and Conciliation Courts, (“an international and interdisciplinary association of family, court, and community professionals dedicated to the constructive resolution of family disputes”) (emphasis in original), at <http://afccnet.org/docs/about.html> (last visited Dec. 14, 2004).

4. AM. LAW INST., PRINCIPLES OF THE LAW OF FAMILY DISSOLUTION: ANALYSIS AND RECOMMENDATIONS (2002).

5. *See, e.g.*, Lee E. Teitelbaum, *Divorce, Custody, Gender, and the Limits of Law: On Dividing the Child*, 92 MICH. L. REV. 1808, 1809 (1994); Carl E. Schneider, *On the Duties and Rights of Parents*, 81 VA. L. REV. 2477 (1995).

6. *Daubert v. Merrell Dow Pharm., Inc.*, 509 U.S. 579 (1993). For additional discussion of the *Daubert* requirements, see *infra* Part IV. *Daubert* interpreted the Federal Rules of Evidence, but because most state evidence rules track the federal rules, the Supreme Court’s ruling has been very influential at both the state and federal levels. *See, e.g.*, Michael J. Saks, *The Aftermath of Daubert: An Evolving Jurisprudence of Expert Evidence*, 40 JURIMETRICS J. 229, 230 (2000).

7. *Daubert*, 509 U.S. at 597 (1993). The Court’s ruling was based on its interpretation of the Federal Rules of Evidence, although the Rules were subsequently amended to more closely track the Court’s interpretation. On remand, Judge Kozinski of the Ninth Circuit Court of Appeals described this new burden as follows:

Our responsibility, then, unless we badly misread the Supreme Court’s opinion, is to resolve disputes among respected, well-credentialed scientists about matters

The connection between family law and social science research raises a number of issues for judges deciding family cases, attorneys preparing cases, and legal scholars working on law reform. In today's world, general and unsupported "facts" about families are hard to justify. A judge would find it difficult, for example, to justify an award of custody to a mother through the reasoning that "[m]other love is a dominant trait in even the weakest of women, and as a general thing surpasses the paternal affection for the common offspring, and, moreover, a child needs a mother's care even more than a father's."⁸ However, courts also receive criticism when they uncritically adopt a theory promulgated by well-known practitioners in a field, such as the psychological parent theory proposed in *Beyond the Best Interests of the Child*.⁹

When judges, attorneys, and policy-makers use social science to develop their decisions, strategies, and recommendations, they need to have the capacity to assess social science research and the ability to use research findings appropriately. Of course, social science does not provide the definitive answer to a policy question or a proper case outcome.¹⁰ A policy choice is a normative choice, formed after weighing competing values and goals. A judge's determination is formed after consideration of the facts of a case and the applicable law; increasingly, social science information is playing a role in this evaluation, albeit an inherently limited role. Understanding this role and assessing social science are important components of modern family law jurisprudence.

This article focuses on courts and recommends that judges, when considering social science in family law matters, broadly employ a gate-keeping model which scrutinizes not only expert testimony but other sources of social science as well. Implementing this model requires an understanding of the complex interrelationship among modes of access, evaluation, and use. Hence, one goal of this article is to examine how courts use and access social science in family law decisions. Another goal is to provide research evaluation standards to assist users of that research.

To accomplish these goals, Part II provides definitions and a typol-

squarely within their expertise, in areas where there is no scientific consensus as to what is and what is not "good science," and occasionally to reject such expert testimony because it was not "derived by the scientific method."

Daubert v. Merrell Dow Pharm., Inc., 43 F.3d 1311, 1316 (9th Cir. 1995).

8. *Freeland v. Freeland*, 159 P. 698, 699 (Wash. 1916).

9. Peggy C. Davis, *'There is a Book Out. . .': An Analysis of Judicial Absorption of Legislative Facts*, 100 HARV. L. REV. 1539 (1987) (criticizing courts' use of JOSEPH GOLDSTEIN, ANNA FREUD, & ALBERT J. SOLNIT, *BEYOND THE BEST INTERESTS OF THE CHILD* (1973)).

10. See *infra* Part IV.B.3.b; see also Sarah H. Ramsey & Robert F. Kelly, *Using Social Science Research in Family Law Analysis and Formation: Problems and Prospects*, 3 S. CAL. INTERDISC. L.J. 631 (1994).

ogy of uses of social science by courts. Part III explores how courts access social science research. Part IV proposes guidelines for judges and others in assessing and using social science research. The article concludes that decision-making in family law cases can benefit from social science research, but judges must take into account the need to critically evaluate the contribution that such research can make.

II. DEFINITIONS AND A TYPOLOGY OF USES OF SOCIAL SCIENCE

A. Definitions

Post-*Daubert* debates have addressed the question: Is social science really “science”?¹¹ Responding to this question requires a common understanding of the meaning of certain fundamental terms, specifically, “science,” “scientific method,” and “social science.” “Science” refers to a method of producing knowledge in which general statements—hypotheses and theories—are tested empirically under controlled conditions with the goal of producing comprehensive explanations about the operation of some system.¹² In the context of this understanding of science, the “scientific method” refers to the rules or standards and community practices by which science proceeds.¹³

Several points should be made with respect to this definition of science. There is a decided emphasis on methodology in the definition rather than on the system or phenomenon studied or the specific substantive research findings. *Daubert* is consistent with this emphasis.¹⁴ Fur-

11. See, e.g., David L. Faigman, *The Evidentiary Status of Social Science Under Daubert: Is it “Scientific,” “Technical,” or “Other” Knowledge?*, 1 PSYCHOL. PUB. POL’Y & L. 960, 961 (1995); Daniel W. Shuman, *What Should We Permit Mental Health Professionals to Say About “the Best Interests of the Child”?: An Essay on Common Sense, Daubert, and the Rules of Evidence*, 31 FAM. L.Q. 551 (1997). For discussion of social science expert testimony based on clinical assessments, see *infra* note 53 and accompanying text, and for discussion of science-based practice claims see *infra* Part IV.B.

12. See, e.g., SHELDON J. LACHMAN, *THE FOUNDATIONS OF SCIENCE* 15-25 (1956); Brief of Amici Curiae American Association for the Advancement of Science and the National Academy of Sciences at 7-13, *Daubert v. Merrell Dow Pharm., Inc.*, 509 U.S. 579 (1993) (No. 92-102). *Daubert* seems to embrace the definition proposed by the brief: “‘Science is not an encyclopedic body of knowledge about the universe. Instead, it represents a *process* for proposing and refining theoretical explanations about the world that are subject to further testing and refinement.’ (emphasis in original).” *Daubert*, 509 U.S. at 590. See also KENNETH R. FOSTER & PETER W. HUBER, *JUDGING SCIENCE: SCIENTIFIC KNOWLEDGE AND THE FEDERAL COURTS* 1-22 (1997).

“Science” has also been defined as a body of knowledge. For example, the authors of *Science in the Law* note that the “classical definition of a science is ‘an orderly body of knowledge with principles that are clearly enunciated.’” DAVID L. FAIGMAN ET AL., *SCIENCE IN THE LAW: FORENSIC SCIENCE ISSUES* § 1-5.1 (2002).

13. See, e.g., LACHMAN, *supra* note 12, at 51-71; DAVID L. FAIGMAN ET AL., *SCIENCE IN THE LAW: STANDARDS, STATISTICS AND RESEARCH ISSUES* § 4-1.2 (2002).

14. See *Daubert*, 509 U.S. at 590 (“a process for proposing and refining theoretical explanations about the world that are subject to further testing and refinement”) (quoting Brief of

ther, the definition is cross-disciplinary in that it applies to any discipline that employs the scientific method to produce knowledge, regardless of the system that the discipline studies.¹⁵ In short, "science" should mean the same thing to a physicist, an embryologist, a psychologist, or a demographer. What matters is the method, not the system under study. In Part IV, we describe the logic, methodological standards, and community practices that constitute the scientific method and discuss their relevance to social science knowledge when it is presented to the courts.

In the context of the preceding discussion, "social science" refers to those disciplines that employ the scientific method to study human social systems.¹⁶ Disciplines generally included under the umbrella of the term "social science" are anthropology, sociology, demography, criminology, economics, political science, and most sub-fields of psychology, such as developmental child psychology and social psychology. Illustrations of social science disciplines and the respective social systems they study would include: economics which studies production and distribution systems of human behavior; anthropology which studies cultural systems such as language and kinship; political science which studies governmental systems; and sociology which studies social systems such as families, schools, and neighborhoods.

As long as the research being considered has employed the scientific method, our answer to the question, "Is social science really 'science'?" is affirmative.¹⁷ However, we think a more useful question

Amici Curiae American Association for the Advancement of Science and the National Academy of Sciences at 7-8, *Daubert v. Merrell Dow Pharm., Inc.*, 509 U.S. 579 (1993) (No. 92-102)). For different views of what constitutes science, see e.g., LARRY LAUDAN, *PROGRESS AND ITS PROBLEMS* (1977); PAUL FREYERABEND, *AGAINST METHOD* (3d ed. 1993). See also sources cited *infra* note 162.

15. For an analysis of science and social science that similarly focuses on methodology, see RICHARD S. RUDNER, *PHILOSOPHY OF SOCIAL SCIENCE* 6-7 (1966).

16. For discussions of the nature of the social sciences akin to this definition, see BERNARD BARBER, *SCIENCE AND THE SOCIAL ORDER* 312 (First Collier Books 1962) (1952); Max Weber, "Objectivity" in *Social Science*, in *THE METHODOLOGY OF THE SOCIAL SCIENCES* 72 (Edward A. Shils & Henry A. Finch eds. & trans., 1949). For a sociological analysis that utilizes this definition, see HARRY C. BREDEMEIER & RICHARD M. STEPHENSON, *THE ANALYSIS OF SOCIAL SYSTEMS* (1962).

Within many social science disciplines there is a humanities-oriented disciplinary perspective that does not emphasize the scientific method, but rather focuses on interpretative/hermeneutic methods more characteristic of the humanities. One-group case studies, ethnographic studies, and phenomenological studies would typically fall within this type of social science. See, e.g., JOHN W. CRESWELL, *RESEARCH DESIGN: QUALITATIVE & QUANTITATIVE APPROACHES* (1994); J. MASON, *QUALITATIVE RESEARCH* (1996). Because research produced within this perspective does not claim scientific status in the terms used in *Daubert*, it is beyond the scope of our analysis. For additional discussions of this approach to social science, see ALEXANDER ROSENBERG, *PHILOSOPHY OF SOCIAL SCIENCE* 21 (1988).

17. For discussion of the scientific method, see *infra* Part IV.A.

from the point of view of a court is: When and how should courts employ the tests of scientific validity to testimony or research based in the social sciences? To answer this question, we need to first consider how courts access and use social science.

B. *Typology of Uses*

Courts use social science in three conceptually distinct, yet interrelated ways, depending on the level of specificity required in a particular case.¹⁸ This tripartite typology is useful in analyzing courts' choices of evidentiary rules, sources of social science, and assessment tools for reviewing social science research.

1. THE GENERAL-RULE LEVEL

First, at the most general level, appellate or trial courts may use social science evidence to develop or clarify a rule of law. Recent court decisions that have relaxed the restrictions on post-divorce custodial parent relocation provide examples of this type of use. The California Supreme Court, for example, appeared to rely heavily on a brief filed by psychologist Judith Wallerstein that presented social science data favoring permissive relocation when the court liberalized its relocation requirements.¹⁹ The New Jersey Supreme Court was more explicit in its reliance on social science research through its acknowledgement that social science research was a significant factor in its decision to liberalize the relocation standard. The court stated that “[m]ost importantly, social science research links a positive outcome for children of divorce with the welfare of the primary custodian and the stability and happiness within that newly formed post-divorce household,” and cited a number of social science studies, many of which were taken from the Wallerstein brief.²⁰ In changing its relocation standard, the Tennessee

18. Professors John Monahan and Laurens Walker have proposed a three-part classification for the courts' use of social science in relation to judicial notice: social authority, social frameworks and social facts. See John Monahan & Laurens Walker, *Social Authority: Obtaining, Evaluating, and Establishing Social Science in Law*, 134 U. PA. L. REV. 477 (1986) [hereinafter *Social Authority*]; John Monahan & Laurens Walker, *Social Frameworks: A New Use of Social Science in Law*, 73 VA. L. REV. 559 (1987) [hereinafter *Social Frameworks*]; John Monahan & Laurens Walker, *Social Facts: Scientific Methodology as Legal Precedent*, 76 CAL. L. REV. 877 (1988) [hereinafter *Social Facts*]. Although the uses we identify have some similarities to those identified by Professors Monahan and Walker in their analysis of judicial notice and “social frameworks,” the uses are sufficiently dissimilar that we have not used their terms to avoid confusion.

19. *In re Marriage of Burgess*, 913 P.2d 473, 483 (Cal. 1996). Unfortunately, the Wallerstein brief has been criticized extensively for a biased presentation of social science research. See *infra* text accompanying notes 112-21 for additional discussion of this brief and subsequent California relocation litigation.

20. *Baures v. Lewis*, 770 A.2d 214, 222 (N.J. 2001).

Supreme Court reviewed decisions from other jurisdictions that had adopted more liberal relocation standards and noted that the decisions “reflect the collective wisdom of both the courts and child psychologists that children, especially those subjected to the trauma of divorce, need stability and continuity in relationships most of all.”²¹

Class actions present another example of courts using social science research to develop a broad rule. *Nicholson v. Williams*, for example, involved a class action by mothers whose children had been removed by the New York City Administration for Children’s Services (which includes child protective services) because the mothers were victims of domestic violence.²² In addition to hearing testimony about the situations of the named plaintiffs, the trial court heard extensive expert testimony on the effects of domestic violence on children, especially the effect of witnessing such violence and the connection between domestic violence and abuse directed against the children.²³ The court also heard expert testimony regarding the effects on children when they are removed from their homes.²⁴ At the close of trial, the district court granted a preliminary injunction to ensure that “battered mothers who are fit to retain custody of their children do not face prosecution or removal of their children solely because the mothers are battered, and . . . [that] the child’s right to live with such a mother is protected.”²⁵

2. THE INTERMEDIATE LEVEL: THE CONTEXT FOR THE CASE

Second, at an intermediate level, a court may use social science research to present a context that is relevant to the parties’ circumstances, but that does not include facts unique to the particular case. In these circumstances, the court is not using social science to develop or explicate a general rule of law, but rather solely to decide the case before it. In *In re A.W.*, for example, a twelve-year-old girl asked the court to substitute counsel of her choice for her state-appointed attorney in a child protective case.²⁶ A child psychologist testified about the capacity of twelve-year-olds to make decisions about a matter as complex as choosing their own attorney.²⁷ However, as the psychologist had not

21. *Aaby v. Strange*, 924 S.W.2d 623 (Tenn. 1996) (citing *Taylor v. Taylor*, 849 S.W.2d 319, 328 (Tenn. 1993)).

22. *Nicholson v. Williams*, 203 F. Supp. 2d 153 (E.D.N.Y. 2002) (certifying a number of state law questions to the New York Court of Appeals while on appeal to the United States Court of Appeals for the Second Circuit in the case of *Nicholson v. Scopetta*, 344 F.3d 154 (2d Cir. 2003)).

23. *Id.* at 197-98.

24. *Id.* at 198-99.

25. *Id.* at 257.

26. *In re A.W.*, 618 N.E.2d 729 (Ill. App. Ct. 1993).

27. *Id.* at 732.

evaluated the girl prior to his testimony, he was only able to provide a general opinion rather than one based upon the capacity of the child in this case.

Syndrome evidence falls into this intermediate category when an expert testifies about a syndrome in general without having clinically evaluated anyone directly involved in the case before the court.²⁸ Indeed, even when an expert has evaluated someone involved in the case, a jurisdiction's evidentiary rules may require that the expert testify only in general terms without relating the general syndrome characteristics to a party in the case; this situation especially arises in criminal cases.²⁹ In *Commonwealth v. Trowbridge*, for example, the appellate court held that the testimony of a pediatric gynecologist, who was an expert witness on sexually abused children, should have been excluded. The expert "not only testified that the child's behavior was consistent with that of sexually abused children, she also testified that the symptoms and physical condition of the child were consistent with the type of non-violent sexual abuse that the child alleged in this case."³⁰ According to the court, expert testimony on general behavioral characteristics of sexually abused children was permissible, but an expert could not compare the child to those general characteristics because "such testimony impermissibly intrudes on the jury's province to assess the credibility of the witness."³¹

Another example of this type of context testimony is the use of an expert's opinion on the suggestibility of child witnesses. In *U.S. v. Rouse*, the court found that the exclusion of a psychologist's expert testimony on the credibility of the child witnesses was proper.³² The trial court had allowed the expert's testimony about his own research on suggestibility and learned or "implanted" memory as it related to improper interviewing techniques used by investigators that taint the testimony of child witnesses. The court noted the "troublesome line" between the application of the expert's general opinions to the case at hand, and the expert's impermissible opinion on the child witnesses' credibility.³³

28. *Pratt v. Wood*, 620 N.Y.S.2d 551 (N.Y. App. Div. 1994) (finding expert testimony on battered wife syndrome admissible evidence); *People v. Taylor*, 552 N.Y.S.2d 883 (N.Y. 1990) (holding that under certain circumstances and subject to certain limitations, evidence of rape trauma syndrome is both relevant and admissible evidence).

29. *Commonwealth v. Trowbridge*, 647 N.E.2d 413 (Mass. 1995); cf. *In re Nicole V.*, 518 N.E.2d 914 (N.Y. 1987) (allowing an expert to testify in child protective case that the child exhibited characteristics of the sexually abused child syndrome for the purpose of corroborating the child's out-of-court statements).

30. *Trowbridge*, 647 N.E.2d at 420.

31. *Id.* (citation omitted).

32. *United States v. Rouse*, 111 F.3d 561 (8th Cir. 1997).

33. *Id.* at 571.

3. THE CASE-SPECIFIC LEVEL

At the most specific level, courts may use an expert's research-based testimony that relates to facts unique to the case before the court. Testimony on a custody evaluation is an example of research-based, specific testimony about individuals involved in the case. The evaluator, for example, would have assessed the family in accordance with a protocol that is based in social science research, often using standardized tests as part of the assessment.³⁴ The expert's testimony, however, might consist almost entirely of a report on the clinical assessment with little evidence being presented on the research supporting the expert's approach.

Syndrome evidence is another example of case-specific testimony, when a court allows testimony about whether a party to a case is exhibiting characteristics of a syndrome and what that implies for the disposition of the case. The testimony on the syndrome itself would be based in research and the characteristics of the individual would be based on a clinical assessment. For instance, in *Soutiere v. Soutiere*, an expert was allowed to testify that the plaintiff wife in a divorce action suffered from "a version of post-traumatic stress disorder (PTSD), commonly known in this context as battered-woman syndrome."³⁵ The trial court had found "the expert's opinions helpful 'to understand the evidence' bearing on property distribution and maintenance."³⁶

This type of syndrome testimony has been allowed in criminal cases as well. In *State v. Nemeth*, for example, the court allowed testimony on the battered child syndrome generally and also allowed testimony that Brian, the sixteen-year-old defendant who had killed his parent, was an abused child.³⁷ The expert psychologist who had examined Brian was also an expert on the battered child syndrome. The court held that the evidence was relevant to elements of Brian's defense, including the determination of whether he "had an honest belief that he was in imminent danger, a necessary element in the affirmative defense of self-defense."³⁸ Further, the court stated:

General information on battered child syndrome would also tend to show that Brian's behavior was consistent with that of an abused child and would lend support to his testimony that he had been

34. See, e.g., *Azia v. DiLascia*, 780 A.2d 992, 996 (Conn. App. Ct. 2001) (using psychologist-conducted clinical interviews, observation, and a battery of tests to evaluate both the parents and the child). See generally JONATHAN W. GOULD, CONDUCTING SCIENTIFICALLY CRAFTED CHILD CUSTODY EVALUATIONS (1998).

35. *Soutiere v. Soutiere*, 657 A.2d 206, 208 (Vt. 1995).

36. *Id.* at 209.

37. *State v. Nemeth*, 694 N.E.2d 1332 (Ohio 1998).

38. *Id.* at 1336.

abused both generally and just prior to the killing. We have held that an expert may not offer an opinion as to the truth of a child's statement. However, an expert may provide testimony that supports "the truth of the facts testified to by the child, or which assists the fact finder in assessing the child's veracity." Expert testimony on battered child syndrome would, in this case, tend to enhance the probability that Brian's account of the facts leading up to the killing was truthful and would lend credibility to his assertion that he was in a state of rage and dissociation at the time of the killing. A diagnosis of battered child syndrome and an explanation of its effects would therefore be relevant in determining whether the case warranted a jury charge on voluntary manslaughter.³⁹

At this case-specific level, particularly where the testimony is focused entirely on the individuals in the case, as in a custody evaluation, questions might be raised about whether the testimony should be evaluated as "science."⁴⁰ Perhaps, following Federal Rules of Evidence, it should be classified as "technical or other specialized knowledge," rather than as "scientific" knowledge.⁴¹ Here, the salient issue is whether the presenters of the evidence are claiming, explicitly or implicitly, that their knowledge has scientific validity. Are they presenting themselves as experts with a gloss of scientific knowledge? Are they employing evaluative practices that purportedly are based in science?⁴² If they are, then their testimony should be scrutinized for scientific validity. Generally, when persons trained in social science are presenting evidence, they are making this claim, at least implicitly, and therefore questions of scientific validity should be addressed. Even though their testimony may be directed at their evaluation of an individual, its context is based in a scientific discipline. We propose that social sci-

39. *Id.* (citations omitted). Cases involving battered women are another example of testimony including both the individual's diagnosis and general information about the syndrome. See, e.g., *State v. Koss*, 551 N.E.2d 970, 975 (Ohio 1990) ("Where the evidence establishes that a woman is a battered woman, and when an expert is qualified to testify about the battered woman syndrome, expert testimony concerning the syndrome may be admitted to assist the trier of fact in determining whether the defendant acted in self-defense.").

40. See, e.g., Daniel A. Krauss & Bruce D. Sales, *The Problem of "Helpfulness" in Applying Daubert to Expert Testimony: Child Custody Determinations in Family Law as an Exemplar*, 5 PSYCHOL. PUB. POL'Y & L. 78, 87 (1999) (suggesting that an expert's opinion based on a clinical assessment has "no science").

41. FED. R. EVID. 702. It is important to note that even if the testimony is classified as "technical or other specialized knowledge," rather than as "scientific" knowledge, a high level of scrutiny should be employed. See *Kumho Tire Co. v. Carmichael*, 526 U.S. 137, 152 (1999) ("[A] trial court should consider the specific factors identified in *Daubert* where they are reasonable measures of the reliability of expert testimony.").

42. In Part IV, we describe such testimony as an illustration of a *science-based practice claim*.

ence-based testimony at all these levels and from all sources, should be assessed using the tools presented in Part IV of this article.

III. SOURCES OF SOCIAL SCIENCE

Courts access social science through three mechanisms: expert testimony, briefs, and judicial notice. Whatever the source of social science or the level of its use, we recommend that courts employ an expansive gate-keeping function and examine the scientific validity of the social science as well as its relevance to the issues in the case. This expansion of *Daubert* is needed if courts are to use social science successfully, particularly because the evidentiary distinctions among sources are not often clear. The following sections consider the sources of social science in relation to the level of its use as outlined in the prior section. Table 1 provides an overview of these relationships.

