

# How Many Words for a Camel? A Perspective on Judicial Evaluation of Social Science Evidence

---

Neil VIDMAR, Ph.D.\*<sup>1</sup>

<b>I.</b>	<b>PSYCHOLOGICAL DIAGNOSIS AND PREDICTION</b> .....	6
<b>A.</b>	<b>The Problem of False Positives and False Negatives</b> .....	6
<b>B.</b>	<b>Convergent and Discriminant Validity</b> .....	7
<b>II.</b>	<b>DRAWING INFERENCES FROM EXPERIMENTAL STUDIES: INTERNAL AND EXTERNAL VALIDITY</b> .....	10
<b>III.</b>	<b>THE CHANGING STATE OF SOCIAL SCIENCE KNOWLEDGE</b> .....	12

---

<sup>1</sup> Professor of Social Science and Law, Duke Law School and The Private Adjudication Center, Durham, N.C.

[In Arabic], [...] [T]here are said to be about six thousand names connected in some way with "camel" [...] These include, for instance, names of classes of camels according to function - milk camels, riding camels, marriage camels, slaughter camels, etc.; names of breeds of different degrees of nobility of lineage, derivation from different lands, etc., names of camels in groups, as several, a considerable number, innumerable, and with reference to their objectives - grazing, conveying a caravan, war expedition, etc.; as many as fifty words for pregnant camels, states of pregnancy, stage at which movement of the fetus is first felt, mothers who suckle and those who do not, those near delivery, [...]<sup>1</sup>

Social science evidence has come to play an increasing role in Canadian and American courts, and its presence is likely to grow as we move toward the twenty-first century. I use the term "social science" broadly, so as to include not only psychology, anthropology, sociology, political science, and economics, but also fields such as psychiatry and social work. As forensic evidence, all of these disciplines can speak to matters concerning human behaviors, states of mind, or patterns of human activity.<sup>2</sup>

The most traditional form of forensic social science evidence has involved such matters as diagnoses of insanity, incompetence or dangerousness. In addition, courts are familiar with survey evidence bearing on legal disputes about obscenity or about consumer confusion regarding trademark infringement. Newer forms of social science evidence involve testimony about child sex abuse, battered women syndrome, and patterns of racial or gender disparity indicating possible discrimination. Survey evidence has been tendered to show pre-trial prejudice in motions for challenges for cause or change of venue,<sup>3</sup> and trial by judge instead of jury.<sup>4</sup> Experimental studies have been offered as evidence to show bias in a police lineup<sup>5</sup> and on the effects of television and print media on public attitudes.<sup>6</sup> The very first Canadian constitutional case<sup>7</sup> set the stage for the introduction of possibly an infinite variety of social science evidence bearing on violations of the Charter and, subsequently, in *R. v. Oakes*<sup>8</sup> it was asserted that it will usually be necessary to introduce evidence in section 1 cases to justify a limitation on a right guaranteed by the Charter.<sup>9</sup> Outside of the trial arena, social science studies are used to help formulate judicial and legislative policies.

It is probably not necessary to make the point that the quality of social science evidence varies from good to bad. Some of it is biased. For example, one psychologist who frequently testifies in Canadian child sex abuse cases told me that he can "always" tell if a child has been sexually abused. This claim was made despite the fact that there is a mountain of psychological literature indicating the fallibility of clinical diagnoses in general<sup>10</sup> and child sex abuse in particular. Indeed, a recent comprehensive review of evidence in child abuse cases in the United States has led the authors to the conclusion that

*so-called experts often make claims that are not in accord with - or are even diametrically opposed to - [extant bodies of psychological] research [...] Experts rarely present a careful summary of the research because doing so would probably force them to attenuate their often strident claims.<sup>11</sup>*

Furthermore, even when the evidence itself may be good, courts may misinterpret it. Professor Carl Baar, for example, has documented how in *R. v. Askov*<sup>12</sup> the Supreme Court of Canada misunderstood and misapplied his research and other data on the pace of criminal litigation in Canadian trial courts.<sup>13</sup> The unfettered adoption of mediation or other forms of alternative dispute resolution based around questionable social science studies is another problematic area.<sup>14</sup>

Do not misunderstand my point. I am not arguing against the use of social science for forensic or other legal purposes. My own professional career is devoted to producing data and conclusions that can shed light on matters of legal interest and be of use in the making of legal policy. I am, however, raising the problems posed for judges in evaluating social science evidence, in separating the good from the bad and understanding when it is or is not relevant to a particular situation.

