Law and Human Behavior, Vol. 26, No. 2, April 2002 (© 2002)

Rethinking the Probative Value of Evidence: Base Rates, Intuitive Profiling, and the "*Post*diction" of Behavior

Deborah Davis[1,2] and William C. Follette[1]

It is argued that American courts may be routinely admitting evidence with little to no probative value and great potential for prejudicial impact. This may be particularly likely with regard to what is essentially "intuitive profiling" or "stereotype" related evidence, defined herein as evidence suggesting that the defendant (or other party), or his (her) behavior, fits intuitive "profiles" (or stereotypes) of the type of person likely to commit the crime or behavior in question. In other words, "intuitive profiling" evidence is admitted to "postdict" behavior. Formal empirically based "profiling" evidence (testimony regarding the fit of a defendant's characteristics or behaviors to formal or scientific profiles of the typical perpetrator of the crime in question) for use to prove guilt is inadmissible in American courts. However, we suggest that everyday use of informal intuitive profiles underlies both judicial determinations of probative value (diagnosticity), and thus admissibility, of evidence, and jurors' use of the evidence in determining guilt. Demonstrations of the use of base rate information to evaluate the probative value of such intuitive profiling evidence both as evidence of guilt and as evidence of innocence are provided. Demonstrations of both how to evaluate the actual *probative value of evidence (when all necessary values are known), and the* theoretical limit *of its probative value (in circumstances where some values are not known) are provided. It is argued that such evaluations may provide the basis for (1) support of motions to either admit or to exclude evidence, (2) testimony to the jury to help them weigh or interpret evidence, (3) exculpatory profiling (profiling evidence of innocence), (4) pretrial research to establish probative versus prejudicial value of evidence, and (5) sufficiency analyses to determine maximum likelihood of guilt, given multiple items of evidence. Among these, the first two are considered most important, as it can be demonstrated that many "profiling" characteristics currently admitted in trial (such as evidence of battery to support a murder charge) are* not *probative of guilt.*

[1]Department of Psychology, University of Nevada, Reno, Nevada.
[2]To whom correspondence should be addressed at Department of Psychology/296, University of Nevada, Reno, Nevada 89557; e-mail: debdavis@unr.nevada.edu.

Each day in thousands of American courts, what is essentially "*intuitive profiling*" evidence is presented to judges and juries, with the assumption that such evidence is "*probative*" (providing useful diagnostic information) of the likelihood of either past (such as a crime) or future (such as proper parenting or future violence) behavior. A parole board or sentencing judge/jury may hear test results allegedly predictive of future violence by the potential parolee/defendant. Family court judges hear psychological evaluations presumed to predict parenting skills. Criminal juries hear evidence that a defendant is in debt—a situation alleged to represent a motive for embezzlement (and thus to provide evidence that the defendant *did* embezzle) or for murder (and thus to provide evidence that the defendant *did* murder).

In each of these circumstances attorneys will argue, at least implicitly, that the party/defendant in question *has the characteristics* of a person likely to engage in the behavior at issue—whether past (such as criminal activity or suicide) or future (such as future criminal activity and parenting skills)—thereby implying that the party/defendant more likely did (will) perform the behavior. Essentially, attorneys invite judges and juries to rely on the representativeness heuristic (Kahneman & Tversky, 1972, 1973; Tversky & Kahneman, 1974, 1982, 1983) to decide their verdicts. That is, they are arguing "*This person has all the characteristics of a murderer/bad parent/thief, therefore, (s)he must (will) be one!*" It is in this sense that we use the term "*intuitive profiling*." We do not refer to technical legal uses of the term profiling to refer to empirically based profiles, but instead refer to intuitive stereotypes or profiles of the types of persons who tend to commit criminal acts, and/or the types of circumstances under which they tend to be committed.

Essentially, the use of intuitive profiling to assess guilt relies on the following logic: "*If persons who commit embezzlement are likely to be in debt, then persons who are in debt are likely to be embezzlers*," or "*If most A's are B's, then most B's are A's*." Logically, of course, these conclusions are erroneous.

Such assumptions regarding contingent probabilities can also be shown to be false by using simple mathematics. In probability terms, the argument is that if the probability of being in debt is meaningfully higher among embezzlers than among nonembezzlers, then the probability of embezzlement among those who are in debt is higher than among those who are not in debt. Such a conclusion may or may not be correct, as we will demonstrate below. Further, the *degree* of error likely to characterize this assumption can be shown to vary predictably with the base rates within the population of both the behavior and profiling characteristic (pre or "*post*dictor" variable) in question. In particular, the greater the discrepancy between the base rate of the predictor variable (i.e., profiling characteristic) and the behavior to be pre(post)dicted, the greater the likelihood of false positive conclusions (in this case, the false conclusion that the person did or will embezzle).

This point was first addressed by Tversky and Kahneman's (Kahneman & Tversky, 1972, 1973; Tversky & Kahneman, 1974, 1982, 1983) demonstrations of the fallacies of use of the representativeness heuristic (i.e., using stereotyping [profiling] to judge category membership) when ignoring base rate information. Essentially, this research demonstrated that no matter how much a person may fit the stereotype for a particular social category, one cannot accurately judge the likelihood that the person actually falls into that category without knowing the base rate of members of that category in the population.

Similarly, a large body of literature in psychology and medicine has recently addressed the accuracy with which low base rate behaviors, events, outcomes, or diseases can be predicted/diagnosed from individual assessments of various kinds (e.g., Swets, Dawes, & Monahan, 2000). Generally, this literature has shown that for populations with an extremely low base rate of the criterion (whether behavior or disease), even using the most accurate medical tests or most predictive psychological assessments, the rate of false positive predictions far exceeds the rate of true positive predictions. Further, the likelihood of error *if one does use such a predictor to form a conclusion* may be calculated reliably, such that it is possible to identify both behaviors and profiling pre(post)dictors for which error is most likely.

Perhaps the area of behavior most frequently addressed by psychological research of this nature is that of dangerousness (aggression). A growing literature in psychology has addressed the difficulties of prediction of dangerousness in populations where violence occurs at low base rates (e.g., Gardner, Lidz, Mulvey, & Shaw, 1996; Grove & Meehl, 1996; Lidz, Mulvey, & Gardner, 1993; Monahan & Steadman, 1994; Steadman, Mulvey, Monahan, Robbins, Appelbaum, Grisco, et al., 1998), calling into question the usefulness of many such predictions in legal contexts such as parole or commitment hearings (see review of the history of this controversy by Litwack & Schlesinger, 1999).

Similar arguments can be made, however, for a wide variety of evidence—not limited to psychological assessments. In particular, many forms of evidence presented in court are designed to demonstrate that the defendant fits the intuitive "*profile*" or stereotype of a person who would commit the crime in question, thereby "*postdicting*" that the person *did* commit the crime. Such evidence is often *minimally* (if at all) probative of criminal disposition or behavior, but does have great potential for prejudicial impact.

The present paper provides a demonstration of the value of use of available base rate information to evaluate the probative value of any form of "intuitive profiling" evidence regarding criminal behavior. To this end, we will describe analyses recently undertaken pertinent to a murder case for which we were retained as experts. We will begin with a brief synopsis of the case and the purpose for which we were retained, and follow with demonstrations of our analysis of the evidence in question.

SYNOPSIS OF THE CASE

We were retained as expert witnesses in the murder trial of a male defendant charged with the murder of his wife. The wife had drowned as a result of a snowmobile crash. She was driving the snowmobile with her husband on the back, lost control, and plunged into a ditch. She and the defendant were found by passersby. She was found face down in the water and apparently drowned; and he sitting face up but not breathing—also apparently drowned or unconscious. CPR was administered to both on the scene, and he was successfully revived, whereas she was not.

Prosecution argued that the defendant had deliberately drowned his wife once they had fallen into the ditch, and that he may have somehow caused the crash, thereafter faking his own unconsciousness/inability to breathe. Physical evidence of each of these assertions was extraordinarily weak, particularly evidence of whether the victim's death was the result of murder or accident.