TABLE 1: SOURCES OF SOCIAL SCIENCE AND SPECIFICITY OF SOCIAL SCIENCE USES IN FAMILY AND CHILD JUDICIAL DECISION-MAKING

1.	<u>Expert Testimony</u>
	a. General Rule Level
	b. Intermediate Level
	c. Case-Specific Level
2.	<u>Briefs</u>
	a. General Rule Level
	b. Intermediate Level
	c. Case-Specific Level
3.	<u>Judicial Notice</u>
	a. General Rule Level
	b. Intermediate Level
	c. Case-Specific Level

A. *Expert Testimony*

Experts rely on social science research to provide information related to family law cases in all three levels of specificity described in Part II. In order for the expert to be permitted to testify, the trial court must decide that the testimony is relevant and will assist the trier of fact, that the expert is qualified, and that the testimony is reliable or grounded in validated scientific procedures.⁴³ This section considers these factors

43. FED. R. EVID. 702, Testimony by Experts.

If scientific, technical, or other specialized knowledge will assist the trier of fact to

and also addresses two additional issues, namely whether an expert should testify on the ultimate issue and what standard of review should be applied to an admissibility determination.

1. RELEVANCE AND ASSISTANCE

According to the *Daubert* interpretation of the Federal Rules of Evidence, the *Rule 702* requirement that the evidence or testimony “assist the trier of fact to understand the evidence or to determine a fact in issue” is related “primarily to relevance.”⁴⁴ An additional aspect of *Rule 702* and of relevancy is the notion of “fit.”⁴⁵ According to the Court, “[f]it” is not always obvious, and scientific validity for one purpose is not necessarily scientific validity for other, unrelated purposes. . . . *Rule 702*’s ‘helpfulness’ standard requires a valid scientific connection to the pertinent inquiry as a precondition to admissibility.”⁴⁶ In other words, a qualified expert who has done high quality research is not allowed to testify on that basis alone; there must be a connection between what the expert knows and what the court needs to determine.⁴⁷

2. QUALIFYING AN EXPERT

The trial court determines whether an individual qualifies as an

understand the evidence or to determine a fact in issue, a witness qualified as an expert by knowledge, skill, experience, training, or education, may testify thereto in the form of an opinion or otherwise, if (1) the testimony is based upon sufficient facts or data, (2) the testimony is the product of reliable principles and methods, and (3) the witness has applied the principles and methods reliably to the facts of the case.

The Rule was amended in 2000 in response to *Daubert* and codifies its general principles. FED. R. EVID. 702 advisory committee’s note. As noted previously, state evidence rules track the federal rules, see *supra* note 6.

Note that the evidentiary term “reliable” has a different meaning than the social science term “reliability.” See *infra* text accompanying note 171.

44. *Daubert v. Merrell Dow Pharm., Inc.* 509 U.S. 579, 591 (1993) (“Expert testimony which does not relate to any issue in the case is not relevant and, ergo, non-helpful.” (quoting 3 JACK B. WEINSTEIN & MARGARET A. BERGER, WEINSTEIN’S FEDERAL EVIDENCE § 702.02 (Joseph M. McLaughlin ed., Matthew Bender 2d ed. 1993))).

45. *Id.* at 591 (“An additional consideration under *Rule 702*—and another aspect of relevancy—is whether expert testimony proffered in the case is sufficiently tied to the facts of the case that it will aid the jury in resolving a factual dispute.” (quoting *United States v. Downing*, 753 F.2d 1224, 1242 (3d Cir. 1985))).

46. *Id.* at 591-92. It has been noted, however, that “[a]rguably . . . the notions of ‘fit’ and ‘helpfulness’ already exist in *Rule 402*’s relevancy requirement,” but that *Rule 702* goes beyond *Rule 402* to also require that the expert be qualified and that the content of the expert’s testimony be scientifically valid. FAIGMAN ET AL., *supra* note 13, at 19. *Rule 401* defines relevant evidence as “evidence having any tendency to make the existence of any fact that is of consequence to the determination of the action more probable or less probable than it would be without the evidence.” FED. R. EVID. 401.

47. For further discussion of the problem of “fit,” see *infra* Part IV.B.3.b.

expert by considering factors such as the individual's "knowledge, skill, experience, training, or education."⁴⁸ Because these factors are interrelated and should be assessed in relation to the proffered testimony, there is not an absolute delineation between qualified and unqualified testimony. A social science expert testifying about a law reform issue, for example, would need training in research methodology, experience in conducting research in the relevant substantive area, and knowledge of the field, but would not need clinical experience. This type of law reform expertise was used in *Baehr v. Miike*, which involved a challenge by gay couples to Hawaii's ban on gay marriages.⁴⁹ Both sides used experts such as psychologists and sociologists to testify on aspects of marriage and parenting.⁵⁰ For this testimony, the experts did not need clinical experience, because they did not need to evaluate the couples who were challenging Hawaii law. Instead, they testified about these issues in relation to the general population. Sociologist David Eggebeen, for example, testified on "changes or trends which have occurred in parenting, child bearing, and labor force behavior."⁵¹ Psychologist Charlotte Patterson's testimony included descriptions of two studies she had conducted on the children of lesbian and gay parents.⁵²

In contrast to these experts who were testifying about research findings, mental health professionals testifying about their evaluation of a parent in relation to a custody matter would need clinical training and experience in custody evaluations, as well as the ability to explain how their observations, assessments, instruments, inferences, and conclusions fit within a conceptual model that is based in scientific research and theory.⁵³

Some proposals for evaluating experts' qualifications seem not to be based on a discipline's criteria, but rather on subjective models of proper techniques and content that apparently the judge should incorporate through judicial notice.⁵⁴ Some seem difficult to apply in the context of a trial, but more importantly, models of evaluative techniques and content should be explicit so that the assumptions about what is "proper"

48. FED. R. EVID. 702.

49. *Baehr v. Miike*, Civ. No. 91-1394, 1996 WL 694235, *1 (Haw. Cir. Ct. Dec. 3, 1996), *rev'd and remanded*, 994 P.2d 566 (Haw. 1999). Interestingly, in this pre-*Daubert* case the expert testimony of one of the state's experts, a psychologist, was allowed but his testimony was found to be "not persuasive or believable because of his expressed bias against the social sciences. . .," as well as problematic due to his "severe views" such as his belief "that there is no scientific proof that evolution occurred." *Id.* at *8-9.

50. *Id.* at *8-9.

51. *Id.* at *6.

52. *Id.* at *12-13.

53. See GOULD, *supra* note 34, at 49-50, 75-76.

54. For discussion of judicial notice, see *infra* Part III.C.

can be examined. In the assessment of the expert's skills, for example, one approach recommended by a national interdisciplinary team is to examine child custody evaluation experts' skills of "commission and omission."⁵⁵ The judge is to determine if the expert is able:

[t]o interview children sensitively so as to understand and appreciate their individuality, their wants and their needs, without mistaking their own responses to those needs for the ones every parent should have, to be able to identify and understand childhood emotional disorders without assuming that any such disorder is disabling and/or inconsistent with maintenance of a satisfactory parent-child relationship.⁵⁶

The proposal did not indicate how a judge could undertake this skills assessment at a preliminary stage in a meaningful way, how counsel or the expert would know what criteria the judge was using, or why these were the most relevant skills-assessment criteria.

Determining the expert's training and knowledge of research in his or her field and assessing the expert's adherence to the field's professional guidelines would be a better, more scientifically-based approach, which could also be more fairly anticipated by counsel and the expert. For example, the experts in *Baehr* should have been knowledgeable about the requirements for literature reviews when they were summarizing the research in a particular area.⁵⁷ A psychologist who performs custody evaluations should be familiar with the American Psychological Association's (APA) *Guidelines for Child Custody Evaluations in Divorce Proceedings*, which include provisions related to personal and

55. NAT'L INTERDISCIPLINARY COLLOQUIUM ON CHILD CUSTODY, LEGAL & MENTAL HEALTH PERSPECTIVES ON CHILD CUSTODY LAW: A DESKBOOK FOR JUDGES § 27:2 (1998).

56. *Id.* Additional suggestions include that:

- a. The evaluator should know how to interview adults fairly, to diagnose any psychopathology in order to determine whether it has an adverse impact on parent-child relationships, to distinguish between clinical conditions and those inevitable and ordinary fears, traumas, and shortcomings of all adults that affect but do not undermine relationships with children and their ordinary growth and development to adulthood, to understand whether a parent can appreciate a child's special needs so as to provide for them.
- b. To evaluate parenting skills and to identify deficiencies in those skills without mistaking an unusual parenting style, or differences between the evaluator's (or "normal" people's) parenting styles and the subject's with a deficiency in parenting skill.

Id.

57. See *infra* Part IV.C.3.

societal biases,⁵⁸ methods of data gathering,⁵⁹ and interpretation of data.⁶⁰ Also of relevance are the *Specialty Guidelines for Forensic Psychologists* which state:

Forensic psychologists have an obligation to present to the court, regarding the specific matters to which they will testify, the boundaries of their competence, the factual bases (knowledge, skill, experience, training, and education) for their qualifications as an expert, and the relevance of those factual bases to their qualification as an expert on the specific matters at issue.⁶¹

This suggests that one aspect of the court's evaluation of an expert's qualifications should be how forthcoming and clear the expert is with

58. American Psychological Association, *Guidelines for Child Custody Evaluations in Divorce Proceedings*, 29 FAM. L.Q. 51, 55-56 (1995), available at 49 AMER. PSYCHOLOGIST 677 (1994) and <http://www.apa.org/practice/childcustody.html>:

The psychologist engaging in child custody evaluations is aware of how biases regarding age, gender, race, ethnicity, national origin, religion, sexual orientation, disability, language, culture and socioeconomic status may interfere with an objective evaluation and recommendations. The psychologist recognizes and strives to overcome any such biases, or withdraws from the evaluation.

Id.

59. *Id.* at 57:

The psychologist strives to use the most appropriate methods available for addressing the questions raised in a specific child custody evaluation, and generally uses multiple methods of data gathering, including, but not limited to, clinical interviews, observation and/or psychological assessments. Important facts and opinions are documented from at least two sources whenever their reliability is questionable. The psychologist, for example, may review potentially relevant reports, *e.g.*, from schools, health care providers, child care providers, agencies, and institutions. Psychologists may also interview extended family, friends, and other individuals on occasions when the information is likely to be useful. If information is gathered from third parties that is significant and may be used as a basis for conclusions, psychologists corroborate it by at least one other source wherever possible and appropriate, and document this in the report.

Id.

60. *Id.*

The psychologist refrains from drawing conclusions not adequately supported by the data. The psychologist interprets any data from interviews or tests, as well as any questions of data reliability and validity, cautiously and conservatively, seeking convergent validity. The psychologist strives to acknowledge to the court any limitations in methods or data used.

Id.

61. The Guidelines are endorsed by the American Academy of Forensic Psychology and are not an official statement of the American Psychological Association. The Guidelines are reprinted in GOULD, *supra* note 34, at 336-45 and in 15 LAW & HUM. BEHAV. 655 (1991). Guidelines have also been promulgated by the Association of Family and Conciliation Courts (www.afccnet.org). Additionally, there are a number of texts on custody evaluation protocol. *See, e.g.*, GOULD, *supra* note 34; PSYCHOLOGICAL EVALUATIONS FOR THE COURTS: A HANDBOOK FOR MENTAL HEALTH PROFESSIONALS & LAWYERS (Gary B. Melton et al., eds., 2d ed. 1997) [hereinafter PSYCHOLOGICAL EVALUATIONS].

regard to his or her own abilities in relation to the questions to be addressed.

3. RELIABILITY

In addition to determining whether an expert is qualified, the court must decide whether the testimony the expert will present is reliable. The issue of what standards courts should use to decide whether scientific expert testimony is admissible has been a subject of lively debate. Trial courts' evaluations differ along two dimensions.⁶² First, judges differ on how active a role they ought to take in screening the evidence that will be heard, as some judges are more willing than others to let expert evidence be presented. Second, courts differ on how to assess the evidence being offered. The two most prevalent models for assessing the proffered testimony in the United States are the "general acceptance test" developed in *Frye v. U.S.*⁶³ and the gate-keeper model posited in *Daubert*.⁶⁴

The succinctly stated *Frye* test does not provide much guidance to judges on how to determine whether something is generally acceptable:

Just when a scientific principle or discovery crosses the line between the experimental and demonstrable stages is difficult to define. Somewhere in this twilight zone the evidential force of the principle must be recognized, and while courts will go a long way in admitting expert testimony deduced from a well-recognized scientific principle or discovery, the thing from which the deduction is made must be sufficiently established to have gained general acceptance in the particular field in which it belongs.⁶⁵

Under *Daubert* general acceptance is not sufficient. Instead judges, as gate-keepers, must make their own assessment of the scientific merit of the testimony:

[T]he trial judge must determine at the outset . . . whether the expert is proposing to testify to (1) scientific knowledge that (2) will assist the trier of fact to understand or determine a fact in issue. This entails a preliminary assessment of whether the reasoning or methodology underlying the testimony is scientifically valid and of whether that reasoning or methodology properly can be applied to the facts in issue.⁶⁶

Although *Daubert* provided more guidance than *Frye*, it left a number of unanswered questions as well. *Daubert* did not clarify, for exam-

62. See FAIGMAN ET AL., *supra* note 13, § 1.10.

63. *Frye v. United States*, 293 F. 1013 (D.C. Cir. 1923).

64. *Daubert v. Merrell Dow Pharm., Inc.*, 509 U.S. 579, 597 (1993).

65. *Frye*, 293 F. at 1014.

66. *Daubert*, 509 U.S. at 592-93.

ple, whether a court could declare expert evidence inadmissible where the expert's methodology, data and application satisfied *Daubert* requirements, but the expert nonetheless reached a conclusion that was not in accord with the scientific community.⁶⁷ *Daubert's* definition of "science" has been criticized, as has the lack of clarity about whether a *Daubert* analysis should apply to expert testimony that falls outside the narrow definition of "science."⁶⁸ Some of these questions have been answered in subsequent cases, but uncertainties remain.⁶⁹

The focus of this paper, however, is not on precise standards for admissibility, but rather on using a gate-keeping model in family law cases generally to evaluate social science. Part IV of this paper provides evaluative tools for judges to use for this purpose. It is important to note briefly here, however, that there are some conceptual problems with the use of a gate-keeper model, which often appear when experts are used in family law cases. First, much of the research that the expert is relying on can present many challenges for the gate-keeping judge as it may be relatively new, may have methodological problems, or may not directly

67. See, e.g., John M. Conley & David W. Peterson, *The Science of Gatekeeping: The Federal Judicial Center's New Reference Manual on Scientific Evidence*, 74 N.C. L. REV. 1183, 1201 (1996). According to the Advisory Committee to the Federal Rules of Evidence, "[t]he Court in *Daubert* declared that the 'focus, of course, must be solely on principles and methodology, not on the conclusions they generate.' *Daubert*, 509 U.S. at 595. Yet as the Court later recognized, 'conclusions and methodology are not entirely distinct from one another.' General Elec. Co. v. Joiner, 522 U.S. 136, 146 (1997)." FED. R. EVID. 702 advisory committee's note. In *Joiner*, the Court went on to note that, "[a] court may conclude that there is simply too great an analytical gap between the data and the opinion proffered." *Joiner*, 522 U.S. at 146.

Michael Saks argues that *Joiner* should be interpreted to be consistent with *Daubert's* methodology and conclusions:

These grounds for excluding the testimony preserve *Daubert's* sensible distinction between methodology and conclusions. A person evaluating a study is capable of assessing the methodology of the study (its design, procedures, and measures) without looking at the study's results. This review of methodology permits a judgment about the weight that should be given [to] the data produced by the study. The person then can evaluate the inferences that the study's author or an expert witness draws from the data. Conclusions cannot stand unless they are rationally supported by the data. And the data cannot stand unless they are generated by sound methodology. One may reject the data and any conclusions based on those data because one finds the methodology unsound. One ought not reject data because one does not agree with the conclusions implied by the data. To the extent that it muddies those distinctions, *Joiner* leads us astray.

Saks, *supra* note 6, at 236.

68. See, e.g., Conley & Peterson, *supra* note 67, at 1201-02. This issue was addressed in *Kumho Tire Co. v. Carmichael*, 526 U.S. 137 (1999). See FED. R. EVID. 702 advisory committee's note ("In *Daubert* the Court charged trial judges with the responsibility of acting as gatekeepers to exclude unreliable expert testimony, and the Court in *Kumho* clarified that this gatekeeper function applies to all expert testimony, not just testimony based in science.").

69. Saks, *supra* note 6.

relate to the question the expert is asked to address.⁷⁰ Second, the expert's testimony is often primarily about the expert's clinical assessment of an individual, rather than about the underlying research that supports the clinical methodology and conclusions.⁷¹ Third, usually the judge, not a jury, is the fact-finder.

The use of the judge as fact-finder is a major distinction between many family law cases and jury-decided cases addressed in post-*Daubert* analyses. In family cases, judges are screening the expert to determine if they themselves should hear the expert's testimony. Many post-*Daubert* analyses address how a judge should proceed in jury trials, recommending for example, that the court hold a separate hearing away from the jury to decide on the admissibility of expert testimony.⁷² Because family cases typically proceed without a jury, the judge in a family matter might be more likely to permit expert testimony, once the expert is qualified, rather than separately assessing whether the expert's testimony should be admissible. The judge would not spend time holding a separate hearing on admissibility (a *Daubert* hearing), but rather at trial the judge could simply give little or no weight to the expert's testimony if it became clear that an evaluation was performed improperly or there were other problems with the testimony.⁷³

In addition to saving time, judges might have other motives for this practice. Judges have noted, for example, that allowing experts to testify has the advantage of granting to litigants who have paid for the experts a sense that they have been fully heard, which is an important component of fairness.⁷⁴ In addition, many appellate courts do not consider the receipt of inadmissible evidence over an objection to be a basis for reversal if there is admissible evidence to support the court's findings.⁷⁵ In such a situation, the appellate court presumes that the judge only considered the admissible evidence.⁷⁶ However, if the trial judge refuses to

70. See *infra* Part IV; see also Margaret F. Brinig, *Promoting Children's Interests Through a Responsible Research Agenda*, 14 U. FLA. J.L. & PUB. POL'Y 137 (2003).

71. See *infra* Part III.A.

72. The separate hearing "entails a preliminary assessment of whether the reasoning or methodology underlying the testimony is scientifically valid and of whether that reasoning or methodology properly can be applied to the facts in issue." *Daubert v. Merrell Dow Pharm., Inc.*, 509 U.S. 579, 592-93 (1993). Indeed, a major justification for having the screening is skepticism about the capacity of the jury. See, e.g., 1 MCCORMICK ON EVIDENCE § 60 (John W. Strong ed., 5th ed. 1999).

73. Judges' abilities to disregard nonscientific supported expert testimony has been questioned. See, e.g., Shuman, *supra* note 11, at 568-69. See generally PETER HUBER, GALILEO'S REVENGE: JUNK SCIENCE IN THE COURTROOM (1991).

74. See TOM R. TYLER ET AL., SOCIAL JUSTICE IN A DIVERSE SOCIETY (1998).

75. 1 MCCORMICK ON EVIDENCE, *supra* note 72, § 60.

76. *Id.*

admit evidence that should have been admitted, reversal is more likely.⁷⁷ This more liberal admissibility approach has the disadvantage that evidentiary rulings may not be as explicit as those resulting from *Daubert* hearings, where the content of the expert's testimony may be hotly contested.

Azia v. DiLascia is an example of a court allowing an expert to testify, but then giving the testimony little weight.⁷⁸ Interestingly enough, the trial court justified its treatment of the expert's testimony by taking judicial notice of professional standards. The expert's lack of adherence to the APA's *Guidelines for Child Custody in Divorce Proceedings* and the *Specialty Guidelines for Forensic Psychologists* was an important factor in the trial court's consideration of the weight to be given to the expert's testimony.⁷⁹ The expert in *Azia* had served in a capacity that was similar to a court-appointed custody evaluator, but subsequently became the child's therapist. The trial court determined that it would accord "her expert opinion little weight," noting the ethical restraints placed on dual roles and the expert's failure to consider these.⁸⁰ On appeal, the plaintiff challenged the trial court's judicial notice of the APA ethical rules.⁸¹ The appellate court held that "[t]he ethical rules applicable to the profession of a witness are permissible for judicial notice because a professional, who is a member of an association, is held accountable to know those ethical rules."⁸² Further, the appellate court held that the defendant in fact had notice that the expert's ethics were at issue because of the questions the judge had asked at trial.⁸³

In contrast, consider the more formal *Daubert*-hearing used in *Dah-*

77. *Id.*

78. *Azia v. DiLascia*, 780 A.2d 992 (Conn. App. Ct. 2001).

79. *Azia v. DiLascia*, No. FA 980548239S, 1999 WL 989461, at *9 (Conn. Super. Ct. Oct. 18, 1999).

80. *Id.* at *9.

81. *Azia*, 780 A.2d at 995.

82. *Id.* at 1005.

83. *Id.* The questions included the following colloquy:

The Court: And you are a member of like the American Psychological Association or something like that?

[Hanley-Kallen]: Yes, Your Honor.

The Court: And do you live by their rules on ethics and stuff like that?

[Hanley-Kallen]: Yes.

The Court: Any organization other than the APA?

[Hanley-Kallen]: Diplomat status in child psychology and child custody evaluation.

The Court: Does that subject you to any other association rules on ethics?

[Hanley-Kallen]: Yes.

The Court: Who's that?

[Hanley-Kallen]: The American Forensic Association. I just became diplomat. I am not sure of the initials.

lin v. Evangelical Child and Family Agency, a tort action by adoptive parents for medical expenses and emotional damages against an adoption agency for nondisclosure of information about their adopted child.⁸⁴ Prior to trial the court heard a number of in limine motions related to admissibility of expert testimony. The plaintiff parents, for example, moved to bar the testimony of an agency expert who was to testify on the history of adoption practice and on his opinion that the child suffered from “reactive attachment disorder,” which the agency characterized as “a condition prevalent among adopted children which makes them [sic] unable to form attachments with other people, including the adoptive parents who care for them.”⁸⁵ The agency apparently was contending that the child’s problems following adoption were attributable to “factors common among adoptive children,” rather than to the child’s own “psychological dysfunction.”⁸⁶ The court held that the expert’s written report, “which was required to ‘contain a complete statement of all opinions to be expressed . . . and the basis and reasons therefor [and] the data or other information considered by [the expert] in forming the opinions,’” failed to provide a sufficient foundation for the expert rendering a diagnosis of the child.⁸⁷ The court examined what the expert had actually done and his professional training and noted that among other failings, the expert had not examined the child and he was not a psychiatrist or clinical psychologist. The court concluded that the expert’s opinion was inadmissible because neither the expert nor the agency had “provided the Court with anything that even suggests, let alone shows, that [the expert’s] opinion regarding [the child’s] condition is ‘the product of reliable principles and methods’ that ‘the witnesses [sic] has applied . . . reliably to the facts of the case.’”⁸⁸ The expert was allowed to testify about “attachment problems generally and how they manifest themselves.”⁸⁹ Unlike the fact-finder judge in *Azia*, the fact-finder jury in *Dahlin* would hear only that expert testimony that had passed *Daubert* screening.

The Court: Did you consult with the ethical rules before making the decision to become [the child’s] therapist?

[Hanley-Kallen]: No, I did not.