What does all of this have to do with camels? As a metaphor, perhaps quite a lot. To me a camel is a camel and a pregnant camel is a pregnant camel. I'm not even sure that I'd know if a camel is pregnant. I have no need or desire to have six thousand, or fifty, words to describe it.

But for an Arab nomad one can quickly see the relevance of such a vocabulary: for example, the recipient of a marriage camel would want to know its lineage to judge the value of the bridal arrangement; the prospective purchaser of a breeding camel would want to inquire about the matters relating to fecundity; beginning a hazardous desert journey most travellers would prefer to be astride a riding camel rather than a slaughter camel.

More than seven decades ago social psychologists and psycholinguists observed that when an object or entity has important relevance or meaning to a culture, its people develop vocabularies that reflect the nuances of their perceptions of it.<sup>15</sup> Camels and their differentiation along many dimensions are important to Arab nomad culture. Similarly, the Inuit language contains many words for snow because of the Inuit need to know precisely how to contend with their harsh environment. The obverse of this language principle also applies. Cultures that have limited vocabularies for certain concepts have greater difficulty communicating and making finer distinctions about objects. Some cultures have number systems that stop after three; more than that is "many". In contrast to the Inuit, the Aztecs used the same basic word for cold, ice, and snow; in the Hopi Indian language the same word was used for insect, airplane, and aviator.<sup>16</sup>

Still, you must be saying, get to the point. How does all of this bear on the evaluation of social science evidence? My point is that often judges have a vocabulary for evaluating the predictive, methodological and conceptual adequacy of social science that is akin to my vocabulary for a pregnant camel. They need the equivalent of the Arab's vocabulary.<sup>17</sup> At minimum, they should know the meaning and appropriate usage of some words and be cognizant of the fact that a more extensive vocabulary exists.

For instance, when called upon to evaluate the merits of scientific evidence, both trial and appellate judges typically refer to its "scientific reliability" when in fact they really are considering its validity.<sup>18</sup> To a social science methodologist "reliability" refers to the degree to which a test, measurement or diagnosis gives consistent results. "Validity" is the term used to refer to the

degree to which a test, diagnosis, or scientific theory measures what it purports to measure or explains what it purports to explain.<sup>19</sup>

A test can be highly reliable but have poor validity. For instance, there is a psychological test (still used by some psychiatrists and clinical psychologists, unfortunately) in which the person being evaluated is asked to draw a house, a tree, and a person. Persons who draw houses with small windows and persons with narrow slanted eyes are often assessed as paranoid by clinicians who use this test. If you retest those same persons later, the likelihood of getting small windows and slanted eyes is high. The problem, of course, is that the theory that drawing slanted eyes proves paranoia is highly suspect. The house, tree, person test is highly reliable but probably has zero validity.

Like the Arabs with their camels, a social science methodologist has a pretty extensive vocabulary that distinguishes types of reliability: for example, inter-rater reliability; intra-rater reliability; test-retest reliability, split-half reliability. There are even more words to describe forms of validity: for example, face validity, predictive validity, criterion validity, concurrent validity; convergent validity; discriminant validity; incremental validity; construct validity; ecological validity; internal validity; external validity. Related concepts such as base rates, selection ratios, and false positive and false negative errors are used to describe the degree to which psychologists and psychiatrists can make accurate diagnoses.<sup>20</sup>

At Duke Law School I teach a whole course on social science evidence, as I did previously at the University of Western Ontario and Osgoode Hall law schools. It is a subject worthy of book-length treatment. I cannot begin to cover the field in the brief space allotted here. However, I have chosen some topics to help illustrate various problems of this field of evidence and indicate how familiarity with some of these methodological terms might assist a judge in assessing the merits of social science evidence.