Given the weakness in physical evidence, the bulk of the prosecution's case consisted of attempts to prove the defendant murdered his wife through evidence that the defendant fit the intuitive profile of a spouse murderer. That is, the prosecution focused on the existence of *motives* for murder, and, indirectly, lack of moral character in the defendant, both of which comport with the intuitive profile of a murderer. In particular, the prosecution provided evidence that (1) the defendant had purchased a large insurance policy on his wife within the year preceding her death (thus suggesting that he killed her for money), and (2) the defendant had been unfaithful to his wife with a number of women since their marriage (thus suggesting he was lacking in moral character, and further may have wanted to be rid of his wife to free him for pursuit of other women).

We were retained to evaluate the latter suggestion. That is, we were asked to provide testimony regarding the relationship between male infidelity and uxoricide (murder of one's wife), with the intention of helping the jury to understand the significance (or in fact, *non*significance/lack of probative value) of the evidence of the defendant's infidelity. It was this assignment that led to our use of base rate information to demonstrate the lack of probative value of information regarding infidelity for assessment of the likelihood the defendant murdered his wife.

ASSESSING THE PROBATIVE VALUE OF EVIDENCE

In legal terms "*probative*" refers to evidence which, if true, makes a conclusion more likely than if the evidence were not true. In terms of our assignment, male marital infidelity is probative for the conclusion of wife murder (uxoricide) if the rate of spouse murder among unfaithful men is meaningfully greater than the rate of spouse murder among faithful men. Thus, in order to answer the question of whether male infidelity is probative of tendency toward uxoricide, it was necessary to determine the difference between the probability of uxoricide among unfaithful husbands and the probability of uxoricide among faithful husbands.

More often than not, when attempting such a calculation, one or both of the necessary two probabilities are unavailable, and must be estimated using base rate information. Nevertheless, if the base rates are known for both the criminal behavior and evidentiary factor in question, the *theoretical maximum* probative value of the evidentiary factor may be calculated (the minimum being, of course, zero—meaning that no difference in the probability of the criminal behavior exists between those with and without the evidentiary factor). We will illustrate this process using our calculations from the murder case described above.

CALCULATING THE *THEORETICAL MAXIMUM* PROBATIVE VALUE OF EVIDENCE

Our examination of the literature on rates of infidelity among men suggested that the best estimate of the rate of at least one incidence of infidelity among married men is approximately 24–26% (see review by Wiederman, 1997). Further, the FBI uniform crime statistics report a rate of uxoricide per year of approximately 4 per million married men (Uniform Crime Reports for the United States, 1993).

To calculate the theoretical maximum probative value of information regarding infidelity, we made the following assumptions:

1. Probative value was defined as the difference between the probability of murder given infidelity and the probability of murder given no infidelity.
2. The probative value of infidelity would be greatest if all husbands who murdered were unfaithful. This is true because (a) the probability of murder among faithful men is at its theoretical floor when it is zero and (b) the probability of murder among unfaithful men is at its theoretical ceiling when all murderers are unfaithful. Thus, the maximum difference in the two probabilities occurs under these conditions.
3. We assumed a maximum length of marriage of 60 years. Thus, to get the maximum potential lifetime incidence of uxoricide, we multiplied 4 (the yearly rate per million) by 60, for a maximum lifetime rate of 240 uxoricides per million men per lifetime. This is, of course, an overestimate, as men are not married for an average of 60 years.
4. Roughly 26% of men are unfaithful with at least one partner. Thus, roughly 260,000 of one million men have been unfaithful.

Based on these assumptions we calculated the following (see illustration in Table 1):

1. *Probability of murder if unfaithful*: Set at theoretical maximum of

   240 (base rate of murder per 1 million)/
   260,000 (base rate of infidelity per 1 million) = .000923

   Note that if the base rate of the evidentiary factor equals or exceeds that of the base rate of the crime, the theoretical maximum rate of the crime

Table 1. Rates of Uxoricide and Infidelity per Million Men

	Faithful	Unfaithful	Total
Kill wife	0	240	240
Do not kill wife	740,000	259,760	999,760
Total	740,000	260,000	1,000,000

Note. Probative value = *p* Kill given unfaithful (240/260,000) − *p* Kill given faithful (0); Ratio of false to true conclusions of guilt = number of unfaithful men who do not kill wife/number of unfaithful men who kill wife (259,760/240); Percent of false conclusions of guilt = number of unfaithful men who do not kill/number of unfaithful men = (259,760/260,000).

given positive standing on the evidentiary factor is equal to the base rate of the crime. However, if the base rate of the crime exceeds that of the base rate of the evidentiary factor, the theoretical maximum probability of the crime given the evidentiary factor is 1 (i.e., every person characterized by the evidentiary factor would commit the crime).

2. *Probability of murder if faithful*: Set at theoretical floor of 0

   Note that the theoretical floor of zero may be set only in cases where the base rate of the evidentiary factor equals or exceeds that of the crime. If the base rate (per million in this example) of the crime is greater, the theoretical floor must be set as follows:

   $$\text{(Base Rate of Crime} - \text{Base Rate of Evidentiary Factor)}/ \\ (1,000,000 - \text{Base Rate of Predictor)}$$

   In other words, this formula uses base rates to calculate the minimum value of

   The number of cases in which the crime occurs but the evidentiary factor does not/The number of cases where the evidentiary factor does not occur

3. Maximum probative value of infidelity:

   .000923 (theoretical maximum probability of murder if unfaithful) – 0 (theoretical minimum probability of murder if faithful) = .000923

Thus, one can conclude that *at maximum* it is .0923% (less than one tenth of 1%) more likely that an unfaithful man will murder his wife at some point in their marriage than it is that a faithful man will murder his wife. If some of those who murder their wives are faithful, this number will decrease as the number of faithful murderers increases. Thus, based solely on this analysis, one can conclude that infidelity is *not* usefully probative of the likelihood of uxoricide.

Although the probative value of infidelity as calculated here is greater than zero (although miniscule), the meaning of this value must be assessed in light of the false positive rate (erroneous conclusions of murder) that would result from concluding murder on the basis of infidelity. To assess this issue, we may calculate the likelihood of error *if* the jury were to conclude murder on the basis of evidence of infidelity, as follows:

1. *Minimum number of times conclusion incorrect for every time correct*:

   259,760 (number of unfaithful men who *don't* murder—i.e. #incorrect inferences)/240 (number of unfaithful men who *do* murder—i.e., #correct inferences) = 1,082.33 incorrect to 1 correct

   Given the parameters assumed in this example, an inference that a man killed his wife because of evidence that he committed adultery would be *incorrect* 1,082.33 times for every single time it would be correct *at a minimum*.

2. *Percent of times conclusion incorrect*:

$$100 - (240/260{,}000\text{—i.e., percent of unfaithful husbands who do kill})$$
$$= 100 - .0923 = 99.907$$

Thus, *at a minimum* an inference that a man killed his wife based on evidence that he committed adultery would be incorrect 99.907% of the time. That is, such an inference would be *incorrect* far beyond a reasonable doubt (almost *certainly* incorrect).

SUMMARY

Based on the above use of base rate information, one can point to the following conclusions:

1. The fact of infidelity is not probative of whether a man murdered or will murder his wife. In fact, the relative increase in likelihood that an unfaithful man will murder his wife, over the likelihood that a faithful man will murder his wife is so infinitesimal (.0923%) as to be totally insignificant.
2. The minimum percent of times a conclusion of murder drawn from the fact of infidelity would be *incorrect* will be more than 99%.
3. For each time such a conclusion is correct, it will be incorrect (at a minimum) more than 1,000 times.
4. All of these numbers *overestimate* the true probative value and probability of accurate inference of murder from infidelity, because all assumptions underlying our calculations were designed to *maximize* probative value.

In order to facilitate application of these calculations to relationships of other evidence to inferences of crime, we have provided Tables 2–4, illustrating the relationships between base rates of the evidentiary factor and criminal behavior and maximum probative value (Table 2), ratio of number of incorrect inferences to each correct inference of criminal behavior from the evidence (Table 3), and the probability of an incorrect judgment (Table 4). Although these three numbers all reflect the same concepts of utility, relevance, or probative value of the evidence, they express these concepts differently, such that they may vary in understandability to judges or jurors.