Id. at 1005 n.15.

84. *Dahlin v. Evangelical Child & Family Agency*, No. 01 C 1182, 2002 WL 31834881, *1 (N.D. Ill. Dec. 18, 2002).

85. *Id.* at *7.

86. *Id.*

87. *Id.* (emphasis in original) (quoting FED. R. CIV. P. 26(a)(2)(B)).

88. *Id.* (quoting FED. R. EVID. 702).

89. *Id.* at *8.

4. TESTIMONY ON THE ULTIMATE QUESTION

Another dispute surrounds the question of whether the expert should be allowed to testify to the ultimate question being considered by the court - a question that has particularly arisen in custody cases.⁹⁰ For example, should the expert testify as to what custody placement is in the child's "best interests" if that is the legal standard for the decision, or should the expert testify only as to such matters as the parent-child relationship, the results of psychological testing, and similar areas which could be grounded in professional, peer-reviewed literature?

Courts and lawyers may want the expert's opinion on the ultimate issue and may characterize it as a valuable piece of information to aid a judge in forming an opinion. Some scholars, however, argue strongly that "mental health professionals ordinarily should refrain from giving opinions as to ultimate legal issues" and suggest that "there is near-unanimity among scholarly commentators" on this point.⁹¹ They also note, however, that most jurisdictions allow and some may require the expert to recommend which placement is best for the child.⁹² The major professional organizations have not taken a position on this issue.⁹³ For example, the APA *Guidelines for Child Custody Evaluations in Divorce Proceedings* state:

While the profession has not reached consensus about whether psychologists ought to make recommendations about the final custody determination to the courts, psychologists are obligated to be aware of the arguments on both sides of this issue, and to be able to explain the logic of their position concerning their own practice.⁹⁴

Those opposed to the expert offering an opinion on "best interests" may note that expert testimony on the "ultimate issue," which was once prohibited at common law, is still characterized as unhelpful and can be excluded on that basis.⁹⁵ Using a *Daubert* analysis should likewise result in the exclusion of the expert's testimony on "best interests." A

90. MARC J. ACKERMAN & ANDREW W. KANE, *PSYCHOLOGICAL EXPERTS IN DIVORCE ACTIONS* § 1.1 (3d ed. 1998).

91. *PSYCHOLOGICAL EVALUATIONS*, *supra* note 61, §1.04, at 17 (emphasis in original).

92. ACKERMAN & KANE, *supra* note 90, § 1.1. A survey of Michigan judges and attorneys found that 84% of the judges thought that custody evaluations should provide recommendations about custody. James N. Bow & Francella A. Quinnell, *Critique of Child Custody Evaluations by the Legal Profession*, 42 FAM. CT. REV. 115 (2004).

93. ACKERMAN & KANE, *supra* note 90, § 1.1.

94. American Psychological Association, *supra* note 58, at 58. The forensic guidelines do not directly address this issue.

95. *See, e.g.*, CHRISTOPHER B. MUELLER & LAIRD C. KIRKPATRICK, *FEDERAL EVIDENCE* § 361 (2d ed. 1994) (noting also that FED. R. EVID. 704 abrogated the common law prohibition on "ultimate issue" testimony). Generally, also, an expert is not allowed to express "legal conclusions." The best interest standard, however, is a blend of fact and law, which makes this objection less salient. *Id.*

decision on the ultimate issue requires legal analysis and consideration of many factors, some of which may not be known to the expert. Because the best interest standard is admittedly indeterminate, it is not possible to critically assess the expert's predictions on which outcome would serve the child's best interests. In other words, the expert's opinion would not have scientific validity and should not be allowed under *Daubert*.

5. STANDARD OF REVIEW

Generally, barring an abuse of discretion on the part of the trial judge, an appellate court does not overturn the trial court's determination that a witness is qualified as an expert.⁹⁶ The standard of review for the decision on admissibility is also abuse of discretion, although several scholars have argued that under a gate-keeper model this should be a de novo review, similar to an appellate court's review of a matter of law.⁹⁷ Under this de novo approach, the appellate court decision would then be binding on lower courts. Michael Saks, for example, argues that general admissibility decisions about scientific theories and their general applications should be the same across cases and that only the specific applications should vary by case.⁹⁸ If admissibility questions are similar across cases, rulings on admissibility should be treated as matters of law to ensure consistency from case-to-case. In his view, this approach would be both more efficient and more coherent.⁹⁹

B. Briefs

In addition to accessing social science through experts, courts also receive social science information in briefs. Indeed, the Supreme Court's use of social science often is dated from its 1908 decision in *Muller v. Oregon*.¹⁰⁰ In *Muller*, Louis Brandeis, representing the state, filed a brief that provided extensive research documentation related to

96. See, e.g., FAIGMAN ET AL., *supra* note 13, § 1-3.3.2; *Gen. Elec. Co. v. Joiner*, 522 U.S. 136 (1997) (holding that abuse of discretion is the proper standard by which to review a lower court's decision to admit or exclude scientific evidence).

97. See, e.g., *In re Adoption of Hugo*, 700 N.E.2d 516 (Mass. 1998) (challenging both the qualifications and the reliability of the expert's methodology, but no abuse of discretion by the trial judge was found with regard to admission of the expert's testimony). The *Hugo* court noted that the expert who testified on the child's placement "had reviewed the case file, interviewed the parties, and gathered information from service providers, a methodology strikingly similar to that used by [the other parties'] own experts" and that "there was adequate support in the record for the judge's conclusion that [the expert's] testimony was based on a reliable methodology." *Id.* at 526. See also, e.g., FAIGMAN ET AL., *supra* note 13, § 1-3.6; *Social Authority*, *supra* note 18.

98. Saks, *supra* note 6, at 231-35.

99. *Id.*

100. *Muller v. Oregon*, 208 U.S. 412 (1908).

the challenged labor law, which provided special protections for women.¹⁰¹ The Court took judicial notice of the material in the brief, which demonstrated that women's physical structure and family responsibilities justified protective legislation. Judicial notice of this "general knowledge" and "widespread belief" allowed the Court to conclude that there was a rational basis for the Oregon legislation.¹⁰²

1. LEVEL OF SPECIFICITY AND CHOICE OF TRIAL OR APPELLATE LEVEL

Courts can receive social science information in briefs at all three levels of specificity: the general-rule level, the intermediate level, and the case-specific level.¹⁰³ Different rules apply, however, depending on whether the brief is introduced at the trial or the appellate stage. The Brandeis brief contained materials that had not been presented at trial; even today there is no procedural bar to including new social science materials in an appellate brief or argument, so long as the material presented is at the general level of law formation.¹⁰⁴ Whether attorneys obtain the best result for their clients by waiting until the appellate stage to present social science or other non-legal materials is an issue debated by advocates. One scholar argues that such information is best introduced at the appellate level in a brief, where the lawyer may use social science research to justify a particular policy choice.¹⁰⁵ She thinks that appellate courts are more likely than trial courts to incorporate these materials into their decision making.¹⁰⁶ An opposing view is that presenting social science evidence at the trial level is more advantageous than waiting until the appellate level, because attorneys who present material at trial have more control over the fact-finding process and lessen the likelihood that judges will seek additional information without the assistance of the parties.¹⁰⁷

When the brief is at the intermediate or case-specific level, rather than the general-rule level, the use of social science information that was not introduced into evidence becomes more problematic. An appellate brief that examines the facts of the particular case in relation to social

101. *Id.* at 419.

102. The decision was in the *Lochner* era, when the Court closely scrutinized and invalidated labor legislation as "an unreasonable, unnecessary, and arbitrary interference with the right and liberty of the individual to contract in relation to his labor, and as such was . . . void under, the Federal Constitution." *Id.* at 418. See *Lochner v. New York*, 198 U.S. 45 (1905).

103. See *infra* Part III, Table 1.

104. See Ellie Margolis, *Beyond Brandeis: Exploring the Uses of Non-Legal Materials in Appellate Briefs*, 34 U.S.F. L. REV. 197, 202, 205 (2000).

105. *Id.* at 210-19.

106. *Id.* at 218.

107. John Frazier Jackson, *The Brandeis Brief—Too Little, Too Late: The Trial Court as a Superior Forum for Presenting Legislative Facts*, AM. J. TRIAL ADVOC. 1, 2-3 (1993).

science could only include facts that were in the trial record. Further, it may be argued that the relationship of the social science to the case facts should also be viewed as a factual issue, which, therefore, could not be presented for the first time at the appellate level.¹⁰⁸

2. BIAS

From the point of view of the judicial system, it is not clear whether experts' testimony or briefs result in a more accurate presentation of social science information. Expert testimony has been criticized for being biased and incomplete.¹⁰⁹ The effectiveness of cross-examination and other adversary system tools designed to eliminate bias have been criticized as well.¹¹⁰ Briefs, however, have also been criticized for bias and incomplete reviews of social science research. Amici briefs filed by special interests groups too often contain "junk social science" which is "characterized by quotes from social scientific research taken out of context, misleading statistical presentations, denigration of studies whose results conflicted with the argument, and anecdotes masquerading as social science findings."¹¹¹

Unfortunately, even briefs that have been written by seemingly neutral social scientists have been criticized for biased presentations. A recent example is the amicus brief filed by Judith Wallerstein in the California relocation case of *In re Marriage of Burgess*.¹¹² Dr. Wallerstein's brief has been criticized for reviewing only ten social science articles, seven of which were from her own research group, rather than doing a comprehensive review;¹¹³ for misinterpreting the findings of

108. See *infra* Part III.C.

109. See, e.g., Davis, *supra* note 9, at 1556.

110. *Id.*

111. Michal Rustad & Thomas Koenig, *The Supreme Court and Junk Social Science: Selective Distortion in Amicus Briefs*, 72 N.C. L. REV. 91, 128 (1993) (examining amici briefs filed on tort punitive damage issues). See, e.g., Brief of Amicus Curiae Rutherford Institute at ix., 1, 13, Cox v. State Dep't of Health & Rehabilitative Servs., 656 So. 2d 902 (Fla. 1995) (No. 82-967) (supporting a ban on homosexual adoption by stating that scientific research "conclusively establishes that homosexual households present an unhealthy, insecure environment for children," but also stating that "studies on the effects of homosexual parenting of children are scant, highly publicized, and conducted largely by lesbian researchers in very limited samples," and that scientific researchers who have reservations about placing a child in a homosexual household "would be unable to develop a rapport with homosexual parents and thus gain the data necessary for a study").

112. *In re Marriage of Burgess*, 913 P.2d 473 (Cal. 1996). Following the case decision, an article was written adapted from the brief. Judith S. Wallerstein & Tony J. Tanke, *To Move or Not to Move: Psychological and Legal Considerations in the Relocation of Children Following Divorce*, 30 FAM. L.Q. 305 (1998).

113. Richard A. Warshak, *Social Science and Children's Best Interests in Relocation Cases: Burgess Revisited*, 34 FAM. L.Q. 83 (2000).

some studies;¹¹⁴ for failing to note the limitations of the social science research she reviewed;¹¹⁵ and for “ignoring the broad consensus of professional opinion” that would not support her conclusion in favor of the custodial parent’s control over relocation.¹¹⁶ In spite of these flaws, Dr. Wallerstein’s brief was relied on in *Burgess* and has been very influential in other relocation cases, such as the New Jersey case, *Baures v. Lewis*, and possibly also the New York case, *Tropea v. Tropea*.¹¹⁷

In a relocation case subsequent to *Burgess*, the California Supreme Court was provided with two competing amici briefs from social scientists and mental health experts.¹¹⁸ Both unfortunately took advocacy positions, rather than emphasizing the limited contribution current social science research can make to the relocation debate. The brief authored by psychologist Richard Warshak, for example, takes the position, which is unsupported by any of the social science research it reviews, that custodial parents’ relocation requests “would not be proposed unless the moving parent placed a higher value on the anticipated gains of the move than on the nonmoving parent’s regular involvement in the fabric of the children’s lives.”¹¹⁹ It recommends that California relocation law adopt a case-by-case comprehensive examination of the potential detriment of relocation, but does not recommend or even address whether the court should consider the potential detriment of non-relocation or of a custodial change. Discouragingly, it relies heavily on a very methodologically weak survey of Arizona State University psychology students whose parents had divorced to support a position that relocation is detrimental. In addition, to support a suggestion that a court’s denial of a relocation request would result in a custodial parent not moving in a majority of cases, it relies on a survey of seventy-two family law attorneys who happened to be attending a continuing legal education pro-

114. *Id.* at 87.

115. *Id.* at 109.

116. *Id.* at 85.

117. *Baures v. Lewis*, 770 A.2d 214 (N.J. 2001); *Tropea v. Tropea*, 665 N.E.2d 145 (N.Y. 1996). See Sanford L. Braver et al., *Relocation of Children After Divorce and Children’s Best Interests: New Evidence and Legal Considerations*, 17 J. FAM. PSYCH. 206, 210 (2003) (noting that the social science articles cited in *Baures* are, with one minor exception, the same as those cited in Wallerstein & Tanke, *supra* note 112); see also Janet Leach Richards, *Children’s Rights vs. Parents’ Rights: A Proposed Solution to the Custodial Relocation Conundrum*, 29 N.M. L. REV. 245, 259-60 (1999).

118. *In re Marriage of LaMusga*, 88 P.3d 81 (Cal. 2004).

119. Amici Curiae Brief of Richard A. Warshak et al. on behalf of LaMusga Children at 15, *In re Marriage of LaMusga*, 88 P.3d 81 (Cal. 2004) (No. S107355), available at http://www.atybrief.com/volexports-/lamusga/warshak_brief.pdf. The amici on the Warshak brief are eighteen social science researchers and ten mental health forensic practitioners. The other brief was authored by Judith Wallerstein, who was joined by five mental health professionals.

gram and were willing to fill out a fifteen minute questionnaire!¹²⁰ Further, the brief supports the recommendations of the one expert in the case on what outcome was in the children's best interests, without noting that "best interests" is an indeterminate legal standard that goes beyond the knowledge of the expert.¹²¹

Briefs filed by social science organizations, such as the American Psychological Association (APA), have been criticized as well. Criticisms of briefs have included a selective use of theories to support a particular position, acceptance of one party's version of disputed facts, and incomplete review of the research.¹²² Even the critics, however, are in favor of organizations such as the APA submitting briefs and feel the briefs can "make an important contribution to the legal process."¹²³ Indeed, the APA has been praised for being one of the few professional associations of social scientists that routinely submits amicus briefs.¹²⁴

A better approach would be to have social scientists, who are committed to being as unbiased as possible, author briefs on family law issues. Absolute neutrality is an unobtainable goal, since there is bias in the formation of research questions, the conduct of research, and the review of research.¹²⁵ However, bias in research and research reviews can be reduced. The brief writer, like the writer of a research review, should clearly indicate the purpose of the brief and the methodology used to select studies, and should assess their substantive findings and methodological characteristics.¹²⁶ The research used in the brief should be judged by its quality, the relation of quality to findings, and its relevance to the policy questions being addressed.¹²⁷ The brief writer should ask peers to review the brief. It has been proposed that professional journals give amicus briefs priority in publication so that the brief's authors would be aware that peer review would be forthcoming,

120. *Id.* at 15-16 (citing Sanford L. Braver et al., *Relocation of Children After Divorce and Children's Best Interests: New Evidence and Legal Considerations*, 17 J. FAM. PSYCH. 206 (2003) and Sanford L. Braver et al., *Experiences of Family Law Attorneys with Current Issues in Divorce Practice*, 51 FAM. REL. 325 (2002)).

121. *Id.* at 18-21. For discussion of expert testifying to the ultimate issue, see *infra* Part III.A.4.

122. Gerald V. Barrett & Scott B. Morris, *The American Psychological Association's Amicus Curiae Brief in Price Waterhouse v. Hopkins: The Values of Science Versus the Values of the Law*, 17 LAW & HUM. BEHAV. 201 (1993).

123. *Id.* at 202.

124. See Rustad & Koenig, *supra* note 111, at 153. A list of APA briefs is available at <http://www.psyclaw.org>.

125. See *infra* Part IV; Ramsey & Kelly, *supra* note 10, at 671 (noting that bias can be reduced, but not eliminated).

126. Ramsey & Kelly, *supra* note 10, at 682.

127. *Id.* at 682-83.

even if it did not occur prior to the submission of a brief to a court.¹²⁸

These briefs should concern themselves with the issues of scientific method discussed in Part IV and they should model themselves on scientific literature reviews, also briefly discussed in Part IV.¹²⁹ In our conclusion (Part V), we suggest how courts might obtain higher quality research review briefs.

C. *Judicial Notice*

The distinction made by Professor Kenneth Culp Davis in the 1940's between "adjudicative facts" and "legislative facts" is still used today to define appropriate content for judicial notice.¹³⁰ Adjudicative facts were those that the court found "concerning the immediate parties - what the parties did, what the circumstances were."¹³¹ When a court, however, "wrestles with a question of law or policy, it is acting legislatively . . . and the facts which inform its legislative judgment may conveniently be denominated legislative facts."¹³²

The Davis adjudicative fact/legislative fact dichotomy has been maintained in the Federal Rules of Evidence with regard to judicial notice. Rule 201 strictly limits the judicial notice of adjudicative facts, but does not apply to the judicial notice of legislative facts. Relying heavily on Davis' work, the Advisory Committee to the Federal Rules of Evidence distinguished between adjudicative and legislative facts as follows:

Adjudicative facts are simply the facts of the particular case. Legislative facts, on the other hand, are those which have relevance to legal reasoning and the lawmaking process, whether in the formulation of a legal principle or ruling by a judge or court or in the enactment of a legislative body.¹³³

Concerning adjudicative facts, under Rule 201 "[a] judicially noticed fact must be one not subject to reasonable dispute in that it is either (1) generally known within the territorial jurisdiction of the trial court or (2) capable of accurate and ready determination by resort to sources whose accuracy cannot reasonably be questioned."¹³⁴

128. Rustad & Koenig, *supra* note 111, at 211.

129. *See infra* note 157 and accompanying text.

130. Kenneth Culp Davis, *An Approach to Problems of Evidence in the Administrative Process*, 55 HARV. L. REV. 364 (1942).

131. *Id.* at 402.

132. *Id.*

133. FED. R. EVID. 201 advisory committee's note, subdivision (a).

134. FED. R. EVID. 201(b), Kinds of Facts. Federal Rule of Evidence 201 further provides:

(c) When discretionary. A court may take judicial notice, whether requested or not.

(d) When mandatory. A court shall take judicial notice if requested by a party and supplied with the necessary information.

In contrast to the restraints on adjudicative facts, the Rules do not restrain judicial notice of legislative facts and, in fact, the Advisory Committee endorsed a broad based authority to consider legislative facts. The committee quoted with approval a law review article by Professor Morgan:

In determining the content or applicability of a rule of domestic law, the judge is unrestricted in his investigation and conclusion. He may reject the propositions of either party or of both parties. He may consult the sources of pertinent data to which they refer, or he may refuse to do so. He may make an independent search for persuasive data or rest content with what he has or what the parties present. . . . [T]he parties do no more than to assist; they control no part of the process.¹³⁵

The Advisory Committee's Note goes on to state:

This is the view which should govern judicial access to legislative facts. It renders inappropriate any limitation in the form of indisputability, any formal requirements of notice other than those already inherent in affording opportunity to hear and be heard and exchanging briefs, and any requirement of formal findings at any level.¹³⁶

Our focus is on the use of social science research, but courts may consider history, science, and other materials "for informing the court's legislative judgment on questions of law and policy."¹³⁷

This broad access to judicial notice of legislative facts in the family law area may create confusion and unfairness. The distinction between legislative and adjudicative facts is not always clear and this problem is exacerbated when the judge, not a jury, is the factfinder. To illustrate the difficulty of distinguishing between adjudicative and legislative facts, consider the following hypothetical case:

Because of her mother's incarceration, Ann went to live with Jane, her mother's sister, when Ann was age one. Jane became Ann's legal

(e) Opportunity to be heard. A party is entitled upon timely request to an opportunity to be heard as to the propriety of taking judicial notice and the tenor of the matter noticed. In the absence of prior notification, the request may be made after judicial notice has been taken.

(f) Time of taking notice. Judicial notice may be taken at any stage of the proceeding.

(g) Instructing jury. In a civil action or proceeding, the court shall instruct the jury to accept as conclusive any fact judicially noticed. In a criminal case, the court shall instruct the jury that it may, but is not required to, accept as conclusive any fact judicially noticed.

FED. R. EVID. 201(c)-(g).

135. FED. R. EVID. 201 advisory committee's note (quoting Edmund J.Morgan, *Judicial Notice*, 57 HARV. L. REV. 269, 270-71 (1944)).

136. FED. R. EVID. 201 advisory committee's note.

137. Davis, *supra* note 130, at 404.

guardian and custodian. Ann has lived with Jane for seven years. The mother has now been released from prison, but Jane has refused to return Ann. Assume that the court must determine custody based on “the best interests of the child.” Assume further that:

1. The court-appointed child psychologist evaluates Ann, Jane, and the mother and provides an extensive written evaluation, which is entered into evidence, and has given expert testimony at trial about Ann’s emotional well-being and the parenting abilities of Jane and the mother.

2. There are child development theories and studies related to the impact of parent-child bonding and attachment, psychological parents, and connection with blood relatives on the child’s well-being, but there is some dispute in the field about the conclusions to be drawn from these studies. The judge is familiar with this literature in a general way, but the theories and empirical work testing those theories are not addressed by the expert’s testimony.

3. The judge, not a jury, will determine the facts and decide the case by analyzing the law related to custody and best interests in relation to the facts of this case and will decide who should have custody.

Based on this hypothetical, can the judge take judicial notice of the material in paragraph two, namely child development theories and studies related to the impact on children’s well-being of parent-child bonding and attachment, psychological parents, and connection with blood relatives? There are four potentially correct answers.

The judge may be able to take judicial notice of the studies because this is part of the court’s consideration of legislative facts, relevant to “the lawmaking process,” as the material will help the judge reach a general rule of law related to balancing the claims of a biological parent and those of a primary caretaker, in relation to the well-being of the child.

On the other hand, these studies could be viewed as adjudicative facts that are specific to this case, which should not be judicially noticed. The judge can take judicial notice only if the facts are not subject to reasonable dispute. In this situation, according to the hypothetical case, there is some dispute over the meaning of these studies. Further, it is likely that there would be even more disagreement about how the judge applied the studies to the situation of the parties.

Another possibility is that the judge could take judicial notice of the studies for some purposes, but not for others. If the judge uses the research as background information when the judge formulates a general rule of law, the research materials would be legislative facts, appropriate for judicial notice. The same facts, however, would be adjudicative

facts if the judge weighs them in deciding this case. The judge must rely on the parties to provide the adjudicative facts and therefore must instruct herself to disregard the paragraph two materials to the extent that they are not part of the record, if she uses the studies as an adjudicative fact.¹³⁸

Alternatively, the judge could conclude that judicial notice is not an issue. In designing and conducting the evaluation, in producing the report, and in testifying, the expert was using the material in paragraph two. If the parties wanted to challenge that material or make it more explicit, they had the opportunity. The judge may therefore use the paragraph two material as the judge sees fit.