## **I. PSYCHOLOGICAL DIAGNOSIS AND PREDICTION**

### **A. The Problem of False Positives and False Negatives**

Twenty years ago Professor David Rosenhan published a study titled "Being Sane in Insane Places" that spoke of the problem of the accuracy of psychiatric diagnosis.<sup>21</sup> In his study a number of mentally normal persons, acting as his research confederates, contacted various psychiatric hospitals with a complaint that they had been hearing voices. At the admission interview these confederates gave false names and occupations and maintained that they heard voices; but otherwise they gave accurate accounts of their life histories. All of the "pseudo patients" were admitted to the hospitals with a diagnosis of schizophrenia. While on the hospital wards they behaved like normal persons, except that they took notes about their experience. Many of the real patients on the wards were immediately suspicious of the pseudo patients and asked them whether they were inspectors or researchers. In contrast, the psychiatric staff continued to view the pseudo patients as mentally ill; some wrote in the medical charts that the patient engaged in "compulsive note taking". Other behaviors that would be viewed as normal in another context were labelled as manifestations of psychiatric disturbance. The patients remained in the hospital for days or weeks, and most were ultimately released with a diagnosis of "schizophrenia in remission".

When they learned about Rosenhan's study, many psychiatrists and other mental health workers were upset with the implications and claimed the study was flawed; it underestimated their predictive prowess. In consequence, Rosenhan conducted a follow-up study. He told the staff of a research and teaching psychiatric hospital that during the next few months one or more pseudo patients would attempt to gain admission to the hospital. The psychiatric staff were thus challenged to uncover the pseudo patients. Of 193 persons admitted for treatment during a three month period, 41 were alleged, with high confidence, to be a pseudo patient by at least one member of the staff; 23 were considered probable pseudo patients by at least one psychiatrist, and 19 were suspected by at least one psychiatrist and one other staff member. In fact none of Rosenhan's pseudo patients ever attempted to gain hospital admission.

The first Rosenhan study helps demonstrate the problem of "false negatives", in this instance the tendency to diagnose a healthy person as sick. The second study demonstrates the problem of "false positives", in this instance to call a possibly sick person healthy.<sup>22</sup> Part of the problem of psychiatric diagnosis in the first study accrues from the fact that "base rates" were high: that is, the overwhelming majority of persons seeking hospital admission do have psychological disturbances. This knowledge undoubtedly affected the psychiatrists' "selection ratio", that is, the proportion of diagnoses of insanity compared to diagnoses of sanity.<sup>23</sup> Notice, however, how the second study changed the assumptions about "base rates" and affected their "selection ratios". The problem of false positives and false negatives and the effect of "base rates" and assumptions about base rates on "selection ratios" is endemic to psychological prediction of all kinds.<sup>24</sup> No test or diagnosis is one hundred percent accurate, but we can often get estimates of its rate of accuracy, that is its tendency to produce false positives and false negatives. Some tests have very high false positive rates. The problem may be particularly acute in the diagnosis of sexual abuse in young children when there is no evidence of physical assault. Experts cannot agree on many of the symptoms or methods of diagnosis of sex abuse, and the problem is exacerbated by assumptions that may affect "selection ratios". Experts contacted to make diagnoses may, like the mental health professionals in Rosenhan's study, begin with assumptions that predispose them toward a high rate of false positives. This leads me to the concepts of "convergent" and "discriminant" validity.

## **B. Convergent and Discriminant Validity**

Convergent validity refers to the ability of different measures or assessments to yield the same prediction about a class of persons. Discriminant validity refers to the ability of those measures or assessments to distinguish between the symptoms or characteristics displayed by that class of persons and persons belonging to another class.<sup>25</sup>

These concepts are particularly cogent in coming to grips with diagnoses of sex abuse. The issue of sex abuse poses major problems for the legal system, and it has caused a bitter and

divisive debate in the discipline of psychology.<sup>26</sup> Many psychologists, psychiatrists, and social workers who work, and testify, in this difficult field are very competent, dedicated professionals. Some, however, are not so good, and the problem is compounded when the expert is influenced by ideology or by expectations about high base rates similar to those made by the health professionals in Rosenhan's study, or when the expert begins as a therapist and is converted to the role of forensic expert.

The concept of convergent validity is pretty easily grasped by the layperson. The sex abuse expert may testify that he or she used various psychological tests, talked to the child's teacher and parents and, in consequence, concluded that all of the sources indicate symptoms consistent with child abuse; that is, the different sources all point in the same direction, they corroborate one another.