One can see that the theoretical maximum utility of a predictor is smallest (and in fact not at all probative) when the base rate of the behavior (crime) is lowest and the base rate of the predictor is highest, and becomes greater as the base rate of the behavior increases and that of the predictor decreases—with maximum value when the base rates are equal. Unfortunately, for many (if not most) of the profiling predictors pertinent in the legal system, the base rate of the predictor far exceeds the base rate of the crime. Thus, the predictor will not be probative—either at all, or sufficiently to outweigh its potential prejudicial value.

Table 2. Maximum Probative Value of Predictor as a Function of Base Rates of Crime and Predictor

Base rate of crime	Base rate of predictor (%)												
	.01	.10	1	5	10	20	30	40	50	60	70	80	90
.01	100	10.0	1.0	.200	.100	.050	.033	.025	.020	.017	.014	.013	.011
.10	*99.9*	100	10.0	02.0	1.0	.500	.333	.250	.200	.166	.142	.125	.111
1	*99.0*	*99.1*	100	20.0	10.0	5.0	3.3	2.5	2.0	1.67	1.4	1.3	1.1
5	*95.0*	*95.1*	*96.0*	100	50.0	25.0	16.7	12.5	10.0	8.3	7.1	6.3	5.6
10	*90.0*	*90.1*	*90.9*	*94.7*	100	50.0	33.3	25.0	20.0	16.7	14.3	12.5	11.1
20	*80.0*	*80.1*	*80.8*	*84.2*	*88.9*	100	66.7	50.0	40.0	33.3	28.6	25.0	22.2
30	*70.0*	*70.1*	*70.7*	*73.7*	*77.8*	*87.5*	100	75.0	60.0	50.0	43.9	37.5	33.3
40	*60.0*	*60.1*	*60.6*	*63.2*	*66.7*	*75.0*	*85.7*	100	80.0	66.7	57.1	50.0	44.4
50	*50.0*	*50.1*	*50.5*	*52.6*	*55.6*	*62.5*	*71.4*	*83.3*	100	83.3	71.4	62.5	55.6
60	*40.0*	*40.0*	*40.4*	*42.1*	*44.4*	*50.0*	*57.1*	*66.7*	*80.0*	100	85.7	75.0	66.7
70	*30.0*	*30.0*	*30.3*	*31.6*	*33.3*	*37.5*	*42.9*	*50.0*	*60.0*	*75.0*	100	87.5	77.8
80	*20.0*	*20.0*	*20.2*	*21.1*	*22.2*	*25.0*	*28.6*	*33.3*	*40.0*	*50.0*	*66.7*	100	88.9
90	*10.0*	*10.0*	*10.1*	*10.5*	*11.1*	*12.5*	*14.3*	*16.7*	*20.0*	*25.0*	*33.3*	*50.0*	100

Note. Values in italic include cases where the base rate of the crime exceeds that of predictor.

Table 3. Minimum Times Incorrect if Conclude Crime to One Time Correct as a Function of Base Rates of Crime and Predictor

Base rate of crime	Base rate of predictor (%)												
	.01	.10	1	5	10	20	30	40	50	60	70	80	90
.01	*0*	9	99	499	999	1999	2999	3999	4999	5999	6999	7999	8999
.10	*0*	*0*	9	49	99	199	299	399	499	599	699	799	899
1	*0*	*0*	*0*	4	9	19	29	39	49	59	69	79	89
5	*0*	*0*	*0*	*0*	1	3	5	7	9	11	13	15	17
10	*0*	*0*	*0*	*0*	*0*	1	2	3	4	5	6	7	8
20	*0*	*0*	*0*	*0*	*0*	*0*	.50	1	1.5	2	2.5	3	3.5
30	*0*	*0*	*0*	*0*	*0*	*0*	*0*	.33	.67	1	1.3	1.7	2
40	*0*	*0*	*0*	*0*	*0*	*0*	*0*	*0*	.25	.50	.75	1	1.25
50	*0*	*0*	*0*	*0*	*0*	*0*	*0*	*0*	*0*	.20	.40	.60	.80
60	*0*	*0*	*0*	*0*	*0*	*0*	*0*	*0*	*0*	*0*	.17	.33	.50
70	*0*	*0*	*0*	*0*	*0*	*0*	*0*	*0*	*0*	*0*	*0*	.14	.29
80	*0*	*0*	*0*	*0*	*0*	*0*	*0*	*0*	*0*	*0*	*0*	*0*	.13
90	*0*	*0*	*0*	*0*	*0*	*0*	*0*	*0*	*0*	*0*	*0*	*0*	*0*

Note. Values in italic include cases where the base rate of the crime equals or exceeds the base rate of the predictor.

Table 4. Minimum Percent Incorrect if Conclude Crime as a Function of Base Rates of Crime and Predictor

Base rate of crime	Base rate of predictor (%)												
	.01	.10	1	5	10	20	30	40	50	60	70	80	90
.01	0	90.0	99.0	99.8	99.9	99.95	99.97	99.98	99.98	99.98	99.99	99.99	99.99
.10	*0*	0	90.0	98.0	99.0	99.50	99.67	99.75	99.80	99.83	99.86	99.88	99.89
1	*0*	*0*	0	80.0	90.0	95.00	96.67	97.50	98.00	98.33	98.57	98.75	98.89
5	*0*	*0*	*0*	0	50.0	75.00	83.33	87.50	90.00	91.67	92.86	93.75	94.44
10	*0*	*0*	*0*	*0*	0	50.00	66.67	75.00	80.00	83.33	85.71	87.50	88.89
20	*0*	*0*	*0*	*0*	*0*	0	33.33	50.00	60.00	66.67	71.43	75.00	77.78
30	*0*	*0*	*0*	*0*	*0*	*0*	0	25.00	40.00	50.00	57.14	62.50	66.67
40	*0*	*0*	*0*	*0*	*0*	*0*	*0*	0	20.00	33.33	42.86	50.00	55.56
50	*0*	*0*	*0*	*0*	*0*	*0*	*0*	*0*	0	16.67	28.57	37.50	44.44
60	*0*	*0*	*0*	*0*	*0*	*0*	*0*	*0*	*0*	0	14.29	25.00	33.33
70	*0*	*0*	*0*	*0*	*0*	*0*	*0*	*0*	*0*	*0*	0	12.50	22.22
80	*0*	*0*	*0*	*0*	*0*	*0*	*0*	*0*	*0*	*0*	*0*	0	11.11
90	*0*	*0*	*0*	*0*	*0*	*0*	*0*	*0*	*0*	*0*	*0*	*0*	0

Note. Values in italic include cases where the base rate of the crime exceeds that of the predictor.

SCIENTIFIC/EMPIRICALLY BASED PROFILES MUST ALSO BE EVALUATED IN THE CONTEXT OF BASE RATES

Before proceeding, we would like to emphasize that even predictors shown empirically to be strongly associated with a crime (i.e., part of an established scientific profile of those who tend to commit the crime) are not necessarily probative of guilt, given that a crime has occurred. For example, among men who commit uxoricide, the most common correlate is a history of spousal battery (e.g., Bixenstine, 1999; Campbell, 1993; Reiss & Roth, 1993; Straus & Gelles, 1990; Websdale, 1999).

Using available estimates of the percent of men who kill their wives who previously battered them, along with the base rates of uxoricide and of battery, we can calculate the probative value of battery as evidence of guilt of uxoricide. Websdale (1999) reported a history of battery among 86.7% of male spouse murderers (an apparently very strong association of battery with uxoricide). The natural tendency is to assume that a history of battery would provide probative evidence of murder. This assumption is clearly false, however.

To illustrate this fallacy, we used Websdale's .86% as the probability of battery among murderers (Websdale, 1999). Estimated base rates of battery range from 10 to 30% (e.g., Davis, Lurigio, & Skogan, 1997), and, as before, the base rate of uxoricide was estimated at 240 per million. Using these numbers, the probative value of battery as evidence of murder is .0638% (at the 30% base rate of battery) and .2022% (at the 10% base rate of battery). In other words, *at maximum* (using the 10% base rate of battery), it is two tenths of 1% more likely that a batterer will murder his wife than that a nonbatterer will murder his. Clearly then, as in this example, even for characteristics known to be strongly associated with a relatively low base rate crime, the probative value of those characteristics as evidence of guilt will be virtually nil.

WHAT IF THE PRIOR PROBABILITY OF GUILT IS KNOWN?