This example illustrates the difficulty in applying these judicial notice rules to many family law cases. The rules of evidence may be more relaxed and more often than not there is no jury, hence the judge is both a fact-finder and a decision-maker.¹³⁹ The distinction between legislative and adjudicative facts is thus a matter of degree, rather than a bright line.¹⁴⁰

Another substantial problem with a court taking judicial notice of

138. Michael J. Saks, *Judicial Attention to the Way the World Works*, 75 IOWA L. REV. 1011, 1017-18 (1990) (noting that “[t]he facts labeled ‘adjudicative’ and ‘legislative’ may be identical,” and “[t]he legal function they serve determines what we call them and what courts may or must do with them”).

139. Consider, for example, a family court judge as the fact-finder, proceeding through the processes Professors Monahan and Walker envision for the use of “social frameworks” in court. They propose that information on social frameworks would come to the court through briefs filed by the parties or from the court’s own research. The judge would evaluate the social science to determine if use of the framework was proper. If the court approved the social science, the court would then communicate its empirical conclusion as a jury instruction, just as the court would communicate its conclusion on the substantive law. The framework instruction would address the following four aspects of the court’s empirical conclusions: “the factual determination that is being framed or put in context for the jury by the research;” “the factors (or ‘variables’) found in the research that bear upon the determination that the jury is to make;” “the form of the relationship that exists between or among the identified factors;” and “the magnitude of the relationship that is addressed in the empirical framework.” *Social Frameworks*, *supra* note 18, at 595.

140. Kenneth C. Davis proposed the following distinction:

When facts used in lawmaking are (1) narrow and specific, (2) central or critical, (3) controversial, (4) unmixd with judgment or policy, (5) provable, and (6) in some degree about the parties or known mainly by them, the parties clearly should have a pre-decision chance to challenge them, and the requirement may even be imposed in the name of due process. But when the facts are (1) broad and general, (2) background or peripheral, (3) noncontroversial, (4) mixed with judgment or policy, (5) not easily provable, and (6) wholly unrelated to the parties, a court or agency may use them without even advertng to any possible problem of procedural fairness. When the six items are mixed - some at one end of the scale, some at the other, some toward the middle - the two questions of how much support the facts should have and of what procedural protections to give the parties may be difficult.

Kenneth Culp Davis, *Facts in Lawmaking*, 80 COLUM. L. REV. 931, 933 (1980).

social science to formulate a general rule of law is that a party may have no opportunity to challenge the court's assessment of social science. Judges obtain knowledge from a variety of sources, such as judicial training programs, books, magazines, television, public lectures, and interaction with friends, colleagues, and family. These off-the-record sources of information can be problematic if judges then use this information as determinative facts in forming their opinions, with no advance notice to the parties. A startling example of the parties being unaware that the judge had taken judicial notice of facts that would be used in the decision was identified by Professor (formally judge) Peggy Davis:

Ross v. Hoffman illustrates both the haphazard way in which extra-record literature infects judicial decisionmaking and the reluctance of judges to control the process. The trial judge's deliberations proceeded as follows:

After reviewing the evidence adduced by the parties, the chancellor [who had decided this custody matter below] explained that his reasoning had, to some extent, been influenced by certain educational background factors to which he had been exposed extra-judicially, as that term applies to the evidentiary confines of this case. He said: "Now, very briefly, there is a book out, which is widely read, by three very well-respected professional doctors, Drs. Goldstein, Freud and Solnick [sic], called 'Beyond the Best Interests of the Child' and in that book they point out that whether any adult becomes a psychological parent over the child is based upon a day-to-day interaction, companionship and shared experiences. And if you look at it from that view, Mrs. Hoffman has had this advantage."

On appeal, Ross complained about the chancellor's reliance upon psychological parent theory, arguing that she had had no notice that the theory was at issue and no opportunity to refute it. Intermediate and ultimate appellate courts both dismissed her protest. These courts found "no error in the fact that a trial judge continues his general education by reading, or that his reasoning is influenced by such education or by his experiences during his lifetime."¹⁴¹

What's to be done about these judicial notice concerns? Professor Davis suggested that:

Some device is needed to focus the attention of judges and litigants upon the possibility that background and legislative fact issues will arise; to encourage judges and litigants to address these issues directly rather than leave them sub rosa; and to suggest, in broad outline, means for assuring that judicial notice is undertaken in a manner that assures fairness and informed deliberation. The *laissez-faire* policy with respect to judicial notice of background and legisla-

141. Davis, *supra* note 9, at 1595-97.

tive facts has failed. Rules of evidence should meet this important issue with something other than silence.¹⁴²

She proposes that the court's judicial notice authority be explicitly recognized and that the court's authority to "entertain or solicit special briefs, argument, affidavits or depositions of experts, or, in the rare case, hearings, as to the propriety of taking judicial notice of a disputable fact" be codified.¹⁴³ Further, she recommends that "[a] trial court's decision whether to exercise this authority should be reviewable and informed by standards."¹⁴⁴

Davis' approach has much in common with suggestions made by Federal District Court Judge Jack Weinstein. Judge Weinstein suggests that "[t]he key in this area is openness and balance. Whenever possible, materials and notices of work and studies should be filed and docketed or announced at sessions with the attorneys and experts. Parties must have the opportunity to counter these extra-judicial sources of knowledge."¹⁴⁵ He further suggests that:

Encouraging and allowing judges to learn as much as they can about the world, their craft and the cases before them is desirable. Some limits and caveats are required to give the parties a fair opportunity to meet the judges' possible misperceptions and to assure the public of an unbiased and fair minded judiciary.¹⁴⁶

To implement these recommendations, Judge Weinstein's suggestions

142. *Id.* at 1603 (footnotes omitted). Professor Davis' article was criticized by Saks, *supra* note 138. His criticism, for the most part, does not really address the questions she raised. He notes that "the peculiar class of cases" that she studied do not tell us much about what judges do generally with certain uses of social science; this may be true, but unfortunately his analysis does not tell us much about how to deal with judicial notice in these "peculiar" cases—namely family law cases. Saks, *supra* note 138, at 1025.

143. Davis, *supra* note 9, at 1603.

144. *Id.*

145. Jack B. Weinstein, *Limits on Judges Learning, Speaking and Acting-Part I-Tentative First Thoughts: How Many Judges Learn?*, 36 ARIZ. L. REV. 539, 560 (1994). See also Judge Weinstein's decision in *Bulova Watch Co. v. K. Hattori & Co.*, 508 F. Supp. 1322 (E.D.N.Y. 1981). In *Bulova* he decided "[i]n view of the extensive judicial notice taken, based partly upon the court's own research," to issue a preliminary memorandum and invite the parties to be heard on the "'propriety of taking judicial notice and the tenor of the matter noticed' upon motion made within ten days." *Bulova*, 508 F. Supp. at 1328. Judge Weinstein suggested that "inviting parties to participate in such ongoing colloquy has the advantage of reducing the possibility of egregious errors by the court and increases the probability that the parties may believe they were fairly treated, even if some of them are dissatisfied with the result." *Id.* He noted that:

This procedure complies with the spirit of *Rule 201(e) of the Federal Rules of Evidence*. . . 'A party is entitled upon timely request to an opportunity to be heard as to the propriety of taking judicial notice and the tenor of the matter noticed. In the absence of prior notification, the request may be made after judicial notice has been taken.'

Id.

146. Weinstein, *supra* note 145, at 565.

include that judicial education programs be provided by neutral or balanced sources and that judges' participation in these programs be disclosed.

Even more importantly, judges themselves must scrutinize the extra-judicial social science material they are considering. The judicial gate-keeping function should be employed not just with experts, but with judicial notice as well.

Generally, courts should be hesitant to take judicial notice of social science research at the case-specific and intermediate levels, because social science findings typically should not be considered "indisputable," and further, there could be disagreement about the connection between the research and the facts of the case.

With our analyses of the way social science research is accessed and used by the courts as a context, we turn to a discussion of standards that judges may employ in assessing social science research. These standards are based in the scientific method and are consistent with *Daubert*.

IV. GUIDELINES FOR ASSESSING THE VALIDITY OF SOCIAL SCIENCE RESEARCH

In this part we describe the scientific method in general and its application to social science research on families. In the context of this analysis, we propose and discuss a set of standards that judges may employ when they must assess social science in deciding cases and developing law. These standards are needed because judges often have no expertise in evaluating science-based claims.¹⁴⁷ Because the standards are explicitly tied to family law, they help to bridge that knowledge gap.

As discussed in Part III, social science research findings may come to the courts from several sources: expert testimony, briefs, and judicial notice. In addition to different sources of social science, it is also important to identify two different types of social science claims that courts may consider.

In the first type of claim, which might be referred to as a *direct research claim*, an expert witness, a brief, a book, an article, or a speaker makes direct reference to findings from a particular study or body of social science literature. For example, in a child protection hearing, it might be claimed that research has shown that family preservation programs have beneficial impacts on children and family functioning. In assessing this claim, as we argue below, the judge should inquire into

147. See, e.g., Saks, *supra* note 6, at 230.

whether the scientific method was applied properly to the production of the family preservation research that is cited.¹⁴⁸

The second type of social science claim is less direct and may be referred to as a *science-based practice claim*. In this instance, an expert, who may or may not be a social scientist, proffers information to the court, usually about a specific party in the case, that is formed from the expert's use of an assessment or clinical strategy based on scientific literature. For example, a clinical social worker or a psychologist may have conducted a custody evaluation and in the course of the evaluation, used one or more assessment instruments for the child or parents. In doing so, there is a claim that the custody evaluation strategy is based on research, that the assessment instruments have been produced and tested scientifically and, therefore, that the expert's evaluation has scientific validity. In this instance, the judge should inquire as to whether the assessment strategies and instruments employed have in fact been produced scientifically.¹⁴⁹

The key question that motivates the analysis in this section is: When a court has information before it that claims social scientific status, what standards should the court bring to bear in assessing the validity of this claim? We propose that regardless of the form or nature of the claim, the court should evaluate it according to standards of the scientific method and come to conclusions in compliance with these standards.¹⁵⁰ Our proposal is in accord with *Daubert*, which calls upon judges to ask four questions of any scientific information proffered to the court by expert witnesses. These questions relate to the scientific method and the scientific community that employs this method.¹⁵¹ Paraphrased, the questions are as follows:

148. Note that this analysis should also be applied to claims that make no explicit reference to any studies, but have the imprimatur of "science," because they are proposed by someone who has expert credentials and wants the work accepted as "science." The book, *BEYOND THE BEST INTEREST OF THE CHILD*, GOLDSTEIN ET AL., *supra* note 9, would be an example of this type of claim. The authors, respected clinicians, did not indicate the basis for their claims, which presumably were formulated from their clinical practices. The book has been highly criticized both for its conclusions and for its lack of attention to scientific method. See Davis, *supra* note 9; see also, e.g., Daniel Katkin et al., *Above and Beyond the Best Interests of the Child: An Inquiry into the Relationship Between Social Science and Social Action*, 8 LAW & SOC. REV. 669 (1974); Peter L. Strauss & Joanna B. Strauss, Book Review, 74 COLUM. L. REV. 996 (1974) (reviewing JOSEPH GOLDSTEIN, ANNA FREUD, & ALBERT J. SOLNIT, *BEYOND THE BEST INTERESTS OF THE CHILD* (1973)).

149. See *infra* Part IV.B.1.

150. For example, Conley & Peterson, *supra* note 67, at 1186, argue that *Daubert* calls upon the court to determine whether "the work of the expert follows the cannons of the scientific method."

151. *Daubert* requires this analysis for expert testimony. We propose that the same analysis should be applied to other sources of social science as well. See *infra* Part V.

1. Can the scientific claim be empirically tested and has it been empirically tested?
2. What is the error rate of the methods or techniques used to test the claim, i.e., how valid are the methods employed in the study or research literature?
3. Has the scientific claim been subjected to peer review and publication?
4. Has the claim achieved general acceptance in the scientific community?¹⁵²

While our discussion in this section does not explicitly track the four *Daubert* questions, we do address each of them in the context of our broader discussion of the standards for assessing the validity of social science research.

At its core, the scientific method consists of three interacting systems: a system of *logic* for developing and empirically testing hypotheses and building theory; a system of *methodological standards* that scientific disciplines employ to guide the research practices used to test hypotheses; and a system of *community socialization, review, and communication practices* that train, police, and coordinate the work of diverse scientists in the day-to-day practice of science.¹⁵³ The scientific method, as it functions, may be conceived of as an ongoing and vast filtering system of claims to scientific knowledge with each of these three systems constituting a level of filtration.¹⁵⁴ Below, we describe each of these systems and how it operates in the social sciences with special reference to family and child research.¹⁵⁵ Based in this description and analysis, Table 2 sets forth validity questions linked to each level of filtration that jointly constitute the scientific method. It is our hope that the table will be a helpful tool for judges and other legal actors concerned with evaluating social science information proffered to the courts.

TABLE 2: QUESTIONS SPECIFIC TO THREE SUB-SYSTEMS OF THE SCIENTIFIC METHOD THAT JUDGES SHOULD ASK IN CONSIDERING FAMILY SOCIAL SCIENCE

SYSTEM I: THE LOGIC OF SCIENTIFIC INFERENCE

152. *Daubert v. Merrell Dow Pharm., Inc.*, 509 U.S. 579, 593-94 (1993).

153. See FOSTER & HUBER, *supra* note 12, for an analysis that also covers these three systems of the scientific method.

154. For a discussion of the scientific method as a set of filters, see Henry H. Bauer, *How Science Really Works*, in SCIENTIFIC LITERACY AND THE MYTH OF THE SCIENTIFIC METHOD (1992).

155. This section substantially updates and provides a more systematic elaboration of themes discussed in Ramsey & Kelly, *supra* note 10.

- Has the *logical system of scientific inference* been followed in the production of the information before the court?
- Is the information the result of empirical testing of falsifiable hypotheses?

SYSTEM II: METHODOLOGICAL DESIGN AND ANALYSIS STANDARDS

1. Measurement Validity

- Have the measures used in the research been validated in the study itself or in prior research?
- Have multiple, validated measures of key variables been used in the research?
- Have experts in measurement, e.g. psychometricians, been consulted where there is doubt concerning measurement validity?

2. Internal Validity and Causal Inference

- Have rigorous *designs* that ensure high levels of internal validity (the controlling of rival hypotheses) been employed in the research before the court?
- Has the research employed well-designed control groups?
- Have experts in research design been consulted where there is doubt concerning internal validity?

3. External Validity: Sampling, Generalizing, and Fit Between Research and Legal Applications

- Has a random sampling method been used in the research before the court or has a non-random sampling method, such as convenience sampling, been employed?
- If sampling has been employed, what is the response rate?
- If a longitudinal sample has been employed, what is the attrition rate?
- Has the magnitude of problems associated with low response rates and/or high attrition rates been established? Have efforts been made to correct for bias created by response rate or attrition rate problems?
- Is there a strong correspondence between the samples or populations used in the study before the court and the legal issues and case being considered by the court?
- Have experts in sample design and execution been consulted where there is doubt concerning sampling and generalizing?

4. Analytic Validity

- Is there an appropriate match between the data used in the research and the analytic methods applied to the data?
- Have the analytic techniques used in the research been applied appropriately?
- Have the results of the analysis used in the research been properly interpreted?
- Have reference manuals geared to judicial practice, such as the Federal Judicial Center's *Reference Manual on Scientific Evidence*, been consulted to address analytic validity questions?
- Have experts, such as applied statisticians, been consulted where there is doubt concerning analytic validity?

SYSTEM III: SOCIAL PRACTICE OF SCIENCE

1. Community Socialization

- To what degree do the social scientists who have authored the research in question possess professional records indicating training and continuing education consistent with the community socialization standards of their discipline?

2. Peer Review

- Has the research successfully passed peer review?
- What is the exact nature of the peer review process used to assess the research under consideration by the court?

3. Scientific Communication

- Has the research proffered to the court as scientific appeared in a reputable scientific publication?
- Have the results of research considered by the court been independently replicated?
- Is the data set upon which the research is based available to other researchers for re-analysis?
- Has the research been placed in the context of a synthetic review of research literatures on relevant issues?

A. *The Logic of Scientific Inference and Progress: Empirical Testing*

Science is in the business of developing general systems of causal statements about how the world around us operates.¹⁵⁶ A statement that outlines a suspected causal relationship between two measurable dimensions of the empirical world is called a *hypothesis*, and the dimensions themselves are called *variables*.¹⁵⁷ A suspected cause is called an *independent variable*, and a suspected effect is called a *dependent variable*.¹⁵⁸ A social scientific hypothesis concerning families and a legal structure, such as joint physical custody, for example, could be:

Children in joint physical custody arrangements subsequent to divorce experience fewer divorce-related negative impacts on their well-being than children in sole custody arrangements.

In this hypothesis, joint custody versus sole physical custody is the independent variable and the child's well-being is the dependent variable.

Systems of interrelated hypotheses that together seek to explain a substantial portion of empirical reality are called *theories*.¹⁵⁹ For example, in developmental psychology, a set of hypotheses jointly constitutes attachment theory, which seeks to explain the relationship between early parent-child interactions and the child's development of later-life

156. See, e.g., LACHMAN, *supra* note 12, at 85.

157. See RODNEY STARK & LYNNE ROBERTS, *CONTEMPORARY SOCIAL RESEARCH METHODS* 13 (2d ed. 1998).

158. *Id.* at 99.

159. For similar statements on the nature of scientific theory, see LACHMAN, *supra* note 12, at 49-56; STARK & ROBERTS, *supra* note 157, at 9-12.

competencies.¹⁶⁰

Thus far we have focused on definitions. Now, we turn to the underlying logic of science. Karl Popper, a leading twentieth-century epistemologist of science, whose thinking is widely cited by practicing scientists, argues that the business of science is not the production of a system of lasting truths, but rather the building of strong theories through the logical process of falsification.¹⁶¹ While other philosophers and historians of science, such as Thomas Kuhn, and other theories of science, such as social constructivism, have challenged Popper's views,¹⁶² Popper's influence in law has increased due to *Daubert's* reliance on the concept of falsification.¹⁶³

For Popper, falsification is the dynamo underlying science. The process works in the following way. First, a hypothesis is developed. Scientists may produce hypotheses from any number of sources. For example, hypotheses may be logical derivations of existing theories or they may be new in their derivation. What is key is not so much the source of the hypotheses, but rather what is done with the individual hypothesis. Second, for Popper and for virtually all scientists, after hypotheses are derived they must next be empirically tested through well-controlled research designs (what we mean by a well-controlled research design is the subject of the next subsection). No matter how compelling or elegantly derived the theory, if there is no empirical testing, there is no science. Abiding by this theory, the Supreme Court in *Daubert* was clearly following Popper in posing its first question: Has the scientific claim been empirically tested?

The next step in Popper's logic of science concerns the nature of the inferences that may be made based on the results of the empirical test(s) of the hypothesis. Recall that hypotheses are general claims about suspected causal relationships in the system of empirical reality that is the purview of the study. When the empirical results of a study testing a hypothesis are known, the scientist and the scientific commu-

160. Robert F. Kelly & Shawn Ward, *Allocating Custodial Responsibilities at Divorce: Social Science Research and the American Law Institute's Approximation Rule*, 40 FAM. CT. REV. 350 (2002).

161. KARL POPPER, *THE LOGIC OF SCIENTIFIC DISCOVERY* (1968). For a more contemporary statement of Popper's view of science and the dynamics of falsification that is geared to a legal audience, see Jan Beyea & Daniel Berger, *Scientific Misconceptions Among Daubert Gatekeepers: The Need for Reform of Expert Review Procedures*, 64 LAW & CONTEMP. PROBS. 327 (2001); see also, generally, FOSTER & HUBER, *supra* note 12, ch. 3, at 37-68.

162. THOMAS S. KUHN, *THE STRUCTURE OF SCIENTIFIC REVOLUTIONS* (1962); KARIN D. KNORR-CETINA, *THE MANUFACTURE OF KNOWLEDGE: AN ESSAY ON THE CONSTRUCTIVIST AND CONTEXTUAL NATURE OF SCIENCE* (1981). See also critics cited *supra* note 14.

163. STATE JUSTICE INSTITUTE, *A JUDGE'S DESKBOOK ON THE BASIC PHILOSOPHIES AND METHODS OF SCIENCE* 28 (1999).

nity want to know how the empirical results relate to the hypothesis that was tested. Popper argues that in a strict sense, empirical results can only tell us if the hypothesis was falsified; we can never know with certainty if the hypothesis is true in the same way that a mathematical proof is true. To understand why this is the case, it is important to consider the types of results that an empirical test might yield. If the results are negative, i.e. the direction of the relationship between the independent and dependent variables runs in the opposite direction of the result predicted by the hypothesis, or if the results show no relationship between the independent and dependent variables, then the hypothesis has been falsified. This is so because we have found an empirical case that negates the general assertion of the hypothesis.

But what if the results of the study testing the hypothesis are positive, that is, the relationship predicted by the hypothesis is indeed found to exist empirically? With such results Popper argues that we cannot say that the hypothesis has been verified. We can only say that it has not been falsified. This is the case because the hypothesis is a general statement, meant to apply to all relevant cases, but the empirical test concerned only the specific cases studied. Hence, it is logically indefensible to generalize from the particular to the general or universal. The most that we can assert based on a positive result from an empirical test of a hypothesis is that the hypothesis has been supported. We cannot say that the hypothesis is true.

This is an important insight because it runs counter to many popular conceptions of science, which suggest that there is such a thing as scientific truth. In reality, the scientific knowledge of any historical moment should be viewed as tentative and humble knowledge for two reasons: first, because scientific knowledge is based on tests that can only falsify and not verify, and second, because the scientific community is always in the process of conducting tests of new hypotheses that may falsify or supercede current theory. There is, however, an important distinction to be made. While it is true that science by definition cannot produce enduring truths, it is also true science has been the source of major advances in knowledge of the empirical world. Indeed, the ability of science to make progress in the production of knowledge is premised on the very system of logic that prevents it from the claiming of enduring truths.

Judges may draw two conclusions from this brief discussion of the logic of the scientific method. First, science requires the empirical testing of hypotheses and theories. If a theory or hypothesis is proffered as scientific and it cannot or has not been empirically tested, it cannot be considered to be scientific. Hence, judges constantly need to inquire as

to the empirical grounds upon which any scientific claim is based. The following are two illustrations of how this standard might operate.

- An expert witness or brief makes a *direct research claim* that foster care generally has negative consequences on the well-being of children. This assertion should reference an empirical research literature indicating that this relationship has been well-tested and supported.
- A clinical social worker makes the science-based practice claim in a custody proceeding that a child exhibits parental alienation syndrome (PAS). As a first step in determining admissibility, the judge should inquire whether studies in which PAS is an independent or dependent variable have been conducted and whether the results of the studies support the claim that PAS is a useful descriptive or diagnostic category.¹⁶⁴

Second, judges need to be aware of the tentative nature of scientific knowledge. Because the logic of falsification is at the heart of science, certainty with respect to the standing of a hypothesis or theory can never be complete. Furthermore, because the process of falsification is ongoing, judges need to be sensitive to the nature of scientific knowledge as a constantly evolving body of empirically based theory. As a result of this constant process of theory revision, there is relatively little “black letter” science.¹⁶⁵

Rigorous empirical testing is at the core of the logic of falsification. If an empirical test of a hypothesis is not rigorous, its results might be used either to incorrectly falsify or to incorrectly support a hypothesis. Hence, the scientific method consists not simply of the logic of falsification, but also of a set of methodological standards that guide the empirical testing of hypotheses. We next turn our attention to the methodological standards by which the empirical testing of hypotheses are judged by the scientific community, for it is with these standards that

164. Such studies might involve tests of models of hypothesized factors that predict the onset of PAS or models which test hypothesized impacts of PAS. For critical assessments of PAS, see the special issue of the Family Court Review dedicated to *Alienated Children in Divorce*, 39 FAM. CT. REV. 246 (2001) and Janet R. Johnston & Joan B. Kelly, *Rejoinder to Gardner's "Commentary on Kelly & Johnson's 'The Alienated Child: A Reformulation of Parental Alienation Syndrome,'"* 42 FAM. CT. REV. 622 (2004). See *infra* text accompanying notes 211-12. On the issue of judges accepting PAS diagnoses in custody proceedings addressed in the same special issue of Family Court Review, see especially R. James Williams, *Should Judges Close the Gate on PAS and PA?*, 39 FAM. CT. REV. 267 (2001).