Discriminant validity appears to be less intuitive to lay persons - and apparently to many experts.<sup>27</sup> Consider the use of anatomically correct dolls. Many experts have formed diagnoses that are heavily weighted by the fact that children expressed "abnormal" interest in them or put them in positions that the expert interpreted as sexual activity.<sup>28</sup> There are problems with inadequate training some experts have had with these dolls, the lack of standardization of the dolls, and the absence of clear diagnostic guidelines about how to interpret the child's responses. Furthermore, until recently, there has been little evidence on how non-sexually abused children would respond to these tests. An emerging body of data seems to suggest that some non-sexually-abused children also show interest in the dolls. The problem posed by this new body of research is the degree to which expressions of sexual interest can discriminate between non-sexually abused and sexually abused children.<sup>29</sup>

In a similar vein Professor Robert Levy has described how the "child sex abuse accommodation syndrome" has inappropriately been used as forensic evidence.<sup>30</sup> This diagnostic description was developed by a psychiatrist who explicitly stated that it was not a test for sexual abuse, but rather was a device to improve therapy and advocacy for children. Some of the



symptoms of the syndrome might easily be displayed by children who are, for example, suffering the result of a bitter divorce or other stresses, but who are not sexually abused. In other words the diagnostic tool may not discriminate between sexually abused and non-sexually abused children. Nevertheless, Levy points out that evidence based on the syndrome was heavily relied upon in court findings that parents had sexually abused their children.

The problem of discriminant validity also occurs with diagnoses of rape trauma syndrome.<sup>31</sup> Consider as one example, a study by Kilpatrick, Vernon and Resnik that, in the authors' words "attempted to set a higher standard for future research"<sup>32</sup> and has been described by one legal commentator as involving "superior methodology".<sup>33</sup> Kilpatrick *et al.*, contacted a number of rape victims in order to validate a "Fear Survey". They compared the reactions of these victims to a group of non-victims: women chosen from the same neighborhood and matched for age and race. By comparing the responses to the two groups of women the researchers sought to demonstrate that the symptoms of rape victims were different than non-rape victims, that is show that their scale demonstrated "discriminant validity". On the surface the study seems methodologically sound, but when we delve a little deeper, serious problems appear. Two of these relate to the discriminant validity. First, the "control" group of non-rape victims was probably inadequate since it is possible that some of the fears expressed by rape victims can also be expressed by women who have been robbed, beaten by spouses, or are experiencing a stressful divorce or separation. In other words, it is not clear that the "Fear Survey" can discriminate between rape victims and non-rape victims having other life stresses. The problem is particularly acute if an alleged rape victim is also subject to some of these other stresses. Second, the rape victim sample itself combined all types of rape victims: 50 percent had unknown assailants, 19 percent were acquaintance rapes, 22 percent involved a "friend", 3 percent occurred on dates, and 5 percent involved assaults by a relative. There are good reasons to believe that victims of acquaintance rape and incest rape might have different - and possibly much more severe - post-rape symptoms than those who are victims of rape by a stranger. In other words the study may fail to provide adequate discriminant validity for subcategories of rape. Since evidence about rape trauma is often introduced in cases involving acquaintance rape, the thrust of this insight is that

testimony based around Kilpatrick *et al.*'s, scale might be slanted toward a "false negative" finding, that is, a conclusion that there was no rape when in fact it did occur.

There is newer research on rape trauma symptoms that answers some of my criticisms of the Kilpatrick *et al.* study.<sup>34</sup> On the other hand, some of the other measures that might be relied on by experts also suffer from similar problems involving lack of discriminant validity. These examples of rape trauma studies and child sex abuse studies show the utility of the concepts of discriminant and convergent validity in evaluating clinical diagnoses or evidence used for other purposes.<sup>35</sup> These concepts apply to all other forms of diagnosis and prediction.