In our example, the issue was both (a) whether death was the result of murder or accident, and (2) if murder, did the defendant commit the murder. If it was known that the woman was murdered, our calculations would arguably have to be different, as it has been shown that 50–80% of murdered women are murdered by their husbands or lovers (e.g., Johnson, 1995). Thus, in the relatively unique circumstance where the probability of guilt for the particular class of defendants is known to be high (husbands in this example), the calculations would have to set the probability of murder at the base rate of guilt (50–80% in this instance), rather than 240 per million, as assumed in our previous illustrations.

With a base rate of guilt of 50%, the probative value of infidelity becomes 67.57% (percent more likely to be guilty given infidelity than no infidelity). With a base rate of guilt of 80%, the probative value of infidelity falls to 27.03%. Thus, ironically, in this case, the higher the prior probability of guilt, the less probative evidence tends to become (as is the case once the base rate of the crime exceeds the base rate of the evidentiary factor [see Table 2]).

It should be noted that although we have illustrated the calculations for the situation where a woman is murdered and uxoricide is suspected, we do not expect that prior probability of guilt is known for most defendants and most crimes. Our use of the uxoricide example is unusual, in that there is data suggesting that one particular person (i.e., the husband or lover) is the convicted murderer in most cases where a woman is murdered. For most crimes, statistics do not point to a particular person in this way. Thus, there would be no reason to adjust our calculations.

Also, with respect to uxoricide, the fact that husbands are convicted in most cases may reflect actual guilt or may reflect unjustified convictions based on the kinds of processes discussed here. Thus, statistics regarding prior probability of guilt are arguably inaccurate. Finally, even if accurate, the use of such probabilities would raise serious issues related to the presumption of innocence.

PROBATIVE OF THE ULTIMATE ISSUE OR PROBATIVE OF A MATERIAL FACT

In our example, we illustrated the calculations for the probative value of evidence of infidelity for the ultimate conclusion of murder. In fact, such evidence could also be admitted as probative of what is deemed to be a "material fact," or a fact deemed to be itself probative of the ultimate conclusion of murder. Infidelity may be considered probative of intention to dissolve the relationship, or to get rid of one's spouse, which in turn would be considered a material fact probative of murder.

We would like to emphasize that base rate analyses can be employed to evaluate probative value of evidence for either another "*material fact*" or for the ultimate conclusion. They may also be used to evaluate the probative value of what is considered to be a material fact. For example, with respect to our uxoricide example, we could show that intention to leave a relationship is not meaningfully probative of uxoricide. Half of marriages end in divorce, as we all know. If half of these are initiated by the male, then 25% of men will intend to dissolve their marriages at some point in the marriage. Given this base rate of 25%, intention to dissolve the marriage is not meaningfully more probative of uxoricide than infidelity at its base rate of 26%.

By way of such analyses, one might show that an allegedly "*material*" fact the party wishes to attempt to prove is actually not "*material*" at all, thus supporting the argument that no evidence offered in support of that demonstrably *im*material fact should be admissible.

SUFFICIENCY EVALUATION THROUGH BASE RATE ANALYSES OF MULTIPLE PREDICTORS

In some circumstances, base rate analyses may be used to inform the ultimate question, rather than to evaluate probative value of specific evidence. Such an analysis may be pertinent, for example, in support of a motion for directed verdict, to prove that combined evidence does not significantly raise the likelihood that the defendant committed a crime—particularly to the level of lack of reasonable doubt. One might be able to prove, for example, that the likelihood a person would commit the crime in question given the entire array of prosecution evidence is only 10% greater than given the lack of all of the evidence.

On the other hand, one might find that although each single predictor or item of evidence may not be probative of guilt (or sufficient proof of guilt) *alone*, they are informative when combined together. In other words, if the defendant has multiple characteristics believed to be associated with the to-be pre(post)dicted behavior, then the combined predictors are probative, or highly suggestive of guilt.

We now turn to an evaluation of these possibilities, and illustration of how the actual or theoretical maximum probability of guilt given the presence of multiple predictors may be calculated. Note, as we engage this problem, that the probative value obtained through these calculations is equal to the maximum likelihood of committing such a crime, given the presence of all predictors. This is because we have set the probability of committing such a crime given the lack of any of the predictors at zero, in order to maximize probative value. Thus, the difference between the maximum likelihood of such a crime given all predictors versus given no predictors will always be equal to the maximum likelihood of the crime given all predictors.

Correlated Versus Uncorrelated Predictors

The marginal utility of additional predictors (and therefore their combined probative value, or the maximum likelihood of guilt given all predictors) will be greatest when the predictors are uncorrelated. To understand this point, consider Tables 5 and 6, which illustrate the calculation of maximum probative value for two perfectly uncorrelated (Table 5) and two perfectly correlated variables (Table 6). For both illustrations, we assume base rates of 25% (or 250,000 per million) for the two predictors, and 400 per million for the crime (a much larger base rate than our previous examples). Further, as in our previous example, to calculate the *maximum* probative

Table 5. Rates of Crime and Two Uncorrelated Predictors per Million Men

	−First predictor	+First predictor	Total
−Second predictor			750,000
Crime	0	0	0
No crime	562,500	187,500	750,000
+Second predictor			250,000
Crime	0	400	400
No crime	187,500	62,100	249,600
Total	750,000	250,000	1,000,000

Note. Probative value of conjunction of two predictors = probability of crime given both predictors − probability of crime given neither predictor = (number who commit crime given both predictors/number of persons with both predictors) − number of persons who *do not* commit the crime given both predictors = (400/62,500) − 0; Ratio of false to true conclusions of guilt = number who *do not* commit the crime given both predictors/number of persons who *do* commit the crime given both predictors = (62,100/400); Percent of false conclusions of guilt = number of innocent persons with both predictors/number of persons with both predictors = 62,100/62,500.

Table 6. Rates of Crime and Two Perfectly Correlated Predictors per Million Men

	−First predictor	+First predictor	Total
−Second predictor			750,000
Crime	0	0	0
No crime	750,000	0	750,000
+Second predictor			250,000
Crime	0	400	400
No crime	0	249,600	249,600
Total	750,000	250,000	1,000,000

Note. Probative value of conjunction of two predictors = probability of crime given both predictors − probability of crime given neither predictor = (number who commit crime given both predictors/number of persons with both predictors) − number of persons who *do not* commit the crime given both predictors = (400/249,600) − 0; Ratio of false to true conclusions of guilt = number who *do not* commit the crime given both predictors/number of persons who *do* commit the crime given both predictors = (249,600/400); Percent of false conclusions of guilt = number of innocent persons with both predictors/number of persons with both predictors = 249,600/250,000.

value/likelihood of guilt, we will assume that all persons who commit the crime are positive on *both* predictors.

Given these assumptions, the following calculations may be made for *uncorrelated* predictors:

1. *The probative value of one positive predictor*:

   (Probability of crime given presence of predictor)
   − (Probability of crime given absence of predictor)
   = (400/250,000) − (0/750,000) = .0016 − 0 = .0016

   Given the assumed parameters, it is .16% (less than two tenths of 1%) more likely that a person characterized by one predictor will commit the crime than one not characterized by the predictor. Stated differently, the maximum probability of committing the crime given one predictor is .0016.

2. *The probative value of two positive predictors*:

   (Probability of crime given presence of two predictors)
   − (Probability of crime given absence of two predictors)
   = (400/62,500) − (0/562,500) = .0064 − 0 = .0064

Given the assumed parameters, it is .64% (less than seven tenths of 1%) more likely that a person characterized by two predictors will commit the crime than one not characterized by the predictors. If we added additional predictors (also with base rates of 25%), the probative values/maximum likelihood of committing the crime would increase to 2.56% (three predictors), 10.25% (four predictors), and 40.96% (five predictors). Thus, the probative value of three combined predictors is virtually nil (less than 3% increment in probability of the crime), but begins to accelerate substantially with additional predictors beyond three. Note, however, that not even with five independent predictors does the probability of the crime given the combined predictors become greater than 50% (which is necessary to meet the standard of "*preponderance of the evidence*," but still clearly not "*beyond a reasonable doubt*"). Further, these numbers represent the probability of committing a crime of the type in question, not the specific crime in question. The latter is necessarily a smaller number.[3]

Determinants of Combined Probative Value

Generally, as with single predictors, the larger the base rates of the predictors and the smaller the base rate of the crime, the smaller the incremental utility of additional predictors.