165. Popper's logic of scientific inference is not relevant to the work practiced by all scientists. For example, some fields of science are highly descriptive and concerned more with improving measurement than with testing hypotheses. Hence, it is probably best to limit Popper's falsification logic to scientific statements that seek to test causal relationships that have been posed as hypotheses. In Part IV.B., we explicitly treat the question of scientific measurement and the assessment of its validity.

judges must also be conversant in their evaluations of social scientific information.

B. *Methodological Standards of Design and Analysis*

There are four standards that scientists generally seek to meet in carrying out their own empirical research and in judging the research of others. Problems related to any of the standards raise the question of what may be called a methods effect or bias in the research design and analysis. As we have noted, the logic of scientific inference is premised on the assumption that valid empirical tests of hypotheses have actually taken place.¹⁶⁶ Such testing involves two steps: the implementation of a research design to collect the necessary empirical data and the subsequent analysis of the data through which hypotheses are tested. When judges consider whether proffered information has been produced by the scientific method, they necessarily must inquire into the degree to which methodological standards of validity in research design and analysis were maintained.

It is with these standards that the Supreme Court was most likely concerned when it instructed judges to ask: What is the error rate of the methods used to test the claim, i.e., how valid are the methods employed in the study or research literature?¹⁶⁷ In this section, we discuss four crucial methodological standards in turn: *measurement validity*, *internal validity*, *external validity*, and *analytic validity*. While the design and analysis standards do not vary from discipline to discipline, the manner in which they are implemented does vary due to the differences in the nature of the empirical system that each scientific discipline studies. Because our concern is with social science research, particularly research related to families and children, we focus on the ways in which social scientists seek to conduct their research in a manner consistent with the four standards and on the challenges that they encounter in doing so. In this section we address the standards themselves, while in the next section, Part IV.C., we consider the social processes used by the scientific community to oversee the use of the standards.

1. MEASUREMENT VALIDITY

Because the scientific method requires that theoretical hypotheses

166. POPPER, *supra* note 161.

167. In *Daubert*, the U.S. Supreme Court noted that “the court ordinarily should consider the known or potential rate of error. . .and the existence and maintenance of standards controlling the technique’s operation.” *Daubert v. Merrell Dow Pharm., Inc.*, 509 U.S. 579, 594 (1993) (citations omitted). While the Court’s wording is somewhat ambiguous, we interpret and paraphrase this statement to mean that courts need to be concerned with the degree to which scientific standards have been maintained with respect to techniques or methods used to test hypotheses.

relating independent and dependent variables be tested empirically, it is necessary to design and implement strategies for measuring these and other variables. The development of empirical measures for theoretical variables is called *operationalization*.¹⁶⁸ Just as a physicist must clearly specify how to measure weight or pressure for the purpose of a study, so too would a social scientist studying the impacts of divorce on children's well-being need to carefully specify how various dimensions of children's well-being would be measured. The standard against which social scientists assess these strategies is called *measurement validity* and it simply refers to the ability of a given measurement strategy to measure what it intends to measure.¹⁶⁹ Measurement validity is a necessary condition for all scientific research because without it subsequent stages in the research process, such as data analysis and hypotheses testing, become suspect. Various methods have been developed for enhancing and assessing measurement validity. Below we focus on the logic behind these methods, rather than the technical aspects of specific methods.

Measuring human social systems is extremely difficult because human social systems are complex and changing realities. Further, ethical and practical concerns, as well as the willingness of subjects to participate in studies, often limit the level of access that social scientists have to subjects for purposes of measurement. Each of these difficulties is heightened further when the study seeks to measure family and child behaviors due to the sensitivity of such behaviors. How is a judge to know if social science research proffered to the court has employed valid measurement strategies? Below, we discuss two measurement issues with which courts should be concerned in making such assessments.

a. The Need for Empirically Validated Measurement Systems

As we have noted, two sorts of social science-based claims may be proffered to courts dealing with family and child matters: *direct research claims* and *science-based practice claims*. In each instance, under *Daubert*, the court would need to determine whether the measurement system used had been validated scientifically. A *direct research claim*, for example, might involve a researcher who measures the variables of *parental conflict* and *children's well-being* as part of a study of

168. Donald T. Campbell, *Definitional versus Multiple Operationism*, in *METHODOLOGY AND EPISTEMOLOGY FOR SOCIAL SCIENCE: SELECTED PAPERS* 31-36 (E. Samuel Overman ed., 1998).

169. For general discussions of measurement validity, see STARK & ROBERTS, *supra* note 157, at 46-49; MARY J. ALLEN & WENDY M. YEN, *INTRODUCTION TO MEASUREMENT THEORY* (Lawrence S. Wrightsman ed., 1979).

the consequences of divorce on children, and that researcher's study is later cited to a court in a disputed custody case. A *science-based practice claim* might involve a parental custodial fitness assessment by a forensic psychologist submitted to the court as part of a disputed custody proceeding that contains a recommendation based in part on results derived from instruments designed to assess parental custodial competence. In each of these examples, part of the court's assessment of the proffered information would necessarily include an assessment of whether the measurement system used in the production of either type of information is scientifically valid. Before addressing how the court would assess measurement validity in cases such as these, it is useful to describe how the social sciences generally assess measurement validity.

In the social sciences, and indeed in all sciences, establishing measurement validity is a collective, community-based process that unfolds over time. Measurement systems in the social sciences may take many forms. For example, where some measurement systems rely upon observational methods such as clinical assessments, others may call upon knowledgeable individuals such as teachers to describe and rate another individual. Still others rely on self-reports to interviewers or on questionnaires, while others may require subjects to perform tasks that are scored. When a new measure is fielded for the first time, measurement theory requires that a series of well-known validity tests be performed to determine the validity of the new measure.¹⁷⁰ The purpose of most tests of measurement validity is to determine (a) whether the new measure gives consistent results when applied to the same or similar subjects under the same conditions (reliability)¹⁷¹ and (b) whether the new measure is related in a predictable manner to other independently derived measures of the same theoretical variable, which is called a construct.¹⁷²

170. For a useful discussion of measurement validity focusing on the uses of psychological testing for forensic purposes, see David Medoff, *The Scientific Basis of Psychological Testing: Considerations Following Daubert, Kumho and Joiner*, 41 FAM. CT. REV. 199 (2003). For a general discussion of measurement theory, see ROBERT J. GREGORY, *PSYCHOLOGICAL TESTING: HISTORY, PRINCIPLES AND APPLICATIONS* (Rebecca Pascal ed., 3d ed. 2000).

171. Measurement reliability is sometimes discussed as a separate concept from measurement validity. We discuss it here as a type of measurement validity because a measure cannot be valid if it is not reliable. Measurement reliability is assessed in manners similar to overall measurement validity.

172. Standard approaches to assessing measurement validity that examine the nature of the measure's relationship to other independently derived measures include *convergent validity* and *discriminate validity*, each being types of *construct validity*, and *concurrent validity* and *predictive validity*, each being types of *criterion validity*. Detailed discussions of these methods are beyond the scope of this paper. Other approaches to measurement validity that do not rely on multiple, independent measures include *face validity* and *logical validity*, both of which are types of *content validity*. For more detailed discussions of measurement validity and reliability, see Medoff, *supra* note 170; ALLEN & YEN, *supra* note 169; EDWARD G. CARMINES & RICHARD A. ZELLER,

After the results of initial measurement validity tests are peer reviewed and published, other researchers typically seek to re-validate and refine the measure by replicating and extending the initial validation work. For important measures, this process continues in regular cycles, all of which are documented in published social scientific literature.

When evaluating a measure, social scientists view measurement validity as a continuum defined by poles of *high* and *low* validity. What determines the placement of a particular measure on the continuum is the degree to which it has been successfully validated and improved by multiple researchers over time. At one end of the continuum are measures that have never been formally validated, even by the researchers who first proposed them. Such measures have little or no validity and could not be accepted as scientific under *Daubert*. At the other end of the continuum are measures that have been validated and improved in multiple studies over a sustained period of time. These measures would have high measurement validity. It is notable that the majority of measures would be placed somewhere between the two poles of the measurement validity continuum. What is key for judges to understand is that the level of validity of a measure is a matter of degree determined by the amount of empirical validation work done on the measure and the degree to which this work has supported and improved the measure over time. Further in this section, we will provide illustrations of measures relevant to family law and discuss their levels of measurement validity.

What does the process of measurement validation imply for judges who must make judgments as to the scientific validity of the measurement systems upon which *direct research claims* or *science-based practice claims* are made? Judges will need to determine if the measures used in the research upon which the claims are based have been validated according to the process described above. This can be done in a number of ways. Perhaps the most obvious path for the judge to follow would be to require the individual who has proffered the claim to provide the court with a brief substantiating the degree to which the measures used have been validated. In this regard, it is notable that scientific publications periodically provide overviews of the current status of validity of many important measures.

The court might also seek the opinion of an independent, court-appointed expert to provide an assessment of the measurement system under scrutiny. At first glance, this sort of judicially sponsored review may appear an onerous and time consuming task for judges to undertake. However, there are only a limited number of measurement systems rele-

vant to the caseloads of a typical court, and an assessment, once performed, would not need to be repeated for a substantial period of time should the same measurement system come into question in subsequent cases.¹⁷³ We next provide brief examples of how the court might implement the requirement for well-validated measures for both *direct research* and *science-based practice claims*.

i. *Direct Research Claims*: One of the key issues studied in the field of family social science in recent decades is the impact of divorce on children's well-being. One of the significant dimensions of children's well-being addressed in much of this research is social adjustment, which might include such factors as fighting, stealing, lying, or inappropriate school behavior. If the research proffered to the court assessed the social adjustment of children, it would be important for the court to inquire as to how the social adjustment of the children in the research was measured with specific attention focused on the validity of the measures employed. In the best research on the impact of divorce on children's well-being, highly validated measures of social adjustment such as the Rutter Home Adjustment Scale, the Bristol Social Adjustment Guide, the Achenbach Child Behavior Checklist and the Behavioral Problem Index have been utilized.¹⁷⁴ Researchers, such as sociologist Andrew Cherlin, have used these scales in highly respected studies of the impact of divorce on children's well-being with justifiable confidence in their measurement validity.¹⁷⁵ Hence, a judge encountering such research could be confident that the research possessed measurement validity. On the other hand, if the research proffered to the court could not substantiate that its method for measuring children's social adjustment had been well-validated, the court could not accept the research as scientific.

ii. *Science-Based Practice Claims*: How might the requirement for validated measures be applied in the case of a custody evaluation that has employed a psychological or other assessment or testing system? Such an evaluation would be an illustration of what we have called a

173. Note also that some scholars have argued that admissibility decisions in some contexts should be treated as a matter of law. *See supra* note 97 and accompanying text.

174. EDUCATION, HEALTH AND BEHAVIOUR (Michael Rutter et al. eds., 1970); D.H. STOTT, THE SOCIAL ADJUSTMENT OF CHILDREN: MANUAL TO THE BRISTOL SOCIAL ADJUSTMENT GUIDES (3d ed. 1969); THOMAS M. ACHENBACH & CRAIG S. EDELBROCK, BEHAVIORAL PROBLEMS AND COMPETENCIES REPORTED BY PARENTS OF NORMAL AND DISTURBED CHILDREN AGED FOUR THROUGH SIXTEEN (46 Monographs of the Soc'y for Res. in Child Dev., No. 1, Serial No. 188, 1981); James L. Peterson & Nicholas Zill, *Marital Disruption, Parent-Child Relationships, and Behavior Problems in Children*, 48 J. MARRIAGE & FAM. 295 (1986).

175. *See, e.g.*, Andrew J. Cherlin et al., *Longitudinal Studies of Effects of Divorce on Children in Great Britain and the United States*, 252 SCIENCE 1386 (1991).

science-based practice claim because there is an implicit or explicit claim that the assessment strategy employed in the custody evaluation has been developed and validated scientifically. Hence, the judge should inquire as to whether a research literature that addresses the validity of the measurement strategy exists and, if it does, whether there is a consensus that the measurement system is valid. If no literature exists, a claim of scientific validity cannot be supported.

If a validation literature for the measure used in the custody evaluation did exist, then its rigor would need to be assessed and its conclusions summarized as a prerequisite for determining whether its claim to scientific validity would be accepted. This is precisely what has been done, for example, in the case of Minnesota Multiphasic Personality Inventory (MMPI-2), a personality measurement tool frequently used in custody evaluations.¹⁷⁶ Specifically, there is a large research literature that has validated the MMPI-2 and courts have accepted the MMPI-2 as a scientifically valid means of measuring personality.¹⁷⁷ Other psychological measures that are well-validated and that might credibly appear in a custody assessment include the Wechsler Adult Intelligence Scale, Third Edition (WAIS-III) and the Personality Assessment Inventory (PAI).¹⁷⁸

On the other hand, if a review of the existing literature revealed that a custody evaluation relied upon a measurement strategy that was not fully validated, that portion of the expert testimony should not be accepted as scientific. Several well-known psychological tests that commonly appear in expert testimony in family and probate courts have not been scientifically validated.¹⁷⁹ These tests include the Thematic Apperception Test, the Tasks of Emotional Development Test, and several versions of the Draw-a-Person Test and the Sentence Completion test.¹⁸⁰

The recently developed Ackerman-Schoendorf Scales for Parent Evaluation of Custody Tests (ASPECT) is an illustration of a measurement strategy that seeks to scientifically assess the relative fitness of parents for custody.¹⁸¹ While substantial and promising work has been done on ASPECT's validation, its scientific validity is still an issue of

176. See, e.g., *People v. Stoll*, 783 P.2d 698 (Cal. 1989).

177. See FORENSIC APPLICATIONS OF THE MMPI-2 (Yossef S. Ben-Porath et al., eds., 2 Applied Psychology: Individual, Social and Community Issues, 1995); David Medoff, *MMPI-2 Validity Scales in Child Custody Evaluations: Clinical versus Statistical Significance*, 17 BEHAV. SCI. & L. 409 (1999).

178. For relevant citations on the validation of these measures, see Medoff, *supra* note 170, at 208.

179. *Id.* at 206, 209.

180. *Id.*

181. MARC J. ACKERMAN & KATHLEEN SCHOENDORF, ASPECT: ACKERMAN-SCHOENDORF SCALES FOR PARENT EVALUATION OF CUSTODY (1992).

some debate.¹⁸² Hence, until a consensus develops as to the validity of ASPECT, the prudent course for a judge to take would be to deny testimony based on ASPECT.

b. The Need for Multiple and Independent Measures of Variables

Beyond the need for well-validated, individual measures, judges should be familiar with two additional propositions of measurement theory when they assess social scientific research claims. First, no single measure or indicator of a variable (e.g., children's well-being) will be fully adequate to empirically capture the entirety of the reality the variable seeks to measure. Second, all individual measures or indicators of concepts contain some amount of measurement error; hence measurement error or bias cannot be fully eliminated.¹⁸³ These propositions and the research that supports them result in the following principle of measurement theory: overall measurement validity is enhanced substantially with the use of multiple, independently derived measures of the variable to be measured.

The existence of multiple independent measures of a variable does not assure perfect measurement, but rather maximizes the likelihood that measurement error will be minimized. Further, when multiple measures of the same variable are employed and the results of a study are consistent across the various measures, overall confidence in the study's results is enhanced significantly. In the social sciences, these propositions were first demonstrated persuasively in the now-classic 1959 article by Donald Campbell and Donald Fiske entitled "Convergent and Discriminant Validation by the Multitrait-Multimethod Matrix."¹⁸⁴ In the analysis, the authors showed that both measurement validity and the overall validity of the conclusions from empirical studies are maximized when researchers employ multiple, well-validated measures of the key variables. Below, we provide illustrations of how judges might apply this principle when direct research claims and science-based practice claims are made.

182. For varying views on the measurement validity of ASPECT, see ACKERMAN & KANE, *supra* note 90; Gary B. Melton, *Review of the Ackerman-Schoendorf Scales for Parent Evaluation of Custody*, in THE TWELFTH MENTAL MEASUREMENTS YEARBOOK 222 (Jane Close Conoley & James C. Impara eds., 1995). Melton is particularly critical of the psychometric properties of ASPECT, that is, its demonstrated measurement reliability and validity.

183. In this discussion our comments are geared mainly to the measurement of complex variables, such as the well-being of children. While in theory the points we make would also apply to more straightforward variables such as age and gender, these variables are generally not subjected to high levels of scrutiny with respect to measurement validity.

184. Donald Campbell & Donald Fiske, *Convergent and Discriminant Validation by the Multitrait-Multimethod Matrix*, 56 PSYCHOL. BULL. 81 (1959).

i. *Direct Research Claims:* To illustrate the application of this principle to the case of direct research claims, consider again a child protection or divorce case in which the court is presented with direct research claims that involve the measurement of children's well-being. The court would want to inquire not only about the degree of validity of individual measures of the children's well-being (see above), but also about whether the study employed multiple, well-validated measures of the well-being of children. A measurement strategy that uses multiple validated measures is more convincing than one that relies on a single measure, even if that single measure is well-validated. Research has shown, for example, that reports of children's well-being may vary significantly depending on whether the sources of the information are parents, teachers, the child, clinicians or other trained observers.¹⁸⁵ It is not so much that these individuals err in their reports, as it is that each has important information to contribute to the overall picture of a child's well-being. Hence, a study that relied solely on a parent's report on a child's well-being would be considered weaker than a study that also had gathered information from other sources, including the child herself and other trained observers.

A significant body of research on the impact of divorce on children that uses multiple measures has been conducted using the National Longitudinal Survey of Youth (NLSY), a large nationally representative sample of youth between the ages of fourteen and twenty-two that began in 1979.¹⁸⁶ Since 1986, the "youth" (now adult women) in the study and their children have been followed intensely with a broad array of measures gathered from a variety of sources designed to monitor the children's well-being. This component of the study is called the Youth Supplement (YS). In the NLSY-YS, extensive and repeated interviews and assessments are conducted with the mothers of NLSY-YS children concerning the child's pre- and post-natal health history. Mothers also provided information for the Behavioral Problem Index (BPI) and for a variety of other child-related attitudes and behaviors. In addition to information from mothers, the NLSY-YS collected information from the children themselves on a range of well-validated cognitive development scales such as the Peabody Assessments¹⁸⁷ and the Wechsler Intelli-

185. Thomas M. Achenbach et al., *Child/Adolescent Behavioral and Emotional Problems: Implications of Cross-Informant Correlations for Situational Specificity*, 101 PSYCH. BULL. 213 (1987).

186. FRANK L. MOTT, THE NLSY CHILDREN 1992: DESCRIPTION AND EVALUATION (rev. ed. 1998); PAULA C. BAKER ET AL., NLSY CHILD HANDBOOK (rev. ed. 1993); OHIO STATE UNIV. CTR. FOR HUMAN RES. RESEARCH, BUREAU OF LABOR STATISTICS, U.S. DEP'T OF LABOR, NLSY79: CHILD & YOUNG ADULT DATA USERS GUIDE (2002), available at <http://www.bls.gov/nlsy79cyaguide/nlsy79cusg.htm>.

187. LLOYD M. DUNN & FREDERICK C. MARKWARDT, JR., PEABODY INDIVIDUAL

gence Scales.¹⁸⁸ Further, trained NLSY-YS observers conducted the Home Observation for Measurement of the Environment (HOME) assessment, a well-validated measure of the quality of home environment.¹⁸⁹ Sociologists Donna Morrison and Andrew Cherlin have used the full range of NLSY-YS measures in one of the most definitive studies yet published of the effects of divorce on children.¹⁹⁰ One of the reasons that this study would deserve to be considered *scientific* by a court derives precisely from the fact that it integrates multiple sources of well-validated measures into a comprehensive analysis.¹⁹¹

Much of the same may be said of law professor Michael Wald's highly regarded field study of proposed innovations in the California child protection system. In order to assess the impact of these innovations, Wald and his colleagues needed to measure children's well-being. To do so, they collected information from a variety of sources using various measurement strategies, including health screening by medical professionals; administration of the Wechler IQ test to the children; assessment of academic achievement based on school records and teacher interviews; reports from mothers and teachers on a variety of child well-being characteristics using validated measures; Child Protective Services records; and interviews with the children themselves using validated measurement systems.¹⁹²

In summary, carefully designed studies, such as Morrison's, Cherlin's, and Wald's, that collect information from multiple sources using multiple, validated measurement strategies meet the scientific standard of measurement validity and should be treated as such by the courts.

ii. *Science-Based Practice Claim*: To illustrate the need for multiple, independently derived measurement strategies in the case of science-based practice claims, consider the issue of assessing parental

ACHIEVEMENT TEST MANUAL (1970); FREDERICK C. MARKWARDT, JR., PEABODY INDIVIDUAL ACHIEVEMENT TEST- REVISED (1989).

188. See *supra* text accompanying note 178.

189. See Robert H. Bradley, *The HOME Inventory: Review and Reflections*, in 25 ADVANCES IN CHILD DEVELOPMENT AND BEHAVIOR 241 (Hayne W. Reese ed., 1994); Robert H. Bradley & Bettye M. Caldwell, *Using the HOME Inventory to Assess the Family Environment*, 14 PEDIATRIC NURSING 97 (1988).

190. Donna Ruane Morrison & Andrew J. Cherlin, *The Divorce Process and Young Children's Well-Being: A Prospective Analysis*, 57 J. MARRIAGE & FAM. 800 (1995).

191. An additional strength of the NLSY-Children data study is that it is longitudinal, in other words, it follows children over time and, as a result, it is able to measure developmental processes inherent in childhood and adolescence. For another divorce impact study that uses multiple sources of information on children's well-being, see P. Lindsay Chase-Lansdale et al., *The Long-Term Effects of Parental Divorce on the Mental Health of Young Adults: A Developmental Perspective*, 66 CHILD DEV. 1614 (1995).

192. MICHAEL S. WALD ET AL., PROTECTING ABUSED AND NEGLECTED CHILDREN (1988).

capacity in the context of child protection proceedings deciding questions such as removal, visitation, reunification and termination of parental rights. In this area, the best-practices models for forensic custody assessments coincide with the scientific standard encouraging the use of multiple, independently derived measurement strategies. Medoff has noted that in the areas of child protection and child custody forensic assessments, a model of professional practices has emerged which calls for the assessor to gather information from multiple sources using multiple assessment methods and to integrate the results of these assessments into a final report.¹⁹³ For example, such reports would optimally be based on multiple sessions in which the evaluating mental health professional would interview parents across a broad range of topics, parent-child observations, administration of relevant tests/inventories for the parent, administration of relevant tests/inventories for the child, review of child protective services' previous reports, and interviewing of collateral sources such as teachers, social workers, family friends, extended kin, and physicians.¹⁹⁴ This approach is entirely consistent with the multi-source/multi-method measurement standard we have discussed, as well as the professional standards of practice developed by organizations such as the American Psychological Association.¹⁹⁵

Unfortunately, this model of professional practice is often at variance with the actual practice of parental capacity assessments in child protection matters, which typically rely on a single session with the parent(s), no in-home visit, few sources of information other than those provided by the parent(s), no reference to previous written reports, and little use of well-validated assessments methods.¹⁹⁶ When judges are faced with such incomplete parental capacity assessments, they should know that the assessments not only fall short of meeting professional standards, but also do not meet the standards of scientific measurement.