## **II. DRAWING INFERENCES FROM EXPERIMENTAL STUDIES: INTERNAL AND EXTERNAL VALIDITY**

The concepts of internal and external validity are used to critique research studies.<sup>36</sup> They help us to evaluate whether the studies allow us to draw reasonable inferences from the questions that were asked. Internal validity refers to the ability of the study to eliminate plausible rival hypotheses about causality. If the researcher wants to determine if A causes B, does the design of the study allow her to eliminate the possibility that, instead, factors C or D caused observed changes in B? Even if the study allows us to conclude that A does cause B, external validity involves the ability to generalize the results over persons, situations, and times in a scientifically justified manner. Professor Baar's analysis of the Supreme Court's interpretation of his study of court delay in *Askov* and its scientifically improper generalization of the conclusions to quite different court circumstances in Montreal and elsewhere is an example of a problem with external validity.<sup>37</sup> In that instance the problem of improper inference did not lie with the social scientist but rather with the Supreme Court.

Let me further illustrate internal and external validity problems in the context of the current widespread enthusiasm for mediation as an alternative to litigation.<sup>38</sup> Many legislatures and courts in the United States and Canada have concluded that mediation of "domestic relations" cases (for

example, divorce, child custody, domestic property disputes, and even spousal assault) is superior to adjudication in its fairness, efficiency, and, incidentally, also in its ability to clear troublesome cases from crowded court dockets. In some jurisdictions, mediation for domestic relations cases has even been made mandatory.<sup>39</sup>

A main impetus behind this development has been the enthusiasm of a substantial number of academic social scientists and lawyers who have cited empirical studies purportedly showing the effectiveness of mediation over adjudication. One of the most frequently cited studies is McEwen and Maiman's study of small claims courts in the state of Maine.<sup>40</sup> Those researchers found that cases that were mediated resulted in more satisfaction and compliance than cases that were adjudicated. Now, one can question whether something that works for small claims court cases can be generalized to domestic relations cases. The external validity issue is intuitively obvious in this instance, but more or less ignored by mediation enthusiasts. However, there is a further problem involving questions of internal validity. My own research on an Ontario small claims court indicated that it is very probable that cases selected for, and resolved by, mediation are quite different than those that are selected to be resolved by adjudication.<sup>41</sup> This internal validity problem raises serious doubts about the claimed superiority of mediation over adjudication. Other studies purporting to show the superiority of mediation over adjudication in the specific context of domestic relations cases are plagued with similar internal validity problems.<sup>42</sup> Interestingly, discussion of the methodological shortcomings of some of the mediation studies and citations of the studies that contradict them are usually absent in the pro-mediation literature.

One serious consequence of this failure by courts and legislatures to critically examine the research and assumptions of mediation advocates is that many victims of spouse abuse may be forced into mediation that jeopardizes their rights and their physical safety.<sup>43</sup> I also want to emphasize that the problem exists in Canada. Professor Desmond Ellis has documented the fact that abused women in Ontario sometimes enter mediation under the threat of withdrawal of the

services of their legal aid lawyers.<sup>44</sup> He has also found that battered women were more likely to be abused after separation if they went through mediation rather than adjudication with lawyers.<sup>45</sup>

### III. THE CHANGING STATE OF SOCIAL SCIENCE KNOWLEDGE

Whether the matter involves the physical sciences or the social sciences, one of the primary points of conflict between law and science is the fact that judges, understandably, would prefer a world in which scientific knowledge is stable and consistent. In fact it is not. One recent example involved the controversy over DNA-typing evidence, in which it was shown that scientific opinion about the accuracy of DNA-typing shifted when certain methodological and technical facts about its forensic uses were known.<sup>46</sup> Another involves the Bendectin cases. When the first cases were brought to the courts, many competent medical experts had reason to believe that Bendectin caused birth defects, but subsequent research changed the opinion of many scientists, posing a substantial problem for courts faced with adjudicating the cases.<sup>47</sup>

Social science knowledge is also prone to change. For instance, consider the "battered spouse syndrome" that was ruled admissible evidence in *R. v. Lavallée*.<sup>48</sup> The original tenants underlying this syndrome involved two core phenomena: the escalating cycle of violence and "learned helplessness" on the part of the victim.<sup>49</sup> As it was originally conceived, the latter notion portrayed the victim as becoming increasingly psychologically helpless as the cycle of violence repeats itself over time, that is, while battered women attempt various strategies to stop the violence early in the relationship, these attempts decline gradually towards total passivity. Recent research appears to at least partially contradict these assumptions.<sup>50</sup> For example, in a sample of battered women who sought orders of protection, my colleague, Professor Karla Fischer, found that rather than becoming more passive the women increased their help-seeking and the number of strategies they used to avoid the violence.<sup>51</sup> Other evidence indicates that frequently women are more likely to be subject to violence during or after attempts to leave the relationship than while staying in it. These newer views on spousal violence have direct implications for attempts to provide remedies for spousal violence. They also have direct implications for a judge who has

to rule on the admissibility of "battered women syndrome" evidence under particular fact patterns. The changing insights on the dynamics of spousal abuse, for example, might encourage a judge to expand the scope of conditions under which such syndrome evidence might be admissible.