It is also true, however, that the greater the correlation between predictors, the less incremental utility one will achieve with additional predictors. This is illustrated in Table 6. Again, we assume base rates of 25% for both predictors, and 400 per million for the crime. Given these parameters, the following calculation of the probative value of positive predictors may be made for *perfectly correlated* predictors:

The probative value of two perfectly correlated positive predictors:

(Probability of crime given presence of both predictors)
– (Probability of crime given absence of both predictors)
= (400/250,000) – (0/750,000) = .0016 – 0 = .0016

Note that this value is identical to the value for one predictor. That is, given the assumed parameters, it is .16% (less than two tenths of 1%) more likely that a person characterized by two perfectly correlated predictors will commit the crime than one not characterized by either predictor.

Greater Utility of Multiple Predictors

In practice, of course, predictors will not be perfectly correlated.[4] However, the degree of correlation between predictors will determine (along with their base rates) the additional probative utility each provides beyond the others (as well as the maximum likelihood of the crime given all predictors), such that probative value or maximum probability of guilt is maximized for uncorrelated predictors and minimized for perfectly correlated predictors.

Generally, pre- or postdiction of behavior will be more accurate when based on an index composed of multiple individual predictors related to the criterion. This has been clearly shown in the clinical forensic literature, where actuarial prediction

[3] With six independent predictors, and the same base rates, the number of people who commit the crime becomes greater than the number who have all six predictors. Thus, maximum probative value, or maximum likelihood that the person would commit such a crime becomes 100% (i.e., all persons with all predictors could commit the crime, and the maximum probability of the crime given all predictors therefore becomes 1). In turn, the minimum probability of the crime given the lack of all six predictors can become zero, if it assumed that all those lacking all six predictors do not commit the crime (i.e., that all those remaining who commit the crime have at least one predictor).

[4] However, we suspect that often multiple items of evidence will be highly correlated. For example, given a case of alleged uxoricide, one might anticipate that evidence that the wife had asked for divorce would be correlated with evidence of fighting and perhaps infidelity.

based on multiple indicators (sometimes including clinical tests) has been shown to be superior to single predictors (whether actuarial or clinical; e.g., Grove & Meehl, 1996; Hanson & Thornton, 2000; Monahan & Steadman, 1994; Steadman et al., 2000).

The Importance of Differential Base Rates

A final point must be made before leaving the discussion of incremental utility of additional correlated predictors. That is, the incremental value of additional predictors will be greater when the additional predictor has a low base rate than when it has a higher base rate. This is true because the base rate of the conjunction of multiple predictors cannot be greater than that of the predictor with the smallest base rate. As the base rate of this conjunction falls, the probative value of the combined predictors goes up. Thus, as one adds predictors with smaller base rates, probative value will go up (up to the point where the variables are perfectly correlated). Conversely, adding predictors with greater base rates will add no probative value beyond existing perfectly correlated predictors, and will add substantially less probative value beyond existing less correlated variables.

CONCLUSIONS

Our analysis has shown the following:

1. The probative value of "*intuitive profiling*" evidence is absolutely dependent on the base rates of the predictors (profiling characteristics/behaviors) and the criterion (material fact or crime).
2. A single profiling predictor is more probative the lower its base rate.
3. A single profiling predictor is more probative the higher the base rate of the criterion (material fact or crime).
4. The incremental utility of multiple profiling predictors is greater the lower the base rates of additional predictors.
5. The incremental utility of additional profiling predictors is less the greater the intercorrelation between predictors.
6. If contingent probabilities are known, the precise probative value of a predictor may be calculated. However, if only base rates for the predictor(s) and criterion are available, the theoretical maximum probative value of the predictor(s) or their combination may be calculated, providing a "*best utility estimate*" of the evidence.
7. Generally, the probative value of any predictor with even a moderately high base rate for pre(post)diction of a low base rate criterion will be insufficient to warrant admissibility in court—particularly when there is reason to believe the predictor may have prejudicial impact.
8. Similarly, the theoretical maximum likelihood of committing a particular type of crime, given one or more evidentiary factors, often will be insufficient to prove guilt. Analyses demonstrating this may be used to evaluate the sufficiency of evidence in support of motions for directed verdicts.

Because the illustrations were structured to maximize probative value, in our examples, the probative value of an evidentiary factor(s) was equal to the probability of the crime given the evidentiary factor(s). When this value is near zero, as it will tend to be with low base rate crimes and high base rate evidentiary factors, we submit that the evidence is not probative, is insufficient to prove guilt, and is instead often likely to be highly prejudicial.

9. Similar calculations may be performed for continuous, rather than categorical pre(post)dictor variables.

The presentation in this paper focused on dichotomous evidence (although the techniques can be applied to categorical evidence with any number of categories). However, decision-making research has identified other statistical prediction rules employing more than one predictor, and continuous rather than (or in addition to) categorical variables. Among psychologists, one of the better known in this category is the Violence Risk Appraisal Guide (VRAG) used to predict the risk of violence among criminal patients (Harris, Rice, & Quinsey, 1993; Quinsey, Harris, Rice, & Cormier, 1998; Rice & Harris, 1995).

Swets et al. (2000) provide an excellent discussion of the use of ROC (receiver operating characteristics) curves to evaluate the probative value (although discussed with other terminology) of standing on the statistical prediction equation for pre(post)diction of behavioral or other categorical criteria e.g., prediction of actual violence [a yes–no categorical outcome] from scores on the continuous VRAG measure. The techniques proposed by Swets et al. (2000) may be employed to provide indices comparable to those proposed herein to assess probative value and relative likelihood of correct versus incorrect verdict decisions based on use of evidence of standing on continuous variables.

CAVEAT: THE IMPORTANCE OF ESTABLISHING PROPER BASE RATES

It is obvious that the value of the calculations we recommend is dependent upon the validity of the base rates and contingent probabilities employed. Thus, it is necessary to determine appropriate base rates *for the relevant population*—in this case, the base rate of the crime among the population demographically (and in other relevant respects) similar to the defendant.

Rogers (2000) has recently addressed this point with regard to risk assessment in forensic practice, noting, for example, that the base rate of violent behavior varies notably (from 7.5 to 66.7%) even between subgroups within a single clinical population (Hiday, 1990). Thus, base rates representing averages across an entire population may be misleading, such that predictions of violence among a given clinical subpopulation (e.g., schizophrenics vs. borderlines) based on the use of the average (rather than group specific) base rate may seriously misestimate the likelihood of violence.

The importance of determining precise base rates varies with the intended use of the data. In our murder example, it was clear that evidence of infidelity would not be probative of uxoricide. Thus, we deliberately (in order to diffuse any potential criticism from the prosecution) chose the most liberal base rate of murder (the base rate of murder over a 60-year period of marriage—i.e., 4 per million per year $\times$ 60 years). Further, we did not employ the differential base rate of uxoricide for the defendant's demographic group (White, 30's, middle class). The base rate would have been reduced had we done so, as murder is more common among non-Whites, younger, and lower class persons, and in common law marriages, where there has been previous physical abuse, and when the spouses are separated (Bixenstein, 1999; Block & Block, 1995; Goetting, 1995; Websdale, 1999; Wilson, Daly, & Wright, 1993; Wilson, Johnson, & Daly, 1995)—all factors that would have reduced the pertinent base rate of murder for our client. Thus, our demonstration showed that with even the most liberal base rate assumption, the evidence of infidelity was not probative. Had it been necessary, however, we could have more precisely determined actual base rates among those demographically equivalent to the defendant—which in this instance would have resulted in a much lower index of probative value.

Such precision may be necessary, where it would increase the index of probative value. It is important, for example, when presenting such indices in court to perform one's calculations using the base rates most favorable to the opposing position (while still being accurate), so that the presentation will not be dismissed as inaccurate or irrelevant. Precision may also be important for "*exculpatory*" base rate analyses (as described below).