The key point to be taken from this section is that judges should inquire not only as to the validity of individual social science research measures, but also as to whether the research proffered to the court has employed multiple, independent measures of key variables. When evidence exists that a multi-method measurement approach has been

193. Medoff, *supra* note 170, at 210.

194. See Karen S. Budd, *Assessing Parenting Capacity in Child Protection Cases: A Clinical Practice Model*, 4 CLINICAL CHILD & FAM. PSYCH. REV. 1 (2001).

195. GUIDELINES FOR PROFESSIONAL EVALUATIONS IN CHILD PROTECTION MATTERS (Am. Psychological Ass'n Comm. on Practices & Standards 1998), available at <http://www.apa.org/practice/childprotection.html>.

196. Karen S. Budd et al., *Clinical Assessment of Parents in Child Protection Cases: An Empirical Analysis*, 25 LAW & HUM. BEHAV. 93 (2001).

employed and that results are consistent across measures, such results should be afforded greater weight by the court.

2. INTERNAL VALIDITY AND THE PROBLEM OF CAUSAL INFERENCE

Based on the prior discussion, let us assume that a researcher has a set of highly valid measures for the key variables needed to test the hypothesis that X causes Y, symbolically $X \rightarrow Y$. The next fundamental methodological issue is to design a study that will allow the researcher to determine with a relatively high degree of certainty whether X has caused Y, for example, whether joint physical custody arrangements cause children to be better off than sole physical custody arrangements.¹⁹⁷ Determining whether causal relationships exist among variables in human social systems in general and in family systems in particular is very difficult for at least two important reasons.

First, to reiterate a point made in the prior section, the social world is a very complex place. Typically, events such as marriage, birth, foster care placement, and divorce come about not for a single reason, but for many reasons acting either independently or in combinations called interaction effects. Hence, social scientists acknowledge that the world they study is a world of many causes, a multivariate world, where it is difficult to parcel out the effects of individual causes acting independently or in combination with other factors.

A second problem faced by social scientists seeking to make causal inferences has to do with the distinction between *correlation* and *causation*.¹⁹⁸ Just because two variables co-vary with each other, in other words they are correlated or associated, does not mean that a causal relationship exists between them. A correlation between two variables may be the result of an underlying causal relationship, but it also may be spurious. A spurious relationship is an artifact produced by the relationships of the two variables under consideration to some third variable whose effect has not been taken into account.¹⁹⁹ Returning to our joint custody example, assume that a study shows a positive relationship between joint physical custody and various measures of post-divorce

197. For a recent review of research on this question, see Kelly & Ward, *supra* note 160, at 350-70.

198. For a discussion of the issue of causation versus association in the context of the *Daubert* decision, see Conley & Peterson, *supra* note 67, at 1218, 1220.

199. For a humorous example of not taking a variable into account, see Eric J. Weissberger, *Biology of the Family Chiaceae (Chia Pets)*, 4 THE ANNALS OF IMPROBABLE RESEARCH (1998) (noting that "the Chia's mating call is usually heard in early November. . .[a]fter being born, juvenile Chias begin life in a purely animal phase, inside a plain cardboard cocoon. . .[i]n an amazing display of synchrony, all Chias hatch on the morning of December 25, and migrate to windowsills, where the plant phase begins to grow"), available at <http://www.improb.com/airchives/paperair/volume4/v4i4/chias.htm>.

children's well-being, that is, in joint physical custody cases scores on measures of children's well-being are significantly higher than in cases of sole maternal or sole paternal physical custody. Such an effect may be the result of the hypothesized benefits of joint physical custody, but it also may be because parents who select joint physical custody have higher incomes or more amicable divorces than parents who settle upon sole physical custody arrangements. Unless a research design can control for or neutralize the effects of income and the level of conflict in the divorce process, there will be uncertainty about the causal nature of the relationship between joint physical custody and children's well-being.

To rigorously determine whether a hypothesized causal relationship exists, a research design must be good at (a) controlling for possible spurious relationships and (b) determining the size and direction of independent and combined effects of individual variables in a multivariate world. Research designs that possess these characteristics are said to have high internal validity, that is, they yield robust answers to questions about hypothesized causal relationships.²⁰⁰ Figure 1 graphically represents the issue that is central to internal validity, namely that we can only know the nature of the true relationship between X_1 and Y if we have controlled for the possible confounding effects of the other variables ($X_{2, 3, 4}$) in the model.²⁰¹ When a judge encounters social science research on families that makes causal assertions, the judge should assess the internal validity of the research design upon which such claims are based. In the following subparts *a* and *b*, we briefly discuss common research designs that a judge may encounter and their strengths and weaknesses in relation to internal validity.

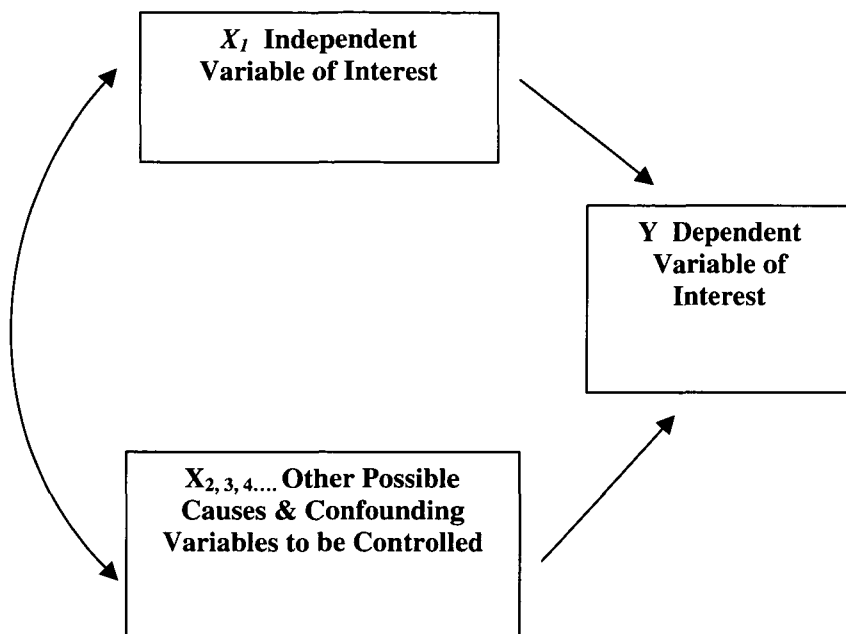
a. The Randomized Experiment

In the best of all worlds, researchers wishing to determine whether a causal relationship exists between two variables will use a *randomized experimental design*. For purposes of illustration, assume that a judge is hearing a child protection case and an expert witness is making the direct research claim that family preservation programs (FPP) have been shown to be effective in treating families where the risk is high that

200. The classic articulation of internal validity in the context of social science research is DONALD T. CAMPBELL & JULIAN C. STANLEY, *EXPERIMENTAL AND QUASI-EXPERIMENTAL DESIGNS FOR RESEARCH* (1963). For a valuable discussion of experimental research designs for public policy analysis, see John Gilbert et al., *Assessing Social Innovations: An Empirical Base for Policy*, in *STATISTICS AND PUBLIC POLICY* 185 (William B. Fairley & Frederick Mosteller eds., 1977). For an excellent discussion on internal validity in the context of evaluating innovations in legal settings, see *EXPERIMENTATION IN THE LAW: REPORT OF THE FEDERAL JUDICIAL CENTER ADVISORY COMMITTEE ON EXPERIMENTATION IN THE LAW* (Fed. Judicial Ctr. 1981).

201. The colloquial expression, *all things being equal* (*ceteris paribus*), captures well the underlying meaning of the concept of internal validity.

FIGURE 1: INTERNAL VALIDITY MODEL: TESTING FOR THE $X_1 \rightarrow Y$ HYPOTHESIS WHILE CONTROLLING FOR RIVAL HYPOTHESES ($X_{2, 3, 4, \dots}$)



children will be removed and placed in foster care because of concerns for the children's safety. The goals of FPP programs are to avoid removal by safely maintaining children in the homes of their parents and to improve the well-being of the children by providing an array of short-term, intensive services.²⁰² A judge faced with such a claim would want to ascertain whether scientific research supported this claim. Generally, if the available research upon which the expert witness based the claim that FPP programs are effective had employed randomized experimental research designs (and it met other validity standards discussed in this section), the judge could be relatively secure in his or her assessment of the internal validity of this direct research claim.

To understand why this is the case, it is important to comprehend the logic of experimental design as it would apply in the case of FPP programs. In principle, researchers studying the effectiveness of FPP programs would want to test hypotheses such as:

202. For a description of these programs and a literature review of research in the U.S. on their impacts, see Robert F. Kelly, *Family Preservation and Reunification Programs in Child Protection Cases: Effectiveness, Best Practices, and Implications for Legal Representation, Judicial Practice, and Public Policy*, 34 FAM. L.Q. 359 (2000).

- Children who participate in family preservation programs will be more likely to safely remain in residence with their parents or guardians than non-participating children receiving status quo child protection services who remain in residence with their parents or guardians.
- Children who participate in family preservation programs will demonstrate higher scores on measures of well-being than non-participating children receiving status quo services.

A randomized experimental design would call for the researchers to create two groups for the purpose of comparison: an experimental group of families that participates in the program and a control group that receives only status quo services.²⁰³ The two groups would then be followed over time and information would be collected on program outcome measures such as removals from home, health, school performance, and family functioning.

These program outcome measures are the dependent variables (Y) in the hypotheses. By creating the two groups, the researcher's design allows for comparisons to be made between the treatment group and the control group to measure the impact of the independent variable (FPP program participation vs. status quo (X_1)) on these dependent variables.

However, for a true experimental design to exist, it is not sufficient to simply create an experimental group and a control group. The experimental and control groups must be created by a random function such as tossing a fair coin. This process is referred to either as randomization or random assignment and it is a necessary condition in creating a classic experimental design.

The importance of random assignment for creating the experimental and control groups derives from the following logic. Assume that we do not know if the experimental and control groups were created through random assignment. Further, assume that the researchers have collected outcome information from the experimental and control groups over the course of two years and that analysis reveals that the experimental group has done much better than the control group. For example, analysis might reveal that many more children in the experimental group have remained safely at home than in the control group. The researchers naturally would be tempted to conclude that the family preservation program had been effective. Given these results, a methodological skeptic immediately would wish to know if any factor or variable besides the experimental treatment could account for the results. The skeptic's question is the *internal validity question*, namely how well has the

203. There are experimental designs that use more than two groups. We rely on the simple, two-group design for purposes of illustration.

research design controlled for rival plausible hypotheses ($X_{2, 3, 4, \dots}$ in the internal validity model) or, alternatively, are the two groups equal in all respects other than the experimental treatment?

Scientists have concluded that the best way to make experimental and control groups equivalent on all characteristics other than receiving or not receiving the experimental treatment is to use random assignment, that is, to allow chance and chance alone to allocate subjects to experimental or control groups.²⁰⁴ Hence, if our hypothetical researchers used randomization to create the two groups, they can respond to the skeptic that yes, within a statistically-known level of certainty, the treatment was the source of improvements in the experimental group.

To gain an appreciation of the advantages of randomized experimental designs, it is helpful to reflect on a hypothetical design that does not employ randomization to create experimental and control groups and the limitations that result with respect to internal validity and causal inference. Again, consider a researcher seeking to evaluate a family preservation program. Such a researcher might be tempted for a variety of reasons, including expense and effort, to create an experimental group by calling for volunteers. The researcher would subsequently administer the new program to the volunteers and then compare the volunteer group with a status quo group of families who had not volunteered for the program.

The problem with such a design from the perspective of internal validity is that volunteers for such a program may differ systematically from non-volunteers on characteristics that are related to outcome measures ($X_{2, 3, 4, \dots}$ in the internal validity model). For example, if volunteers have fewer problems to begin with or if they have a more cooperative attitude toward service providers and programs, one would expect them to do better on outcome measures independent of FPP program participation. As a result, we would be uncertain if the program caused improvements or if the volunteer-related factors caused the improvements. Consequently, we would be on weak ground in drawing any causal inferences related to program-effect hypotheses. Note, however, that if the researcher had used a randomized experimental design, there would be a high level of certainty that prior to the initiation of the experiment, the average number of problems faced by families and the average level of cooperativeness in the experimental and control groups would be about the same and, as such, not a concern from the perspective of internal validity.²⁰⁵

Judges may conclude from this discussion that a study presented to

204. CAMPBELL & STANLEY, *supra* note 200.

205. Kelly, *supra* note 202, limited his review of FPP program evaluations to mainly

the court that has employed a randomized experimental design offers the highest level of internal validity, which is the ability of a design to determine if a causal relationship exists. Such research, if otherwise competently executed, should be given greater scientific credence by the courts than research that does not use a randomized experimental design.

b. Alternatives to the Randomized Experiment:
Quasi-Experimentation

One of the problems with carrying out randomized experimental research designs is that for ethical, political, and practical reasons they cannot always be employed with human populations. For example, randomly assigning some children to joint physical custody, others to paternal custody, and others to maternal custody for a study of the impacts of custody arrangements on children would meet the requirements for a rigorous experimental design, but ethical and legal objections quite rightly would prevent such a design from ever being carried out. Given these limits, social scientists and policy researchers often employ alternatives to pure experimental designs that do not involve random assignment, but still retain many of the internal validity advantages of randomized designs. Such alternatives generally are referred to as quasi-experimental designs.

One such alternative, called a matched-group design, involves first identifying research subjects who agree to be exposed to an innovation or who have been exposed to the innovation naturally and using them as a treatment group. Next, this quasi-treatment group is matched with a quasi-control group consisting of subjects who have not been exposed to the innovation and who are matched to the quasi-treatment group. In principle, subjects in the treatment and control groups should be matched on all important characteristics that are thought to influence the relationship between the innovation under investigation and outcome measures. Hence, in matching, comparison groups are produced based on the researcher's knowledge of the population and the processes being investigated, rather than the more scientifically rigorous procedure of random assignment. Matching creates a valuable comparison group with which to assess a scientific hypothesis, legal reform, or innovative program so long as great care is taken in matching the experimental and control groups. Thus, while the internal validity of a matched group design would not be as strong as that of a randomized experimental design, it could be nearly as strong and it will clearly be superior to a design that did not employ a control group. Matched-group designs

experimental studies and found that some FPP programs with specific characteristics had moderate to small positive impacts, but that many programs had no effect.

have been used effectively to study legal innovations such as child support reform.²⁰⁶

Judges assessing matched-group studies may conclude that if a matched-group design has been well-executed, that is, matching has in fact occurred on all or most of the background characteristics known to be important, then credence should be given to the study's results, but not the same high level of credence as would be afforded results from a randomized experiment.

In addition to experimental and matched-group designs, a third research design often used in family social science for studying causal relationships should be considered, namely the survey. Surveys typically involve two major steps: the sampling of a population of interest, ideally using a probability sample (see the following section), and the administration of some type of questionnaire or interview to collect information from the members of the sample.

In family research, social scientists increasingly conduct longitudinal surveys in which the same respondents are repeatedly queried over time. Examples of well-designed longitudinal surveys of the population of United States families, children and youth include the National Survey of Children, the National Survey of Families and Households, and the National Longitudinal Survey of Youth (NLSY).²⁰⁷ Fortunately, it is feasible to use survey data, especially if they are longitudinal, to study causal hypotheses relevant to family law issues.

Recall that the key to making any causal inference is the development of a control group that is equivalent to the group that experiences the treatment or event that is being assessed. Assume that once again we wish to assess the potential impact of joint legal custody arrangements relative to sole legal custody arrangements, this time on the quality of post-divorce co-parenting. With a longitudinal survey of divorce cases, we can construct a control group of divorced families that has not exper-

206. Irwin Garfinkel et al., *The Wisconsin Child Support Assurance System: Estimated Effects on Poverty, Labor Supply, Caseloads, and Costs*, 25 J. HUM. RESOURCES 1 (1990); Irwin Garfinkel & Marieka M. Klawitter, *The Effect of Routine Income Withholding of Child Support Collections*, 9 J. POL'Y ANALYSIS & MGMT. 155 (1990); Irwin Garfinkel, *Utilization and Effects of Immediate Income Withholding and the Percentage-of-Income Standard: An Interim Report on the Child Support Assurance Demonstration*, 4 INST. FOR RES. ON POVERTY, SPECIAL REP. NO. 42 (1986).

207. For general descriptions of these data sets, see respectively NICHOLAS ZILL ET AL., 1976-1987 NATIONAL SURVEY OF CHILDREN: WAVES 1, 2, AND 3 (CHILD TRENDS, INC.) (1992); JAMES SWEET ET AL., UNIVERSITY OF WISCONSIN-MADISON, THE DESIGN AND CONTENT OF THE NATIONAL SURVEY OF FAMILIES AND HOUSEHOLDS (NSFH Working Paper No.1, 1988), available at <http://www.ssc.wisc.edu/nsfh/home.htm>; NLSY79: CHILD & YOUNG ADULT DATA USERS GUIDE, *supra* note 186. For the Canadian equivalent of the NLSY, see Statistics Canada's National Longitudinal Survey of Children and Youth (2003), available at <http://www.statcan.ca/english/sadds/4450.htm>.

enced joint legal custody and a treatment group that has experienced joint legal custody. But how comparable are these groups? How much internal validity would conclusions concerning the impact of joint legal custody on children possess if they are based on comparisons between these groups?

To answer these questions, it is important to recognize that statistical techniques can be employed with survey data to control for, that is, make equivalent, background differences between groups we wish to compare; differences that might confound our ability to make the sorts of comparisons needed for strong causal inferences.²⁰⁸ This is precisely the approach taken by psychologist Eleanor Maccoby and law professor Robert Mnookin in the Stanford Child Custody Study, a highly regarded analysis of post-divorce custody arrangements and their impacts in two highly diverse California counties.²⁰⁹

Maccoby and Mnookin used a longitudinal survey design in which court caseloads of divorces were sampled soon after initial filings and the divorcing parents were interviewed on three occasions during the three-year period following the divorce. Among the issues that Maccoby and Mnookin explored in these analyses was the effect of joint legal custody arrangements on post-divorce parenting, visitation, and child support. In making comparisons between joint legal custody and other custodial arrangements, the authors developed statistical models that controlled for potentially confounding background factors such as pre-divorce parental income, education and conflict ($X_{2,3,4...}$ in Figure 1). For example, in controlling for pre-divorce family income, the authors were able to account for the possibility that higher income families may be both more likely to choose joint legal custody and more likely to experience or report better post-divorce adjustment. In fact, Maccoby and Mnookin found that joint legal custody had little effect on post-divorce parenting, visitation, and child support. These findings are credible in terms of internal validity because of the well-developed statistical

208. For a general discussion geared to legal audiences of the *statistical regression techniques* used to control for confounding factors, see Daniel L. Rubinfeld, *Reference Guide on Multiple Regression*, in REFERENCE MANUAL ON SCIENTIFIC EVIDENCE 179 (2d ed. Fed. Judicial Ctr. 2000). For a general discussion of survey research geared to legal audiences, see Shari Seidman Diamond, *Reference Guide on Survey Research*, in REFERENCE MANUAL ON SCIENTIFIC EVIDENCE 229 (2d ed. Fed. Judicial Ctr. 2000). For discussions of the value of longitudinal data in assessing causal relationships, see JAMES S. COLEMAN, *LONGITUDINAL DATA ANALYSIS* (1981); Jay D. Teachman, *Methodological Issues in the Analysis of Family Formation and Dissolution*, 44 J. MARRIAGE & FAM. 1037 (1982); and PAUL D. ALLISON, *EVENT HISTORY ANALYSIS: REGRESSION FOR LONGITUDINAL EVENT DATA* (1984).

209. ELEANOR E. MACCOBY & ROBERT H. MNOOKIN, *DIVIDING THE CHILD: SOCIAL AND LEGAL DILEMMAS OF CUSTODY* (1992); Robert H. Mnookin et al., *Private Ordering Revisited: What Custodial Arrangements Are Parents Negotiating?*, in *DIVORCE REFORM AT THE CROSSROADS* 37 (Stephen D. Sugarman & Herma Hill Kay eds., 1990).

models used to control for differences between divorced families with joint legal custody and those with other custody arrangements.

From the perspective of judges evaluating such evidence, it is important to recognize that statistical controls of the type used by MacCoby and Mnookin would not be considered as robust as the controls produced by well-designed randomized or quasi-experimental matched-group designs, but comparisons derived from powerful statistical models do provide reasonably strong grounds upon which to draw causal inferences.²¹⁰

Several general implications may be drawn from this discussion of internal validity. It is essential that judges critically examine the grounds upon which researchers and practitioners make causal assertions. In doing so, judges will need to inquire as to the measures taken to control for rival causal hypotheses as plausible explanations for study results. Such an inquiry will inevitably lead judges to ask questions about control groups and other design methods used to control factors that may confound the researchers' abilities to make strong causal inferences. A crucial implication of this analysis is that judges should be particularly skeptical of small one-group designs with no control group that do not address systematically the problem of plausible alternative explanations of study findings.²¹¹

As an illustration of the difficulties associated with one-group designs, consider the fact that small, mainly one-group clinical studies were used to identify the parent alienation syndrome (PAS).²¹² Because most of these studies did not employ control groups or other rigorous control mechanisms, their ability to identify the causes and effects of the syndrome has been extremely limited. The same may be said with respect to evaluating the impact of treatment modalities geared to PAS.²¹³ Like PAS research, early research on family preservation pro-

210. The use of survey research in the social sciences to do causal analysis is very similar to the epidemiological techniques used in biomedical science. See, e.g., Michael D. Green et al., *Reference Guide on Epidemiology*, in REFERENCE MANUAL ON SCIENTIFIC EVIDENCE 333 (2d ed. Fed. Judicial Ctr. 2000).

211. Hypothetical illustrations of one-group designs might include: a child psychologist who employs a new therapy for treating teen depression in four of his/her adolescent clients with similar symptoms and who finds that there is improvement in each of the teens; or a single school district that alters its approach to providing support services to children experiencing divorce and finds that the students' test scores improve. The problem with one-group designs is that they do not allow for comparisons with subjects who were not exposed to the treatment or intervention under consideration. Hence, there is no way to tell whether the improvements found are the result of the treatment or factors unrelated to the treatment. Among research designs, the one-group design is the weakest in terms of internal validity. See CAMPBELL & STANLEY, *supra* note 200.

212. RICHARD A. GARDNER, *THE PARENTAL ALIENATION SYNDROME: A GUIDE FOR MENTAL HEALTH AND LEGAL PROFESSIONALS* (1992).

213. See Kathleen Coulborn Faller, *The Parental Alienation Syndrome: What Is It and What*

grams was largely based on small, one-group studies, which generally reported very optimistic results. Later research employing more controlled designs reported more mixed and nuanced results.²¹⁴

Hence, to reiterate, all social science research designs are not equal in their ability to control for alternative explanations and rival hypotheses. The proper use of social science findings requires disciplined inquiry concerning questions of internal validity and causal inference.