Social science knowledge about the reliability of child witnesses is another example. It changes monthly as the new issues of psychology research journals appear on our desks. Much of the new knowledge has direct relevance to evaluating experts' testimony on the issues.<sup>52</sup> As a final example, this month, October 1993, the National Research Council in the U.S. will publish a report on child abuse and neglect.<sup>53</sup> Among other findings, the report, authored by a panel of sixteen experts from different disciplines, challenges some of the most widely held beliefs about the causes and consequences of child abuse. In particular the panel concluded that the "intergenerational" explanation that sexually abused children themselves become sexually abusive parents is not scientifically supportable. Existing research studies on the matter lack both internal and external validity.

The theme of this paper, judging social science, has naturally caused me to emphasize the negative side of social science evidence. Consequently, I want to reiterate my view that much social science evidence is good and can withstand methodological scrutiny. Good social science can assist courts in reaching more accurate and fair decisions, whether the application is in the form of forensic evidence or in the design and evaluation of the procedural mechanisms that help the law fulfil its goals. The end result can be enhanced legitimacy of courts in the eyes of the public.<sup>54</sup>

But what is my message for judges? They do not have the time, nor probably the inclination, to get a Ph.D. in a social science discipline in order to critique the evidence. Even if they did, should they also get degrees in biochemistry, medicine, and engineering to evaluate these other types of expert evidence? Of course that is impractical.

I also do not want to leave the impression that I am saying that, absent a Ph.D., judges are incapable of intuiting some of the methodological and conceptual problems that may attend social science evidence or studies, even if they do not have the sophisticated vocabulary of the methodologist. After learning that it is possible to make fine distinctions, I can now conceive of different stages of camel pregnancy. In fact, if someone would loan me a fecundated beast I might be able to cognitively apprehend at least some of the stages. I might even borrow an English-Arabic dictionary to learn some of the words for these stages so as to better understand and communicate with others about them. And if I decided to take a journey with some desert nomads, I would certainly want to know some of the distinctions between camels so as to not end up on an unsuitable mount.

Judges are already on a journey with social science. It is here to stay and they need to understand and evaluate it. Mason, for instance, reviewed appellate court decisions in child sex abuse cases and charged that "courts are unconcerned about scientific knowledge, but are willing to accept the testimony of expert witnesses as long as they have significant clinical experience".<sup>55</sup> This will not suffice - in child abuse cases or in any other cases involving social science evidence. Knowing that there are distinctions and that there is a vocabulary to aid in the task is an important step. The exploration of other steps judges could take would require a much longer paper. In the meantime, I urge judges to think about the many words for scientific reliability and validity.