IMPLICATIONS

Presently, where admissibility of evidence in trial is dependent upon its "*probative value*," the decision is made by the judge, based on unknown subjective criteria. Similarly, when a motive to dismiss is dependent upon the judge's subjective certainty of guilt, the criteria by which (s)he arrives at that judgment are subjective. Strong experimental evidence collected across a wide variety of subject populations suggests, however, that such decisions are frequently compromised by susceptibility to the representativeness heuristic (Kahneman & Tversky, 1972, 1973; Tversky & Kahneman, 1974, 1982, 1983). That is, judges' decisions would be driven partially or wholly by stereotypes concerning the *type of person* likely to commit a particular crime, or the *type of circumstances* likely to produce crime, and so forth, without sufficient consideration of the importance of base rates. Like other professional decision makers as diverse as doctors, stock brokers, astrologers, graphologists, or clinical psychologists (see Gilovich, 1991; Gilovich & Savitsky, 1996, for examples), who fall prey to the representativeness heuristic, judges' will be victims of such bias when ruling on probative value (and thus admissibility) of evidence. That is, judges are certain to fall prey to the tendency to assume that if an item of evidence (whether physical/factual, or personal/situational characteristic) fits their intuitive stereotype or causal theory of those associated with a specific criminal behavior, the evidence is usefully probative of guilt (i.e., that for defendants and situations where such evidence is true, the defendant is more likely to be guilty than if the evidence was not true).

It is important to note that judges are using what is essentially "*intuitive profiling*" to determine relevance (probative value) and therefore admissibility of evidence. This practice is in stark contrast to the inadmissibility of formal profiling evidence in trial (Kirkpatrick, 1998).

As Kirkpatrick notes, there has been prolific growth in the use of empirically based profiling for criminal investigation over the last two decades. Among the more common are profiles of airplane hijackers, car thieves, drug couriers, arsonists, serial killers, abusive spouses, or battering parents. Such empirically based profiles are commonly used to identify persons who should be the subject of special questioning or further investigation. However, they are generally not admissible in court proceedings to prove guilt (Kadish, 1997). When offered to prove guilt, they are thought to be

> too sweeping and over inclusive, and hence potentially misleading to juries and unfairly prejudicial to defendants. For example, if the profile of a drug dealer included characteristics, such as wearing flashy clothes, driving a fast car, and carrying a cellular telephone, such characteristics would also describe a significant number of attorneys. (Kirkpatrick, 1998, p. 255)

Kirkpatrick (1998) provided a case example in which the court reversed the conviction of a teacher accused of "*deviate sexual intercourse*" with a student, based on the "*erroneous*" admission of testimony regarding the behavioral profile of interactions of pedophiles with their victims. The court held as follows:

> Detective Robson testified to what might be described as a "profile" of a non-violent child abuser who is unrelated to the child: physical and psychological "testing" of the child, giving gifts, showing affection, praising, making the child feel comfortable in the abuser's presence, etc. That child abusers use these techniques has no bearing on whether a person who does these things is a child abuser. For example, it is probably accurate to say that the vast majority of persons who abuse children sexually are male. This says little, if anything, however, about whether a particular male defendant has sexually abused a child . . . Given the lack of probative value of Detective Robson's testimony on this point, the danger of unfair prejudice to defendant from the unwarranted inference that, because the defendant engages in acts that sexual child abusers engage in, she, too, is a sexual child abuser is simply too great. It was error to admit this testimony . . . (State v. Hansen, 1987; see also reversals in US v. Small, 1996; US v. Simpson, 1990)[5]

Although not stated in these terms, Kirkpatrick and the court recognize intuitively the problem of base rates, the fact that behaviors or characteristics with high population base rates do not "*post*dict" criminal behavior, and the potential for great prejudicial impact of profiling evidence. Although the problem of prejudicial impact remains equivalent, this understanding is often lacking with respect to evidentiary rulings on individual items of evidence that comport with judges' *intuitive* profiles.

Although not recognized as such, this evidence is still "*profiling*" evidence (although intuitive rather than formal or scientific). The judge assesses relevance and admits the evidence based on his or her intuitive profile of persons or situations associated with the crime, and then the jury uses that evidence in combination with

[5] In line with the court's decision, Melton, Petrila, Poythress, and Slobogin (1997) illustrated the point that depending on the base rates, even a measure with 90% accuracy (profiling tendency toward child sex abuse) may result at a minimum in 68% false positive classifications.

their own intuitive profiles to render a verdict. Both decisions simplistically employ the representativeness heuristic, in the absence of appropriate consideration of base rates and the true diagnostic value of information they are using.[6]

Scientific profiles, although portraying accurate correlations between offender characteristics and specific crimes, will still lack probative value for determination of guilt (for low base rate crimes). In contrast to scientific profiles, however, intuitive profiles can both lack probative value *and* be highly inaccurate, assuming associations between specific behaviors or characteristics and particular crimes that either do not exist or are the reverse of actual associations (see Wrightsman, 2000, chap. 4 on criminal profiling). Thus, evidence intended to address guilt by likening a defendant to a profile or stereotype of those likely to commit the crime in question has great potential for introducing bias and error. Most such evidence is certain to have prejudicial impact, yet will more often than not lack probative value.

In most cases, intuitive profiling evidence will bias the jury toward finding guilt. However, Wrightsman (2000) pointed out that intuitive profiles may include those for both "*criminals*" and "*noncriminals*," such that some defendants may be wrongfully exonerated as a result of demographics, personality, or demeanor inconsistent with profiles of perpetrators of the crime in question.

USES OF EMPIRICAL/MATHEMATICAL ASSESSMENTS OF PROBATIVE VALUE

Given the lack of recognition of everyday use of intuitive profiling for evidentiary decision making among the judiciary, social scientists may be useful in at least the following respects.

Support of Motions to Admit or Exclude "*Intuitive Profiling*" Evidence

The techniques discussed herein can establish the probative value (or lack thereof) for evidence, whether offered *as* a material fact, or *to prove* a material fact. Further, as noted earlier, where base rate analyses successfully demonstrate that a fact assumed to be "*material*" is actually not meaningfully probative, all evidence intended to prove that fact may be deemed inadmissible.

In some cases, such as our murder case, base rate analyses can demonstrate that neither a particular item of evidence nor a presumed "*material fact*" are meaningfully probative. Particularly when combined with evidence of probable prejudicial impact,

[6]Some have recently questioned the extent to which use of the representativeness heuristic leads to erroneous judgments, as implied by Tversky and Kahneman's (1974) focus on the "*base rate fallacy*" (e.g., Gigerenzer, Todd, & the ABC Research Group, 1999), arguing instead that simple heuristics often result in superior accuracy. The present analysis clearly illustrates the conditions under which accuracy will be minimized or maximized. As Table 4 illustrates, the likelihood of error when using the representativeness heuristic to conclude guilt is fully determinable using our suggested base rate analyses. Gigerenzer et al. (1999), however, would disagree, suggesting that in order to evaluate our arguments, we should use real world accuracy (How often are our judgments actually correct in distinguishing the actually guilty from actually innocent?), which they term "performance criteria" rather than Baysian logic (termed "coherence criteria") as the criterion of accuracy.

testimony demonstrating lack of probative value may be effective in securing a ruling of inadmissibility. As a rule of thumb, attorneys considering the use of such testimony are reminded that actual probative value will be minimal when the base rate of the evidentiary factor is higher and that of the crime is low.

In contrast, where the base rate of the evidentiary factor is lower and that of the crime higher, the maximum probative value of the evidentiary factor is higher (see Tables 2–4). In these circumstances, by mathematically/empirically proving the high probative value of the evidence, testimony in evidentiary hearings may support the admission of an intuitive profiling variable that may otherwise be ruled unduly prejudicial.

Testimony to Help the Jury Weigh or Interpret Evidence

The purpose for which we were retained in our example murder trial was to provide a context to help the jury understand the irrelevance of the prosecution's excessive presentation of testimony regarding defendant's infidelity. This evidence was sure to be prejudicial, and the defense attorney hoped to provide proof to the jury that infidelity was not meaningful evidence of propensity toward or motive for murder.

Generally, base rate analyses such as those provided here can be provided by way of expert testimony to help the jury understand that the intuitive profile or stereotype telling them the evidence is probative of guilt is misleading. It remains, however, to demonstrate that such explanations can effectively reduce juror use of the profiling evidence. Our experience in discussions of these ideas—even with intelligent colleagues and students—has been that they understand the math, but nevertheless find it difficult to accept. Further, there is substantial evidence that jurors do not understand and properly utilize probability estimates and other statistical information, even when explained clearly (e.g., Faigman, & Baglioni, 1988; Smith, Penrod, Otto, & Park, 1996). Thus, we anticipate that it will be difficult to reduce or eliminate juror use of nondiagnostic information, particularly when they intuitively believe it is diagnostic. This problem, however, is an important reason for the judiciary to begin to understand and recognize the validity of base rate indices of probative value, and to exclude evidence that is truly not probative, yet carries great potential for prejudicial impact.