3. EXTERNAL VALIDITY

Assume for the purposes of our analysis that a family social science study is presented to a court and that its measurement validity and internal validity are robust. The court should be concerned with a third type of design validity, external validity. External validity refers to the degree to which a study's results may be generalized properly to relevant populations and conditions beyond the actual subjects of the study.²¹⁵

In the social sciences the issues of external validity and generalizability frequently lead to concerns about sampling. Because of cost and other considerations, social scientists typically study sample members of populations rather than studying entire populations. Hence, unless a researcher is studying an entire population through census methods, a major challenge in family research is developing representative samples from which to generalize to the population of interest. This is a challenge in all sample-based research whether the sample is a survey, a matched-group study, an ethnography, an experimental design, or any other type of data collection design.

a. Sampling and Generalizing

The optimal method for achieving a representative sample is to use a random function to select members for the sample from a universal listing of all members of the population that the sample is expected to represent. A sample selected in this manner is called a *simple random sample*. There are many technical variations of simple random samples (generally referred to as probability samples), but the key notion for the assessor of family social scientific research to keep in mind is that non-probability samples are likely biased, and if they are biased, they will present a distorted picture of the population. In contrast, well-designed

Data Support It?, 3 CHILD MALTREATMENT 100 (1998); Joan B. Kelly & Janet R. Johnston, *The Alienated Child: A Reformulation of Parental Alienation Syndrome*, 39 FAM. CT. REV. 249 (2001); see also sources cited *infra* note 164.

214. See Kelly, *supra* note 202.

215. See Medoff, *supra* note 170, at 201; CAMPBELL & STANLEY, *supra* note 200, at 5.

and well-implemented probability samples provide the best means available for minimizing the risk of bias, and for this reason they possess high levels of external validity or generalizability. Further, statistical measures of significance and levels of confidence may be properly employed only when a study utilizes a probability sample.²¹⁶

Hence, the first question that a judge should ask in attempting to assess the external validity of any sample-based family study is: was a random process used to select the sample? If it was not, then the judge should have serious concerns about the study's representativeness and any use of statistical analysis in the study would be substantially limited.

For a variety of reasons, many of which are related to time and cost, some researchers rely on small, non-random samples in their work. For example, in their much cited work on the impact of divorce on children, Judith Wallerstein and Joan Kelly used a sample of families with children experiencing divorce who presented themselves to a family center for divorce-related counseling and who subsequently agreed to participate in the study.²¹⁷ This type of sample is often referred to as a *convenience sample* because it capitalizes on social situations where groups of interest to researchers are concentrated. In the case of the Wallerstein and Kelly study, this was a clinical setting. Unfortunately for research purposes, however, the subset of divorcing or divorced families who would present themselves to a clinic for assistance is likely to be a group with characteristics atypical of the general population of divorcing couples with dependent children. While we cannot be certain that the study is unrepresentative, the odds are great that it contains substantial systematic sampling bias that will seriously compromise its external validity. The use of convenience samples may sometimes be justified when research on a given topic is in its early and exploratory stages and research funding is limited. While the results of such studies may be of interest to the scientific community because of their novelty and therefore merit publication, the applicability and utility of such research findings are extremely limited in a legal setting.²¹⁸

Even when a study intends to use a random sample, substantial dif-

216. For a general introduction to sampling theory in the context of social science surveys, see GRAHAM KALTON, *INTRODUCTION TO SURVEY SAMPLING* (John L. Sullivan & Richard G. Niemi eds., Sage University Paper Series: Quantitative Applications in the Social Sciences, No. 07-035, 1983).

217. JUDITH S. WALLERSTEIN & JOAN BERLIN KELLY, *SURVIVING THE BREAKUP: HOW CHILDREN AND PARENTS COPE WITH DIVORCE* (1980).

218. In addition to the sampling problems discussed above, the Wallerstein and Kelly study has been criticized because it did not include a control group of children who had not experienced divorce. Without such a control group, the authors' well-publicized conclusions concerning the negative impact of divorce on children are subject to criticisms related to internal validity. See discussion *infra* Part IV.B.2.

difficulties may arise in implementing the design. One such difficulty relates to *response rates*.²¹⁹ Once subjects are randomly selected for the sample, it is crucial that the researcher maximize the response rate of these individuals, households, or families, if the benefits of a random sample are to be realized. For example, in a divorce study researchers might identify all families that had divorced in a given year by using county-level court records. From such a listing, the researchers then would randomly select the cases for their sample. Once this sample selection occurs, the researchers are faced with the task of securing the participation of the selected families. The response rate is the number of families who participate divided by the number randomly selected for the sample.²²⁰ Securing a high response rate typically involves time-consuming tracking and follow-up efforts and, in some instances, inducements for prospective participants.²²¹ Low participation rates are problematic because non-participants may differ from participants in systematic ways that make the sample unrepresentative or biased and thereby jeopardize the generalizability of the study. Such unrepresentativeness is often referred to as a *selection bias*.²²² For example, if parents experiencing high-conflict divorces were significantly less likely to participate in the study described above, the resulting sample would not accurately represent relevant divorce processes, such as the use of joint physical or legal custody arrangements.²²³

How should judges assess response rates in sample-based family studies? First, it is crucial that judges inquire as to the exact response rate of any study brought before the court. A judge should be seriously concerned about a study that does not present information on its response rate. If this information cannot be ascertained, the study should be devalued. Second, a response rate in the range of seventy percent or better is generally considered to be quite good because it is unlikely, although not impossible, that selection biases will be large enough to seriously bias the results of the study.²²⁴ Third, regardless of

219. See generally SURVEY NONRESPONSE (Robert M. Grove et al., eds., 2002).

220. See THE AM. ASS'N FOR PUBLIC OPINION RESEARCH, STANDARD DEFINITIONS: FINAL DISPOSITIONS OF CASE CODES AND OUTCOME RATES FOR SURVEYS (3d ed. 2004).

221. Examples of studies of divorce that have used this approach include: MACCOBY & MNOOKIN, *supra* note 209; Maria Cancian & Daniel R. Myer, *Who Gets Custody?*, 35 DEMOGRAPHY 147 (1998).

222. For a general introduction to these problems, see Richard A. Berk, *An Introduction to Sample Selection Bias in Sociological Data*, 48 AM. SOC. REV. 386 (1983); KALTON, *supra* note 216.

223. For comparisons of response rates in recent divorce studies, see I-Fen Lin, *Perceived Fairness and Compliance with Child Support Obligations*, 62 J. MARRIAGE & FAM. 388 (2000).

224. DAVID B. LARSON ET AL., U.S. DEP'T OF HEALTH & HUM. SERV., THE SYSTEMATIC REVIEW: AN INNOVATIVE APPROACH TO REVIEWING RESEARCH (1992), available at <http://aspe.os.dhhs.gov/daltcp/reports/sysrevia.htm>. See also SURVEY NONRESPONSE, *supra* note 219.

whether the response rate is above or below seventy percent, judges should look for a discussion of possible bias related to the study's response rate. Generally, it is possible to compare the major characteristics of a sample with information from other sources about the population with which the study is concerned. For example, in the Stanford Child Custody Study, Mnookin and Maccoby compared their sample of divorce cases in California with information for the whole of the United States drawn from the U.S. Census Bureau's Current Population Survey. Such comparisons give the reader of a research report a reasonable sense of whether selection bias actually exists and, if so, its degree.²²⁵

If it appears that sample selection bias is a serious concern, there are various methods available that may be used to partially or fully remedy such problems, including (a) weighting the cases in the sample so that they better reflect the population to which the sample will be generalized and (b) creating measures called instrumental variables that explicitly gauge selection bias and control for it. Discussing these methods is beyond the scope of the present article, but it is important for judges to be aware that these methods exist and to seek expert assistance when questions of sample bias arise.²²⁶

Increasingly, family social science employs longitudinal survey designs in which a sample is selected at a certain point in time and then the individuals in the sample are followed over time. We briefly described how such designs may be used for causal analyses in Part IV.B.2.b. In such studies the problem of *sample attrition*, which is akin to response rate problems, may arise. Sample attrition occurs when members of a study at its beginning leave the study while it is still in progress for any number of reasons, such as death, moving, or unwillingness to continue participation. Sample attrition may result in sample bias if the people who leave the study differ from those who continue to participate in systematic manners that are associated with the social processes under investigation. In essence, sample attrition produces the same issues as a problematic initial response rate because it creates concerns about representativeness. Fortunately, as with response rate problems discussed above, methods have been developed that may be used to mitigate bias associated with sample attrition.

The important point to be taken from the preceding discussion is that when judges encounter research that employs a longitudinal data collection design, they should inquire as to the degree of attrition that has occurred during the course of the study and, if it is substantial, they

225. MACCOBY & MNOOKIN, *supra* note 209.

226. *See* Berk, *supra* note 222; KALTON, *supra* note 216.

should ascertain what efforts (if any) were undertaken to adjust for the effects of attrition.

b. The "Fit" Between Research and Legal Application

Beyond technical issues of sampling, there is a second type of external validity problem with which judges should be concerned when assessing direct research claims or science-based practice claims in court, namely the extent to which the research is related to the legal question the court is addressing. From an evidentiary point of view, this issue concerns whether there is a "fit," or a valid scientific connection between the research and the question.²²⁷

In family law, this kind of external validity question is often raised because social science research related to a legal question may not be at a very sophisticated level at the time the legal issue needs to be addressed. Often, this may be so because the legal issue precedes the research. For example, when courts were first confronted with joint custody questions, there was little research on post-divorce, joint-custody arrangements compared to other forms of custody arrangements, because joint custody was relatively new. Further, at this early stage the joint custody couples studied were those who had insisted on joint custody, even though it was not the norm at the time. Also, the divorced parents in these studies typically volunteered to participate rather than being randomly selected into the studies. Consequently, these early studies could not be used to generalize to more recent populations of divorcing parents for whom joint custody is an established option.²²⁸ In contrast, consider the current state of research on the impact of divorce on children. Divorce has been occurring among large numbers of families with children since the late 1960s and studies of the impact of divorce have become increasingly sophisticated since that time. Now when policies about divorce are debated, there are numerous high quality studies to consider.²²⁹

Recent debate about same-sex couple adoption provides another illustration of problems related to the generalizability of social science research and its fit in the context of legal decision-making.²³⁰ Because

227. See *infra* Part III.A.1.

228. For recent reviews of the social science literature on joint custody, see Kelly & Ward, *supra* note 160; Robert Bauserman, *Child Adjustment in Joint-Custody Versus Sole-Custody Arrangements: A Meta-Analytic Review*, 16 J. FAM. PSYCHOL. 91 (2002).

229. See, e.g., P. Lindsay Chase-Lansdale & E. Mavis Hetherington, *The Impact of Divorce on Life-Span Development: Short and Long-Term Effects*, 10 LIFE-SPAN BEHAV. & DEV. 105-50 (1990); Paul R. Amato, *Children's Adjustment to Divorce: Theories, Hypotheses and Empirical Support*, 55 J. MARRIAGE & FAM. 23 (1993); Morrison & Cherlin, *supra* note 190.

230. See, e.g., Michael S. Wald, *Same-Sex Couple Marriage: A Family Policy Perspective*, 9 VA. J. SOC. POL'Y & L. 291 (2001); Charlotte J. Patterson, *Adoption of Minor Children by*

adoption by same-sex couples is a relatively new occurrence, there is not a large pool of gay families who have adopted children and there are relatively few studies of adoptive gay parents. When directly related research is lacking, courts may be presented with related research. For example, there is research on gay parents, particularly custodial lesbian mothers raising their biological children.²³¹ Judges would need to consider whether results from studies of lesbian biological mothers are relevant to a population of both male and female same-sex couples who seek to adopt children to whom they are not biologically related.²³² Similarly, judges might need to consider whether research on heterosexual couples and adoption may be used to make inferences about gay adoption. Research that is clearly not relevant may also be presented to the court.²³³

In the context of foster care adoptions, however, a court might determine that the relevant comparison should be the likely impact of adoption compared to foster care, rather than of gay adoptive parents compared to hetero-adoptive parents. There seems to be substantial agreement that adoption is better than staying in foster care, and this research might be persuasive to a judge who is concerned about the relative lack of research on gay adoptive parents.²³⁴ The choice of a comparison group in this context is a normative decision, which could be justified by taking into account governmental policies encouraging adoption.²³⁵

Even if there are a number of excellent and relevant studies, social science research, in itself, does not provide a definitive answer to legal questions. For example, assume that high quality, national social science studies demonstrated that, on average, married parents spent twenty percent of household income on supporting one child. Further assume that the legislature amended the child support laws to require that upon

Lesbian and Gay Adults: A Social Science Perspective, 2 DUKE J. GENDER L. & POL'Y 191 (1995); Lynn D. Wardle, *The Potential Impact of Homosexual Parenting on Children*, 1997 U. ILL. L. REV. 833 (1997); Carlos A. Ball & Janice Farrell Pea, *Warring with Wardle: Morality, Social Science, and Gay and Lesbian Parents*, 1998 U. ILL. L. REV. 253 (1998).

231. See, e.g., Charlotte J. Patterson et al., *Children of Lesbian and Gay Parents: Research, Law, and Policy*, in CHILDREN, SOCIAL SCIENCE, AND THE LAW 176 (Bette L. Bottoms et al. eds., 2002).

232. *Id.*

233. See, e.g., Brief of Amicus Curiae The Rutherford Institute, *Cox v. Fla. Dep't of Health & Rehab. Serv.*, 656 So. 2d 902 (Fla. 1995) (No. 82-967).

234. See, e.g., Marianne Berry & Richard P. Barth, *A Study of Disrupted Adoptive Placements of Adolescents*, 69 CHILD WELFARE 209 (1990) ("adoption is generally preferred to foster care as a lasting and developmentally superior choice of home for all children and youths").

235. See, e.g., Karen Spar, *Adoption Promotion Legislation in the 105th Congress*, CONGRESSIONAL RESEARCH SERVICE No. 97-491 (The Library of Congress 1997) (discussing the Adoption and SAFE Families Act of 1997).

divorce, non-custodial parents of one child had to pay twenty percent of their income for child support. A judge who ordered a twenty percent payment of child support in a divorce case could feel confident that the order, applied to the individual in the case, was the correct order because it followed the law. In contrast, what if the judge was trying to “apply” the finding of the social science study? The judge should not assume that the divorcing couple before him had allocated their resources during their marriage in the same way as the average parent in the national studies. Further, the studies may have addressed only the behavior of married parents, not of divorced parents. A judge who applied a twenty percent rule based on his or her knowledge of the study would be going beyond the study and making a normative decision that consideration of the behavior of a sample population of married parents was a fair way to allocate child support for the specific divorcing parents before the judge.

In summary, judges need to be concerned with whether research can be generalized beyond the subjects of the actual study to the population of people who will be affected by a rule of law that the court is considering. Further, judges may need to determine the relevance of a study to a decision about the particular parties before the court.

4. ANALYTIC VALIDITY

Each of the methodological standards discussed thus far largely concerns the design used to collect the empirical data for a social science study of families or children. These standards are fundamental to good social science; indeed they are fundamental for all science, because high-quality data sets are a necessary condition for testing hypotheses and theories that can be used to expand the scientific knowledge base of the discipline. Even so, once the data for a study are collected, there remains the task of analyzing the data with the goal of rigorously testing the hypotheses that motivated the study.

Methods of analysis used in social science vary substantially depending largely on the type of data that have been collected. The techniques of statistical analysis used for experimental data, for example, differ from those used for longitudinal survey data, which in turn differ from those used to analyze ethnographic data. Hence social scientists, indeed all scientists, are concerned in their work with a fourth type of validity, analytic validity. Analytic validity refers to: (a) the appropriate match between the data to be analyzed and the technique(s) used in the analysis, (b) the appropriate application of the techniques to the data, and (c) the appropriate interpretation of the results of the analysis. Judges assessing the scientific validity of social science research should be attentive to issues of analytic validity because the use of inappropri-

ate techniques, the misapplication of appropriate techniques, or the misinterpretation of technically correct results may result in biased or misleading reports of results. In this subsection, we briefly illustrate problems of analytic validity with two examples and then suggest how problems of this sort may be identified without the courts themselves becoming experts in statistical and other forms of analysis.

a. Statistical Significance

When testing for a relationship between two or more variables in a sample-based data set, researchers typically report the level of statistical significance of the relationship. Statistical significance simply refers to a measure of the likelihood that the relationship between two or more variables in a sample has occurred by chance (due to the use of the random function to select the sample), rather than by the fact that the relationship actually exists in the population. Put differently, statistical significance tells the researcher the odds that a relationship among variables found in a sample data set is an artifact of studying a sample, rather than studying an entire population.²³⁶ Two points are key in assessing the use of significance tests in social science research. First, if the study's sample was not selected randomly or if the randomness of the sample has been seriously compromised, it is inappropriate to report the levels of statistical significance for relationships among variables found in the sample. For example, reporting the significance level of the relationship between joint-physical custody arrangements and post-settlement litigation in a nonrandom convenience sample of cases collected at a mediation center would be inappropriate. Second, a corollary to the first point is that the appropriate use of tests of statistical significance has everything to do with whether the sample was randomly selected and has relatively little to do with the size of the sample. Studies using large nonrandom samples have no more legitimate call to use tests of statistical significance than small convenience samples.²³⁷

236. For general introductions to the concept of statistical significance and significance tests, see LAWRENCE B. MOHR, UNDERSTANDING SIGNIFICANCE TESTING (Sage University Paper Series: Quantitative Applications in the Social Sciences, No. 07-073, 1990); RAMON E. HENKEL, TESTS OF SIGNIFICANCE (Eric M. Uslaner ed., Sage University Paper Series: Quantitative Applications in the Social Sciences, No. 07-004, 1976).

237. A topic related to statistical significance is that of *confidence levels and intervals*. While significance tests concern relationships between two or more variables, confidence levels assess the likelihood of error and the degree of expected error when we use a single sample variable to estimate a population variable. For example, if a researcher uses data from a national survey of divorces to estimate the proportion of all divorce settlements that call for sole paternal custody, he or she might say the following: *I am 95% certain that 10.6% (+/- 2%) of all divorce settlements call for paternal sole physical custody*. The 95% figure is the confidence level, while the +/- 2% figure is the confidence interval. In essence, the researcher is saying that if we believe that the true proportion of divorces with paternal custody is between 8.6% and 12.6%, we will be correct

b. Specification of Models and Relationships

Depending on the type of information that researchers collect from their research subjects, different methods are required for appropriate analysis of the data. For example, producing a statistical model of relationships among variables that are rankings of a respondent's views, such as levels of satisfaction with various terms of a divorce settlement decree, requires an analytic technique that differs from what would be used to model relationships among variables that are counts of respondent's characteristics such as age, income, or child support payments. Specifically, each of these types of data would require a different form of a statistical modeling technique generically referred to as multivariate regression analysis.²³⁸ The key point here is that if a form of regression analysis is misapplied to a particular type of data set, the resulting analysis runs the risk of misestimating the underlying relationships in the data. Regression analysis so commonly appears in social science and other scientific information proffered to courts that the Federal Judicial Center's *Reference Manual on Scientific Evidence* includes an entire chapter on regression analysis.²³⁹

How then is a judge to know if the appropriate analytic techniques have been applied to the data in a study presented to the court? At one level, materials such as the *Reference Manual on Scientific Evidence* are available and provide highly useful overviews geared to legal audiences on issues related to analytic validity.²⁴⁰ However, it is also true that analytic validity can be and frequently is a highly technical area requir-

95% of the time. For a study in which such an estimate is made, see Robert F. Kelly & William Rinaman, *The Structure and Prediction of Classes of Divorce Settlements Involving Dependent Children in a National Sample*, 38 J. DIVORCE & REMARRIAGE 1 (2003). What is key here is that the use of confidence levels/intervals involves the same statistical reasoning as significance tests and that each requires a randomly drawn sample in order to be validly employed.

238. For discussions of various types of regression analyses and their appropriate uses, see LARRY D. SCHROEDER ET AL., UNDERSTANDING REGRESSION ANALYSIS: AN INTRODUCTORY GUIDE (John L. Sullivan & Richard G. Niemi eds., Sage University Paper Series: Quantitative Applications in the Social Sciences, No. 07-057, 1986); SCOTT MENARD, APPLIED LOGISTIC REGRESSION ANALYSIS (Sage University Paper Series: Quantitative Applications in the Social Sciences, No. 07-106, 2d ed. 2001); JOHN H. ALDRICH & FORREST D. NELSON, LINEAR PROBABILITY, LOGIT, AND PROBIT MODELS (Sage University Paper Series: Quantitative Applications in the Social Sciences, No. 07-045, 1984); MELISSA HARDY, REGRESSION WITH DUMMY VARIABLES (1993).

239. See sources cited *supra* note 208.

240. See FRANK M. ANDREWS ET AL., A GUIDE FOR SELECTING STATISTICAL TECHNIQUES FOR ANALYZING SOCIAL SCIENCE DATA (2d ed. 1981); HANDBOOK OF DATA ANALYSIS (Melissa Hardy & Alan Bryman eds., 2004); GERALD VAN BELLE, STATISTICAL RULES OF THUMB (2002); PHILLIP I. GOOD & JAMES W. HARDIN, COMMON ERRORS IN STATISTICS (AND HOW TO AVOID THEM) (2003); and standard texts treating law and statistical analysis such as DAVID W. BARNES & JOHN M. CONLEY, STATISTICAL EVIDENCE IN LITIGATION: METHODOLOGY, PROCEDURE, AND PRACTICE (1986).

ing expertise that many judges simply will not have the time to develop in the course of their careers. Ideally judges would have access to staff or impartial experts who could assist in assessing the analysis.²⁴¹ As a short-cut, however, judges could rely on the underlying quality-assurance procedures of science, such as the peer-review process used by scientific journals. Through these processes the scientific community scrutinizes research reports on issues such as analytic validity. Specifically, we recommend that when questions of analytic validity arise, the courts should inquire both whether the research in question was published in a respected medium of scientific communication and to what levels and degree of review the research was subjected prior to publication. In the next section, we provide a detailed discussion of these scientific community social processes and ways in which the courts may rely upon them.

C. *The Social Practice of Science: Community Socialization, Review, and Communication*

The scientific method is constituted by the fundamental logic for developing and testing hypotheses and theories²⁴² and by a set of standards for the collection and analysis of empirical data,²⁴³ but there is a third essential system that is key to the scientific method, namely the social practices that together compose the scientific community. In this section, we briefly describe three basic social practices that characterize the scientific community, *socialization*, *review*, and *communication*, and the relevance of these practices to judges assessing social science research.

1. COMMUNITY SOCIALIZATION

Judges who assess social scientific research need to understand the processes through which social scientists are socialized, because these processes are among the most elemental in the scientific community's efforts to regulate the quality of research. Social scientists with Ph.D.s learn the logic and the standards of the scientific method in many ways. Most commonly, they receive graduate training comprised of a combination of course work and supervised research. This training requires the student-scientist to pass qualifying examinations prior to the beginning

241. Suggestions for aids to the Supreme Court have included social science special masters to consider social science factual disputes and a national center that would assist the Court by conducting needed research. See, e.g., Rustad & Koenig, *supra* note 111, at 159-61; see also Kenneth Culp Davis, *Judicial, Legislative and Administrative Lawmaking: A Proposed Research Service for the Supreme Court*, 71 MINN. L. REV. 1, 15 (1986).