## FOOTNOTES

1. O. Klineberg, *Social Psychology* (New York: Henry Holt, 1940) at 44-45.
2. For a more extensive survey of applications of social science evidence than covered in this paper see N. Vidmar, "Social Science Evidence and Data" in G.M. Chayko, E.D. Gulliver & D.V. Macdougall, eds., *Forensic Evidence in Canada* (Aurora: Canada Law Book, 1991) 641; J. Monahan & L. Walker, *Social Science in Law*, 2d ed. (New York: Foundation Press, 1990).
3. For a partial review of cases see Vidmar, *supra* note 2.
4. *R. v. McGregor* (1992), (Ont. S.C.) [unreported].
5. *R. v. Oughton* (18 September 1986), (B.C.S.C.) [unreported].
6. *R. v. Kenny* (1991), (Nfld S.C.) [unreported].
7. *Law Society of Upper Canada v. Skapinker* (1984), 11 C.C.C. (3d) 481 at 501-502, 9 D.L.R. (4th) 161 (S.C.C.).
8. *R. v. Oakes* (1986), 24 C.C.C. (3d) 321 at 347-348, 26 D.L.R. (4th) 200 (S.C.C.).
9. For a review, see K. Swinton, "What Do the Courts Want from the Social Sciences?" in R.J. Sharpe, ed., *Charter Litigation* (Toronto: Butterworths, 1987). Professor Swinton has noted three overlapping ways in which social science evidence can be used to interpret the charter: to help prove the denial of a right; to justify the limitation of a right under section 1; and to challenge the reasonableness of the limitations of a right. It may also be used to help fashion a remedy under section 24. See also Baar, *infra* note 13; C. Baar & E. Baar, "Diagnostic Adjudication in Appellate Courts: The Supreme Court of Canada and the Charter of Rights" (1989) 27 Osgoode Hall L.J. 1; *Ford v. Attorney General of Quebec*, [1988] 2 S.C.R. 712, 54 D.L.R. (4th) 577; L.E. Weinrib, "The Morgentaler Judgment: Constitutional Rights, Legislative Intention, and Institutional Design" (1992) 42 U.T.L.J. 22.
10. Faust & Ziskind, "The Expert Witness in Psychology and Psychiatry", (1988) 241 *Science* 31.
11. Ceci & Bruck, "Suggestibility of the Child Witness: A Historical Review and Synthesis" (1993) 113 *Psychological Bulletin* 403; Mason, "A Judicial Dilemma: Expert Witness Testimony in Child Sex Abuse Trials" (Fall-Winter 1991) *Psychiatry and Law* 85; R.J. Levy, "Using 'Scientific' Testimony to Prove Child Sexual Abuse" (1989) 23 *Family L.Q.* 383. Levy and Ceci and Bruck discuss pseudo-scientific claims including the claim that children never lie.

12. *R. v. Askov*, [1990] 2 S.C.R. 1199, 74 D.L.R. (4th) 355.
13. C. Baar, "Court Delay as Social Science Evidence: The Supreme Court of Canada" and "Trial Within a Reasonable Time" (Paper presented at the Annual Meeting of Law and Society Association, Chicago, Illinois, 27-30 May, 1993).
14. See Fischer, Vidmar & Ellis, "The Culture of Battering and the Role of Mediation in Domestic Violence Cases" (1993) 46 SMU L. Rev. 2117.
15. See Klineberg, *supra* note 1 at 42-50.
16. *Ibid.*
17. See E.G. Gerjuoy, "Science and Technology Resources for the Courts" (1991) 14 The Justice System Journal 358; S. Jasanoff, "What Judges Should Know About the Sociology of Science" (1992) 32 Jurimetrics Journal 345; S. Jasanoff & Nelkin, "Science, Technology, and the Limits of Judicial Competence" (1981) 214 Science 1211.
18. I trace this tendency in large part back to the language in the leading case on the admissibility of scientific evidence, *Frye v. the United States*, 293 F. 1013 (D.C. Cir. 1923). Canadian courts have made reference to the *Frye* case: see Vidmar, *supra* note 2 at 671. The problem may also result, in part, from the tendency for judges to equate the reliability of civilian witness testimony with expert testimony.
19. See for example, P.J. Runkel & J.E. McGrath, *Research on Human Behavior* (Montreal: Holt, Rinehart and Winston, 1972); J.M. Carlsmith, P.C. Ellsworth & E. Aronson, *Methods of Research in Social Psychology* (Don Mills: Addison-Wesley, 1976).
20. See Vidmar, *supra* note 2.
21. Rosenhan, "On Being Sane in Insane Places" (1973) 179 Science 250.
22. Of course, the findings from the first Rosenhan study suggest the possibility that in the latter study some of the real patients were sane rather than sick.
23. See Vidmar, *supra* note 2 at 667; Faust & Ziskind, *supra* note 10.
24. See Faust & Ziskind, *supra* note 10. To take a single example, it is well-documented that polygraph experts are prone to conclude that a person is lying when in fact the person is telling the truth; see Vidmar, *supra* note 2 at 667 for references.
25. See for example, Runkel & McGrath, *supra* note 19 at 163-167.
26. See Ceci & Bruck, *supra* note 11.