Exculpatory Profiling

The potential exists to use base rate analyses to provide evidence that a specific characteristic or behavior *is probative*, but of *innocence* rather than of guilt. This situation would exist for example, where (in contrast to intuitive beliefs) the likelihood of the crime given *presence* of the evidentiary factor is actually lower than the likelihood of the crime given *absence* of the evidentiary factor. Where these contingent probabilities are known (or determinable), base rate analyses might be presented to either support a motion to exclude the evidence (to prevent prejudicial impact) or to show the jury that the evidentiary factor is actually *exculpatory* rather than incriminating.

Further, base rate analyses can be presented to identify factors that mitigate the importance of an evidentiary factor that, when averaged across the general population, is actually probative. For example, an Axis II psychiatric evaluation is associated with violence. However, when mitigating factors are present (such as positive social relations, self-esteem, religious beliefs, and parents' acceptance of the patient [Plutchik, 1995]), such a diagnosis is much less associated with violence. In other words, base rates of violence in the psychiatric population are much lower, given the presence of mitigating factors, and in such a situation the diagnosis is much less probative of violence.

Pretrial Research to Assess Probative Versus Prejudicial Value of Evidence

In some cases the trial team may anticipate (1) that a particular item of evidence will be damaging and (2) that the opposition will oppose a motion to exclude the evidence. In such circumstances, it may be possible to empirically establish both the maximum or actual probative value of the evidence, *and* its probable prejudicial impact.

If the necessary base rates and/or conditional probabilities can be found in available literature, the maximum and/or actual probative value of the evidence would be assessed using the calculations described in this paper. For some crimes, existing research has already demonstrated prejudicial impact of specific evidence. For example, substantial research demonstrations are available regarding factors that affect verdicts in rape cases (e.g., Olsen-Fulero & Fulero, 1997).

More typically, probable prejudicial impact of the evidence would be assessed through mock jury research. Two versions of the case presentation would be presented to two representative groups of mock jurors—one in which the evidentiary factor is presented, and one in which it is not. Comparison of the rate of conviction among the group exposed to the evidentiary factor with that of the group *not* exposed to the evidentiary factor would provide the index of probable prejudicial impact.

Imagine, for example, that in our murder case we had been able to show (as we did) that the likelihood of spouse murder among unfaithful men is less than one tenth of 1% greater than among faithful men. Thus, the fact of infidelity is not usefully probative of guilt. However, imagine in addition that we had conducted a mock jury involving 100 jurors, with two very complete presentations of the case (one with and one without the evidence of infidelity), and that we had found a rate of 40% guilty votes among the 50 jurors not seeing the evidence of infidelity and a rate of 70% guilty votes among the 50 jurors who did see it. This contrast—of less than one tenth of 1% *actual* greater likelihood of guilt among unfaithful men, as opposed to a 30% greater likelihood of guilty votes among those who hear of the infidelity—clearly indicates that the prejudicial impact of the evidence is greater than its probative value.

Sufficiency Evaluations

As illustrated by our analyses of multiple predictors, base rate analyses may contribute to sufficiency evaluations relevant, for example, to motions for a directed verdict. In such a case, the attorney would want to argue that the sum total of the evidence presented by the opposition is not sufficient to prove the case, even in the absence of an opposing presentation. Our calculations of the combined probative value of multiple predictors (equivalent in our examples to the probability of committing such a crime given the combined evidence) are illustrative of the manner in which base rate analyses can be presented for sufficiency evaluations.

Imagine, for example, that the prosecution has presented evidence against a male defendant accused of killing his wife including motives of infidelity and insurance benefits (as in our case), and an eyewitness who saw a man with hair color and approximate height matching the defendant walking away from the site of the murder within the appropriate time frame. The base rates of each of these evidentiary factors—infidelity, male-initiated relationship dissolution, insurance benefits, height, and hair color—are all available. The theoretical maximum probative value of this evidence can thus be calculated. Given the low base rate of uxoricide, this combined evidence would demonstrably fall short of proof of murder beyond a reasonable doubt. Thus, given prosecution reliance solely on the listed evidence, the defense could justifiably argue for dismissal.

WHAT DOES IT MEAN TO BE "PROBATIVE"?

Our demonstrations and suggestions have highlighted the importance of the question of "*degree*" of probative value, when deciding whether to admit evidence. As is the case in many decision contexts, however, one cannot meaningfully determine the predictive or diagnostic value of a predictive assessment without consideration of *errors* (see Swets et al., 2000). When determining the value of a medical test, for example, one must consider *both* the percent of true cases of disease identified by the test (or in the legal context, the percent of correct determinations of guilt based on the evidentiary factor), *and* the percent of *false positive diagnoses* (or findings of guilt when the party is innocent). Further, one must consider the *importance and implications* of the rate of false positives to true positives.

In the medical context, one might consider the relative benefits of detecting true cases of AIDS, versus the costs of falsely informing a patient that (s)he is infected. In legal contexts, one must weigh the benefits of convicting those who actually commit the crime against the costs of false conviction. To some extent, this is an issue of values. However, in determining the level of probative value at which evidence should be admissible, it must be recognized that when a lenient criterion is used, the rate of false guilty conclusions will increase.

Traditionally, determination of probative value has been intuitive, as has determination of prejudicial value. In neither case has the probable ratio of true to false conclusions been formally addressed. Our analysis provides the means to bring objectivity to the process.

Just as the courts have come to recognize the conditions under which eyewitness testimony has limitations, it must take notice of the fact that human beings do not intuitively understand the importance of base rate information. Judges often make the decision about the admissibility of evidence by saying "*that is a matter for* *the jury to weigh*." The point of this paper is that without instruction, juries cannot properly understand the information value in some kinds of evidence often admitted into testimony. If judges take it upon themselves to understand the issues, much of this (mis)information will never be presented. If it is, it should only be with the proper context of the extremely high risk of false inferences. Even then, it is difficult for juries to put aside their propensity to use the representativeness heuristic improperly.

There is no doubt that it will be difficult for courts to adapt to the ideas presented here. Our analysis started in the context of a murder trial where motive, means, and opportunity are standard issue evidence presented by the prosecution. The fact is that many of the traditional offerings of motive would have virtually no actual probative value, but be highly prejudicial—in that motive would be perceived by the jury as evidence of guilt. The number of divorces, arguments, insurance beneficiaries, and even threats far, far exceed the number of murders. Many of these postdictors are not only highly correlated with one another, but are often *perfectly* correlated and thus do not increase the actual strength of evidence of a crime. For example, if a couple were in the middle of divorce, were fighting, and even threatening economic or physical harm to one another, these facts are hardly three independent pieces of evidence of motive, as divorce generally entails fights and threats of various sorts. It is much easier to imagine a constellation of all these behaviors than one in isolation.

CONCLUSIONS

We have demonstrated that the probative value of evidence can be mathematically/empirically established—subject to the availability and reliability of pertinent base rates or contingent probabilities or both in the appropriate population. Further, we have suggested that the objectivity/fairness of the trial process would be improved by use of these techniques to reduce bias in evidentiary rulings and jury decision making.

We realize that there will be great resistance in the judiciary to the presentation or acceptance of such analyses to support motions to exclude (or include) evidence, or to dismiss on the basis of insufficient evidence—both as a result of misunderstanding of statistical indices of probative value or sufficiency, and of intuitive disbelief of their implications. (Ironically, the judge in our murder case refused to allow testimony to the jury—either on the base rate of infidelity among husbands, or on the implications of this and the base rate of uxoricide for the probative value of evidence of infidelity—on the grounds that *the testimony would be prejudicial*). However, it is our hope that just as forensic/clinical practitioners have become increasingly aware of the pitfalls of clinical predictors of low base rate behaviors (and increasingly willing to adopt alternative predictive strategies—e.g., Grove & Meehl, 1996; Hanson & Thornton, 2000; Monahan & Steadman, 1994; Steadman et al., 2000), the judicial system will ultimately understand and become increasingly accepting of empirical/statistical techniques as aids for evidentiary rulings.