242. See *infra* Part IV.A.

243. See *infra* Part IV.B.

of dissertation research. Then there is the dissertation itself: a book-length piece of original research that must be publicly defended and approved by a committee of senior professors. Importantly, during their graduate training social scientists intending to pursue careers involving research will apprentice as research assistants on research projects headquartered at their graduate schools and led by senior professors. Further, it is increasingly common for research-oriented social scientists to complete post-doctoral fellowships for two or three years after they receive their degrees. In these "post-docs," the professionally-young social scientist may learn more specialized research skills, hold a leadership position on a research project, and intensively pursue a specialized research project alone or as a member of a team. While there is substantial variation in the prototypical training of a social scientist that we have described schematically, there is also enough uniformity in the process so that a judge faced with assessing a study produced by a social scientist ought to be able to ascertain whether the researcher has in fact been properly trained in the logic and methodological standards of the social sciences.

It is also important to note that graduate social science training typically includes seminars that deal with professional behavior and the code of ethics of the discipline.²⁴⁴ As related to the current discussion, these codes include descriptions of problems associated with researcher bias and how best to minimize its effect on the quality of research. For example, when the research is submitted to a journal for publication, such codes require the researcher to fully disclose funding sources for the research project, the institutional affiliation of author, and potential conflicts of interest. The goal of such mechanisms of transparency is to reduce researcher bias.²⁴⁵ Hence, judges should be mindful that, like the bar, each social science discipline operates with a code of ethics and that each code provides standards that are used in the research review process.

The professional socialization and training of social scientists do not end with a Ph.D. or a post-doctoral fellowship, just as legal education does not end with the granting of the J.D. and the passing of a bar examination. Social scientists, like other scientists, continue training in their substantive fields of specialization and their disciplines' methodologies by maintaining an active voice in their disciplines' scientific com-

244. *See, e.g.,* AM. SOCIOLOGICAL ASS'N, CODE OF ETHICS AND POLICIES AND PROCEDURES OF THE ASA COMMITTEE ON PROF'L ETHICS (1997).

245. Because human beings are the typical research subjects of social scientific research, social scientists are bound by ethical codes governing the treatment of human subjects in research. Training in such codes is typically part of graduate methodological training and journals typically require that researchers have complied with their institution's human subject policies.

munity. We discuss the nature of these communication practices below, but in the context of our discussion of continuing social scientific training, it is important to note that the professional resumé of the author or co-author of a study may be examined for indications of the degree to which the researcher has continued his or her professional training and socialization. Such indications might include the following activities: presenting research papers at professional meetings, publishing research papers in peer-reviewed journals, reviewing papers for presentation at professional meetings or for publication in journals, reviewing research proposals for government and other funding organizations, and serving on editorial and advisory boards. Again, as with graduate training, it should be possible to ascertain from researchers' resúmes the degree to which they have continued to train in their discipline. While indicators of training and continuing education will not be sufficient in themselves to determine the scientific merit of a piece of research proffered to the court, they are clearly relevant and should be assessed with care.

2. COMMUNITY REVIEW

One of the distinguishing features of the scientific method and the type of knowledge it produces is the intense review procedure through which research reports must travel before and after formal communication of results to the discipline. Indeed, when scientists use the term *objectivity*, they typically are not referring to the achievement of a state of the absolute absence of bias, but rather to the degree to which research reports have successfully undergone the process of critical community review. More formally, objectivity is the process by which the scientific community continuously scrutinizes the methodological and theoretical merits of research results and explanations presented by its members with the goal of reducing bias.²⁴⁶ Two key aspects of this understanding of objectivity should be noted. First, the scientific community can never declare a research finding to be free of bias. Rather, implicit declarations are made as to the degree to which findings have been scrutinized for bias and the degree to which the findings have withstood this scrutiny. Second, objectivity is inherently a collective or social process. It is the scientific community that produces objectivity, not individual researchers.

A judge who is reviewing a social science study should inquire into how rigorously the study was reviewed, because in doing so the judge is inquiring into the objectivity of the research. It is precisely for this reason, we believe, that the third question that the Supreme Court in

246. See, e.g., JOHN ZIMAN, *RELIABLE KNOWLEDGE: AN EXPLORATION OF THE GROUNDS FOR BELIEF IN SCIENCE*, 107-09 (1978); Beyea & Berger, *supra* note 161, at 338.

Daubert instructed judges to ask of information proffered as scientific is: *Has the scientific claim successfully passed peer review?*²⁴⁷ We next turn our attention to describing this process.

The centerpiece of the process of scientific objectivity is pre-publication peer review.²⁴⁸ Generally, peer review operates as follows. A research report is submitted to a scientific journal for possible publication.²⁴⁹ The norm is that a paper may be under review only at a single journal at a time, a norm that differs greatly from that typically followed by law reviews in the United States, which allow authors to make multiple, simultaneous submissions. After a preliminary review by the journal's editor, any indication of the author's identity and affiliation is removed from the paper and copies are sent to reviewers who have expertise in the areas covered by the paper. In this process of blind peer review, there are typically three or more reviewers who are asked to assess the paper for its scientific merit. In evaluating the research paper and findings, reviewers are expected to employ the standards of the scientific method that we have previously described. Specifically, they are to determine whether the logic of scientific inquiry was followed in the research.²⁵⁰ Reviewers then examine the merit of the work in terms of methodological design and analysis standards, specifically examining measurement validity and internal, external, and analytic validities.²⁵¹ Reviewers are also mindful of professional ethical standards²⁵² in their review, particularly as they relate to potential researcher bias and to the protection of human subjects.

After the journal receives the reviews, a judgment is made on the research report. Judgments generally fall into the following categories: accept for publication; accept pending minor revisions; reject with an invitation to undertake major revisions and then to resubmit for further reconsideration; and reject. In the social sciences, most papers are either rejected or given "revise and resubmit" decisions at this point. For papers other than those rejected, editors will make the necessary inquir-

247. *Daubert v. Merrell Dow Pharm., Inc.*, 509 U.S. 579, 593 (1993).

248. See FOSTER & HUBER, *supra* note 12, at 163-205.

249. Procedures similar to those described above are typically also used for papers submitted to professional meetings for presentation or research proposals submitted to funding agencies, although procedures do vary. The review process for research reports submitted as book manuscripts is more difficult to describe. For university presses, the process is very similar to that described in the text with respect to social science journals. For trade publishers, however, there may be little if any scientific peer review with a greater emphasis placed on public appeal than scientific merit. Hence, in assessing research presented in book form, judges need to be active in ascertaining the exact nature of pre-publication review that has taken place.

250. See *infra* Part IV.A.

251. See *infra* Part IV.B.

252. See discussion of professional socialization *infra* Part IV.C.1.

ies related to professional ethics, such as those related to full disclosure of funding sources. For authors who receive a revise and resubmit decision and who in fact do revise and resubmit, the process described above continues until the paper is either acceptable or the determination is made that it does not merit publication. Publication rates vary among social science disciplines, but for lead journals published by professional associations of social scientists, the annual acceptance rate is likely in the range of ten to fifteen percent.²⁵³

The conclusion that judges should unflinchingly take from this discussion is that it is essential to inquire as to whether any research under consideration by the court has undergone peer review and, if so, the exact nature of the peer review. Research that has been published after thorough peer review should be given substantially greater weight than that which has not.

Beyond peer review, other methods are used to enhance objectivity in social science, two of which are particularly relevant to our analysis, namely replication and public access to data for re-analysis. As a general rule, peer reviewers and the scientific community give greater credence to research findings that have been replicated than to those that are reported for the first time. Replication generally means that researchers have been able to reproduce the same result under the same or very similar circumstances in separate studies.

Pure replication can be difficult to achieve in the social sciences because most research is conducted in ongoing social systems rather than highly controlled laboratory settings, where the conditions needed for replication are relatively easy to achieve. Because social systems are dynamic, it is difficult to meet the expectation inherent in replication: that research conditions across replication studies will be the same or very similar. For example, even if three studies of custody arrangements in divorce settlements were conducted at the same time in California, Wisconsin, and Michigan, they would not be pure replications of each other because the laws regulating divorce in the three states would differ.²⁵⁴ These studies then at best would be quasi-replications. If they produced similar results, the results would likely be given the enhanced status of having been replicated in spite of the differences between the

253. To illustrate this point, we calculated the acceptance rate for the four lead journals of the American Sociological Association in 1996. It was 11.5%. For important descriptions and critiques of the peer-review process, see Elizabeth Knoll, *The Communities of Scientists and Journal Peer Review*, 263 JAMA 1330 (1990); David F. Horribin, *The Philosophical Basis of Peer Review and the Suppression of Innovation*, 263 JAMA 1438 (1990); Arnold S. Relman & Marcia Angell, *How Good is Peer Review?*, 321 NEW ENG. J. MED. 827 (1989).

254. See Greer Litton Fox & Robert F. Kelly, *Determinants of Child Custody Arrangements at Divorce*, 57 J. MARRIAGE & FAM. 693 (1995).

states. On the other hand, if the findings were not replicated across the three studies, it would be difficult to draw strong inferences because differences between the states in their legal systems or in other important characteristics might plausibly account for the differences in the study results.

In general then, replication in the social sciences operates in a slower and much less efficient manner than in the natural sciences, because the social sciences cannot rely upon laboratory controls to provide equivalent research settings for the replication of studies.²⁵⁵ Nonetheless, judges assessing research findings should inquire as to whether or not research findings before them have been replicated and the degree to which the findings have been replicated. Findings that have been replicated repeatedly should be given greater credence than those that have not been replicated. As with peer review, this analysis of replication is consistent with *Daubert* in which the Supreme Court requires judges considering information proffered as scientific to inquire as to whether the information has acquired widespread acceptance: the fourth *Daubert* question.²⁵⁶ Within the scientific community, one of the necessary conditions for widespread acceptance of a finding is that it has been replicated.

It is increasingly common in the social sciences to require researchers to make the data sets upon which their published analyses are based available to other researchers for possible re-analysis.²⁵⁷ The agency that originally funded the research, for example, may require the researcher to make the data available as a condition of receiving the grant. Family social science projects in which this was the case include the National Survey of Families and Households, the National Survey of Children, and Maccoby and Mnookin's Stanford Child Custody Study. Journals may also require that researchers whose work is to be published provide other researchers with access to their data sets at the time of publication. This is often done by placing the data on a web site and providing the web site's URL in a footnote to the article. Highly regarded journals that have such a policy or have experimented with it include *Science*, the premier journal of the American Academy for the Advancement of Science, which publishes both natural and social science research; *Demography*, the highly regarded journal of the Population Association of America; and the *American Political Science Review*, the lead journal of the American Political Science Association.

255. The major exception to this statement is psychology in which well-controlled lab-based experiments are common and therefore replication is more likely to occur.

256. *Daubert v. Merrell Dow Pharm., Inc.*, 509 U.S. 579, 597 (1993).

257. See generally Kenneth J. Meier, *The Value of Replicating Social Science Research*, CHRON. HIGHER EDUC., Feb. 7, 1997, at B7.

The underlying rationale for making data publicly available is that researchers who know that other researchers may reanalyze their data will be more cautious in data collection, analysis, and reporting of results.

In this regard, it should be noted that one of the most influential social science studies on the impact of divorce reforms in the 1970s and 1980s in the United States, Lenore Weitzman's *The Divorce Revolution*, was heavily criticized and many of its findings were cast into serious doubt because, when reanalyzed, the data sets that Weitzman provided to other researchers did not yield the same results as she reported in her book.²⁵⁸ Given the increasingly common requirement that data be made available to the scientific community, it would be wise for judges assessing social science research to inquire whether the data sets upon which research reports are based have been made available to the scientific community. When the data sets have been made available, greater credence should be lent to the research.²⁵⁹

The description of standard, community-review procedures in the social sciences that we have provided should not be interpreted as a claim that biased or even fraudulent work goes unpublished. The history of science tells us otherwise. Rather, the point we wish to make is that the scientific community and its institutions have a strong commitment to the process of auto-criticism and that this commitment distinguishes the scientific community from many other social institutions. Further, it is likely that this commitment to auto-criticism is one of the cornerstones upon which the impressive historical progress of scientific knowledge is built. Hence, judges reviewing social scientific research should scrutinize it for distinguishing signs indicating that it has been thoroughly filtered by the review processes of science.²⁶⁰

3. COMMUNITY COMMUNICATION

Once research reports are reviewed and found to merit the attention of the scientific community, they must be communicated to the commu-

258. See LENORE J. WEITZMAN, *THE DIVORCE REVOLUTION: THE UNEXPECTED SOCIAL AND ECONOMIC CONSEQUENCES FOR WOMEN AND CHILDREN IN AMERICA* (1985); Richard R. Peterson, *A Re-Evaluation of the Economic Consequences of Divorce*, 61 AM. SOC. REV. 528 (1996); Lenore J. Weitzman, *The Economic Consequences of Divorce Are Still Unequal: Comment on Peterson*, 61 AM. SOC. REV. 537 (1996); Richard R. Peterson, *Statistical Errors, Faulty Conclusions, Misguided Policy: Reply to Weitzman*, 61 AM. SOC. REV. 539 (1996).

259. Organizations such as SOCIOMETRICS collect important family data sets and make them available in easy-to-analyze formats. See, e.g., Sociometrics "American Family Data Archive," at <http://www.socio.com>; see also the University of Michigan's Inter-University Consortium for Political and Social Research, at <http://www.icpsr.umich.edu>.

260. See Bauer, *supra* note 154, for a description of the scientific review process as a huge knowledge filter.

nity. This communication is also a cornerstone of scientific progress. Community communication is fundamental because the systems and problems that scientists seek to understand are large and complex, resembling an immense jigsaw puzzle, and no one researcher or research team could possibly work on the entire puzzle alone.²⁶¹ Hence, the day-to-day work of researchers typically focuses on mid-sized and smaller problems or systems, that is, parts of the puzzle. Under these conditions, the only hope for the scientific community as a whole to advance the understanding of big systems and problems is for the researchers who constitute the community to continuously communicate with each other. It is precisely for this reason that scientific communication, typically through journal articles, is so central to the enterprise of science. Indeed, it has been argued credibly that the pace of scientific progress did not accelerate in the modern era until the means of scientific communication became more rapid.²⁶²

The recognition of the centrality of scientific communication has an important implication for judges considering social scientific family research. Research that has been published in social scientific journals should be given greater credence not only because it has passed the scrutiny of the peer review process of journals, but also because the results have been communicated to the broader scientific community and will be tested with respect to its contribution to the larger puzzle-solving enterprise. Research that is isolated from this essential communication loop because it has not been published may represent interesting information, but it is not scientific knowledge because it has not been integrated into the essential practice of scientific communication. In many ways the problem of such isolated research is analogous to a hypothetical situation in which an appeals court writes a precedent setting decision, but fails to communicate it to lower courts. Is such a decision, however thoughtful and innovative, precedent? In this regard, it is notable that one of the major criticisms of the concept of parental alienation syndrome (PAS) is that much of the work upon which it is based is self-published and therefore out of the scientific communication loop. Hence, the concept and its use in divorce-related research and clinical practice lacks the credibility that would normally be attributed to concepts that had undergone peer-review and succeeded in being published in a scientific journal.²⁶³

Beyond research reports communicated in the form of journal arti-

261. We borrow the jigsaw puzzle metaphor from Michael Polyani, *The Republic of Science: Its Political and Economic Theory*, 1 *MINERVA* 54 (1962).

262. See *id.*

263. See, e.g., Kelly & Johnston, *supra* note 213.

cles, there is another type of scientific communication that bears directly on the work of judges who must assess social science research, namely the literature review. Periodically, journals and other publications such as annual reviews of professional associations publish literature reviews that seek to summarize the present state of knowledge in an area of scientific research. For example, the *Journal of Marriage and Family* produces a decade review in which the most important research published in fundamental sub-fields of family studies over the past ten years is synthesized in fifteen to twenty integrative articles.²⁶⁴ Similarly, the American Sociological Association produces an *Annual Review* in which synthetic articles are commissioned in fundamental areas of sociological research.²⁶⁵ The goal of such articles is to present a holistic picture of where knowledge stands in a particular field and to thereby provide researchers with a summary of what is well-known and well-replicated, what is less certain, and what crucial problems need to be addressed by researchers. In short, literature reviews communicate to researchers the current state of work on the jigsaw puzzle and, in so doing, the literature review seeks to provide guidance for future work.

Literature reviews may be qualitative or quantitative in nature. In the case of the former, a highly regarded researcher gathers the relevant studies, summarizes them, and then produces an integrative statement based on the researcher's judgment.²⁶⁶ Over recent decades researchers have increasingly used quantitative methods to produce literature reviews that may be categorized by the broad title of *meta-analyses*. Meta-analyses gather all relevant studies on a topic, code the research on multiple dimensions of its methodology, categorize key findings that are comparable across studies, and then enter this information into statistical models that assess whether hypothesized effects have been demonstrated to exist.²⁶⁷ An important recent meta-analysis relevant to family law is Bauserman's analysis of a large number of studies that assessed the

264. The JOURNAL OF MARRIAGE AND FAMILY is the research journal of the National Conference of Family Relations, the largest association of family professionals in the United States. For the two most recent decade reviews, see 62 J. MARRIAGE & FAM. 873-1307 (2000) and 52 J. MARRIAGE & FAM. 807-1172 (1990).

265. For a recent volume containing review articles on older children leaving home, the consequences of divorce for children and stepfamilies in the United States, see 20 ANNUAL REVIEW OF SOCIOLOGY (John Hagan & Karen S. Cook eds., 1994).

266. See Ramsey & Kelly, *supra* note 10, at 675-83.

267. For discussions of these methods, see MORTON HUNT, HOW SCIENCE TAKES STOCK: THE STORY OF META-ANALYSIS (1997); THE HANDBOOK OF RESEARCH SYNTHESIS (Harris Cooper & Larry V. Hedges eds., 1994); FREDRIC M. WOLF, META-ANALYSIS: QUANTITATIVE METHODS FOR RESEARCH SYNTHESIS (Sage University Paper Series: Applications in the Social Sciences, No. 07-059, 1986). For a discussion of the utility of meta-analysis for the legal system, see Jeremy A. Blumenthal, *Law and Social Science in the Twenty-First Century*, 12 S. CAL. INTERDISC. L.J. 1, 38-46 (2002).

impact of joint-custody arrangements on children.²⁶⁸

The role of research literature reviews, whether qualitative or quantitative, in the process of scientific communication makes it plain that the single research article, while important in its own terms, gains its larger significance and ultimate scientific meaning only when placed in the context of related work by other members of the scientific community. This fact is important in the context of social science research proffered to the courts. Judges should not rely on any single piece of research in evaluating a social scientific issue or question, but rather should, to the degree possible, supplement specific pieces of research with literature reviews that provide context for specific research findings presented to the court.²⁶⁹ It is in doing so that judges will be able to assess the degree to which direct research claims and science-based practice claims have achieved general acceptance. Recall that the fourth inquiry that judges are required to pursue in considering research findings or theory proffered to the court as scientific under *Daubert* concerns the degree to which the research findings or theory have gained general acceptance.²⁷⁰

V. CONCLUSION

Social science research and theory may be used properly and profitably by courts in their handling of family law cases. Our general conclusion in many ways follows *Daubert* in that we urge judges to examine whether social science proffered to the court has been produced in manners consistent with the scientific method. In Parts II and III we developed typologies of the uses and sources of social science knowledge claims that come to the courts (see Table 1). In Part IV, we sought to provide a detailed map of questions and standards that judges may employ in making such methodological assessments (see Table 2). In this section, we offer concluding reflections on these themes and make suggestions for how the judicial and social scientific communities might better cooperate in the future.

First, the application of *Daubert* and its progeny is limited to expert testimony. From our analysis, however, it should be clear that social science research and theory come to the courts in additional forms, such as briefs and judicial notice. We believe that all social science research

268. Bauserman, *supra* note 228.

269. Other illustrations of such reviews on topics relevant to family and child law include: Kelly, *supra* note 202, on the effectiveness of family preservation and reunification programs, and Paul Amato, *The Consequences of Divorce for Adults and Children*, 62 J. MARRIAGE & FAM. 1269 (2000), on the impact of divorce on children and parents.

270. In ways that are very similar to the literature review, the best graduate and undergraduate textbooks perform a similar literature review function.

and theory used by the courts in whatever form should be subjected to the same level of scrutiny that we have articulated in Part IV. In this, we clearly extend *Daubert*. We do so because we expect that the scrutiny of all scientific claims entering a judicial proceeding is likely to enhance procedural fairness and reduce the risk of unvalidated scientific claims influencing the substantive outcomes of such proceedings.

Second, a practical question that arises from our analysis and our proposal to extend *Daubert* concerns strategies for assisting the courts in the task of scrutinizing the methodological validity of social science research and theory. Specifically, beyond asking the questions and applying the standards outlined in Part IV, what can judges do to improve their use of social science short of becoming social scientists themselves? We have two suggestions that we believe will be helpful in this regard.

One suggestion is that the family law judiciary should better communicate its research needs in a systematic and regular manner to social scientists. Left to their own devices, social scientists select topics for research that interest them and then conduct research on these topics in accord with their discipline's framing of the topic. The research that results, however, may not address relevant legal questions faced by judges. To provide for better communication between the judiciary and social scientists, committees of judicial or other legal associations should work with social scientists to formulate research questions that derive explicitly from the experience of judges and reflect the judges' needs. Research questions could then be forwarded to social science professional associations for publication. They might also be forwarded to governmental agencies and foundations with research funding functions with the expectation that the lists would influence research funding priorities.

Another suggestion is that courts should consider commissioning nonpartisan research organizations to produce research reviews summarizing the state of current research on social science topics that regularly come before the courts, especially the topics about which there is doubt concerning conclusions that may be validly drawn from the literature. In conducting such reviews, the staff of the research organization could be instructed to explicitly address the key issues in *Daubert* as we have outlined them in Part IV. Such reviews would go a long way toward assisting judges in objective assessments of social science research.²⁷¹

271. Many not-for-profit, nonpartisan research organizations already release research reviews relevant to family law matters on a regular basis. Illustrations include: the Child Trends Research Brief series (<http://www.childtrends.org>), the Board on Children Youth and Families of the National Academies (<http://www.bocyf.org>), and the journal, *The Future of Children*, published

Third, judges need to keep in mind that social science cannot provide complete answers to the difficult questions they confront every day. There are basic tensions between social science knowledge and judicial decision-making. Social scientists study populations or classes of people (or samples thereof) and their studies typically allow for inferences only at the aggregate level, not at an individual level. Trial judges, however, need to decide cases involving specific individuals. This problem is exacerbated by the fact that judges deciding family matters frequently need to make a decision based on a prediction of what will happen in the future to these individuals. In using social science, the judge cannot treat a social science finding like a rule of law and apply it to an individual. Instead, the judge must determine the extent to which the characteristics of the population studied might be relevant to the case to be decided and what normative values are relevant as well.

The new gate-keeping role of judges provides many challenges, particularly in the family law area, because family law jurisprudence is increasingly influenced by social science. Judges need to be knowledgeable consumers of social science research, so that they can critically assess not only expert testimony, but also briefs, training programs, books, and other judicially noticed materials that they consider in law formation and adjudication. We are cautiously optimistic that a mutually critical and profitable relationship can be forged between judges and social scientists and that families will be the better for it.