27. *Ibid.* See also Faust & Ziskind, *supra* note 10.
28. See Ceci & Bruck, *supra* note 11 at 423; Levy, *supra* note 11 at 397.
29. See Ceci & Bruck, *supra* note 11.
30. Levy, *supra* note 11 at 393.
31. Burgess & Holmstrom, "Rape Trauma Syndrome" (1974) 131 *Am. J. Psychiatry* 981; P. Frazier & Borgida, "Juror Common Understanding and the Admissibility of Rape Trauma Syndrome Evidence in Court" (1988) 12 *Law and Human Behavior* 101; K. Fischer, "Defining the Boundaries of Admissible Expert Psychological Testimony on Rape Trauma Syndrome" (1989) *U. Illinois L. Rev.* 691.
32. Kilpatrick, Veronen, & Resick, "Assessments of the Aftermath of Rape: Changing Patterns of Fear" (1979) 1 *Journal of Behavioral Assessment* 113.
33. M.E. Donohue, "Another Closed Door: Rape Trauma Syndrome" (1987/88) 23 *Gonzaga L. Rev.* 1.
34. See for example, Fischer *supra* note 31 for a partial review.
35. Rape trauma syndrome may also be introduced in court as "social framework" evidence that is intended to provide judge or jury a context for interpreting contested adjudicative facts. See Vidmar, *supra* note 2; Monahan & Walker, *supra* note 2; Vidmar & Schuller, "Juries and Expert Evidence: Social Framework Testimony" (1989) 52 *Law and Contemporary Problems* 133.
36. See D. Campbell & J. Stanley *Experimental and Quasi-Experimental Designs for Research* (Chicago: R. McNally, 1963).
37. See Baar, *supra* note 13.
38. See Fischer, Vidmar & Ellis *supra* note 14.
39. *Ibid.*
40. McEwen & Maiman, "Mediation in Small Claims Court: Achieving Compliance Through Consent" (1984) 18 *Law and Society Review* 11. See Fischer *et al.*, *supra* note 14, for documentation.
41. Vidmar, "The Small Claims Court: A Reconceptualization of Disputes and an Empirical Investigation" (1984) 18 *Law and Society Review* 515; Vidmar, "Assessing the Contributions of Case Characteristics and Settlement Forums on Dispute Outcomes and Compliance" (1986) 21 *Law and Society Review* 155.

42. See Fischer, Vidmar & Ellis, *supra* note 14.
43. *Ibid.*
44. Ellis & Stuckless, "Postseparation Abuse, Mantal Conflict Mediation, and Postseparation Abuse" (1992) 9 *Mediation Q.* 205.
45. Ellis, "Post-separation Woman Abuse: The Contribution of Lawyers as 'Barracudas', 'Advocates', and 'Counsellors'" (1987) 10 *Int'l J. L. and Psychiatry* 403.
46. See for example, Jasanoff (1992), *supra* note 17.
47. J. Sanders, "The Bendectin Litigation: A Case Study in the Life Cycle of Mass Torts" (1992) 43 *Hastings L.J.* 301.
48. *R. v. Lavallée* (1988), 65 C.R. (3d) 387 (Man. C.A.) 393 (1990), 55 C.C.C. (3d) 97 (S.C.C.); for a review and critique see R.A. Schuller, "The Impact of Battered Woman Syndrome Testimony on Jury Decision Making: *Lavallee v. R.* Considered". (1990) 10 *Windsor Yearbook of Access to Justice* 105.
49. See Schuller, *supra* note 47.
50. See for example, Fischer, Vidmar & Ellis, *supra* note 14.
51. *Ibid.* at 2133-2137.
52. See for example, Ceci & Bruck, *supra* note 11.
53. See American Psychological Society Observer, "What Does Science Know About Child Abuse?" in C. Tower, ed., *Understanding Child Abuse and Neglect* (Boston, Toronto: Allyn and Bacon, 1993). For example, prior studies are based on retrospective reports that may not be reliable. Additionally, longitudinal studies that permit inference about the percentages of abused children who become adequate parents do not exist.
54. See T.R. Tyler, *Why People Obey the Law* (New Haven: Yale University Press, 1990); Vidmar, "Procedural Justice and Alternative Dispute Resolution" (1992) 3 *Psychological Science*, 224; Vidmar, "The Origins and Consequences of Procedural Fairness" (1990) 15 *Law and Social Inquiry* 877.
55. Mason, *supra* note 11 at 20.