ACKNOWLEDGMENTS

The authors would like to thank the editor, two anonymous reviewers, Dr. William O' Donohue and Dr. Elizabeth Loftus for their comments on previous drafts of this article.

REFERENCES

Bixenstine, V. E. (1999). Spousal homicide. In H. V. Hall (Ed.), *Lethal violence: A sourcebook on fatal domestic, acquaintance and stranger violence* (pp. 231–257). New York: CRC Press.

Block, C., & Block, R. (1995). Trends, risks, and interventions in lethal violence. *Proceedings of the third annual spring symposium of the Homicide Research Working Group*, Atlanta, GA.

Campbell, J. C. (1993). The Danger Assessment Instrument: Risk factors of homicide of mate by battered women. In C. Block & R. L. Block (Eds.), *Questions and answers in lethal and non-lethal violence*. Proceedings of the first annual workshop of the Homicide Research Working Group. Washington, DC: U.S. Department of Justice, National Institute of Justice Research Report.

Davis, R. C., Lurigio, A. J., & Skogan, W. G. (1997). *Victims of crime* (2nd ed.). Thousand Oaks: Sage.

Faigman, D. L., & Baglioni, A. J. (1988). Bayes' theorem in the trial process: Instructing jurors on the value of statistical evidence. *Law and Human Behavior*, *12,* 1–17.

Gardner, W., Lidz, C., Mulvey, E., & Shaw, E. (1996). A comparison of actuarial methods for identifying repetitively violent patients with mental illness. *Law and Human Behavior*, *20,* 35–48.

Gigerenzer, G., Todd, P. M., & the ABC Research Group. (1999). *Simple heuristics that make us smart.* New York: Oxford University Press.

Gilovich, T. (1991). *How we know what isn't so: The fallibility of human reason in everyday life*. New York: The Free Press.

Gilovich, T., & Savitsky, K. (1996). Like goes with like: The role of representativeness in erroneous and pseudoscientific beliefs. *Skeptical Inquirer*, March/April, 34–40.

Goetting, A. (1995). *Homicide in families.* New York: Springer.

Grove, W., & Meehl, P. (1996). Comparative efficiency of informal (subjective, impressionistic) and formal (mechanical, algorithmic) prediction procedures: The clinical–statistical controversy. *Psychology, Public Policy, and Law*, *2,* 293–323.

Hanson, R. K., & Thornton, D. (2000). Improving risk assessments for sex offenders: A comparison of three actuarial scales. *Law and Human Behavior*, *24,* 119–136.

Harris, D. H., Rice, M. E., & Quinsey, V. L. (1993). Violent recidivism of mentally disordered offenders: The development of a statistical prediction instrument. *Criminal Justice and Behavior*, *20,* 315–335.

Hiday, V. A. (1990). Dangerousness of civil commitment candidates: A six-month follow-up. *Law and Human Behavior*, *14,* 551–567.

Johnson, H. (1995). Risk factors associated with non-lethal violence against women by marital partners. In C. Block & R. Block (Eds.), *Trends, risks, and interventions in lethal violence*. Proceedings of the third annual spring symposium of the Homicide Research Working Group (pp. 151–168), Atlanta, GA.

Kadish, M. J. (1997). The drug courier profile: In planes, trains, and automobiles, and now in the jury box. *The American University Law Review*, *46,* 747.

Kahneman, D., & Tversky, A. (1972). Subjective probability: A judgment of representativeness. *Cognitive Psychology*, *3,* 430–454.

Kahneman, D., & Tversky, A. (1973). On the psychology of prediction. *Psychological Review*, *80,* 237–251.

Kirkpatrick, L. C. (1998). Profile and syndrome evidence: Its use and admissibility in criminal prosecutions. *Security Journal*, *11,* 255–257.

Lidz, C., Mulvey, E., & Gardner, W. (1993). The accuracy of predictions of violence to others. *JAMA*, *269,* 1007–1011.

Litwack. T. R., & Schlesinger, L. B. (1999). Dangerousness risk assessments: Research, legal, and clinical considerations. In A. K. Hess & I. B. Weiner (Eds.), *The handbook of forensic psychology* (pp. 171–217). New York: Wiley.

Melton, G. B., Petrila, J., Poythress, N. G., & Slobogin, C. (1997). *Psychological evaluations for the courts: A handbook for mental health professionals and lawyers* (2nd ed.). New York: Guilford Press.

Monahan, J., & Steadman, H. (Eds.) (1994). *Violence and mental disorder: Developments in risk assessment*. Chicago: University of Chicago Press.

Olsen-Fulero, L., & Fulero, S. M. (1997). Commonsense rape judgments: An empathy-complexity theory of rape juror story making. *Psychology, Public Policy, and Law*, *3,* 402–427.
Plutchik, R. (1995). Outward and inward directed aggressiveness: The interaction between violence and suicidality. *Pharmacopsychiatry*, *28*(Suppl.), 47–57.
Quinsey, V., Harris, G., Rice, M., & Cormier, C. (1998). *Violent offenders: Appraising and managing risk.* Washington, DC: American Psychological Association.
Reiss, A. J., Jr., & Roth, J. A. (1993). *Understanding and preventing violence.* Washington, DC: National Academy Press.
Rice, M. E., & Harris, G. T. (1995). Violent recidivism: Assessing predictive validity. *Journal of Consulting and Clinical Psychology*, *63,* 737–748.
Rogers, R. (2000). The uncritical acceptance of risk assessment in forensic practice. *Law and Human Behavior*, *24,* 595–605.
Smith, B. C., Penrod, S. D., Otto, A. L., & Park. R. C. (1996). Jurors' use of probabilistic evidence. *Law and Human Behavior*, *20,* 49–82.
State v. Hansen, 304 Or. 169, 176, 743 P.2d 157 (1987).
Strauss, M. A., & Gelles, R. J. (1990). *Physical violence in American families.* New Brunswick, NJ: Transaction.
Steadman, H., Mulvey, E., Monahan, J., Robbins, P., Appelbaum, P., Grisso, T., et al. (1998). Violence by people discharged from acute psychiatric inpatient facilities and by others in the same neighborhoods. *Archives of General Psychiatry*, *55,* 393–401.
Steadman, H. J., Silver, E., Monahan, J., Appelbaum, P. S., Robbins, P. C., Mulvey, E. P., et al. (2000). A classification tree approach to the development of actuarial violence risk assessment tools. *Law and Human Behavior*, *24,* 83–100.
Swets, J. A., Dawes, R. M., & Monahan, J. (2000). Psychological science can improve diagnostic decisions. *Psychological Science in the Public Interest*, *1,* 1–26.
Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. *Science*, *185,* 1124–1131.
Tversky, A., & Kahneman, D. (1982). Judgments of and by representativeness. In D. Kahneman, P. Slovic, & A. Tversky (Eds.), *Judgment under uncertainty: Heuristics and biases* (pp. 82–98). Cambridge: Cambridge University Press.
Tversky, A., & Kahneman, D. (1983). Extensional versus intuitive reasoning: The conjunction fallacy in probability judgment. *Psychological Review*, *90,* 293–315.
Uniform Crime Reports for the United States (1993). Federal Bureau of Investigation, U.S. Department of Justice, Washington, DC: U.S. Government Printing Office.
US v. Small, 74 F.3d 1276, 1283 (D.CC. Cir. 1996).
US v. Simpson, 910 F.2d 154, 156–158 (4th Cir. 1990)
Websdale, N. (1999). *Understanding domestic homicide.* Boston: Northeastern University Press.
Wiederman, M. W. (1997). Extramarital sex: Prevalence and correlates in a national survey. *Journal of Sex Research*, *34,* 167–174.
Wilson, M., Daly, M., & Wright, C. (1993). Uxoricide in Canada: Demographic risk patterns. *Canadian Journal of Criminology*, *35,* 263–291.
Wilson, M., Johnson, H., & Daly, M. (1995). Lethal and nonlethal violence against wives. *Canadian Journal of Criminology*, *37,* 331–361.
Wrightsman, L. S. (2000). *Forensic psychology*. Belmont, CA: Wadsworth.