WHY JURIES CAN’T FOLLOW INSTRUCTIONS

By

Adam Gifford, Jr.* and Melanie Stallings Williams**

I. INTRODUCTION

There are a myriad of studies analyzing the extent to which juries fail. Given the cautions to withhold judgment until the conclusion of the case, followed by jury instructions on how to proceed based on their factual findings, practitioners and researchers have bemoaned the effect of other factors on jury deliberations. Empirical studies have analyzed the effect of the parties’, the jury’s and the case’s characteristics and concluded that all of these factors have impacts on juries’ outcomes despite judicial admonitions to the contrary. Not surprisingly, studies show that juries typically reach a conclusion about the case long before the jury deliberations begin. But it does not follow that because jurors cannot follow instructions the legal system has failed. We argue that it is not only impossible but undesirable for jurors to withhold conclusions until the end of a trial.

Analyzing the problem requires understanding that there are biological, physical and social constraints that prevent jurors from following the standard instructions, i.e. to refrain from forming opinions about a defendant’s guilt or liability until all of the evidence has been presented. In part the argument is based upon an understanding of the complexity of the problem. First, there is the difficulty of accurately perceiving and understanding the meaning of the evidence presented. Second is the problem of combining the evidence in an optimal manner.

* Professor of Economics, California State University, Northridge. Ph.D., University of California, San Diego, 1972; A.B. San Diego State University, 1969.
Third, there are severe memory constraints that make an “after all the evidence has been heard” combinatorial process difficult if not impossible.

II. JURORS DO NOT WITHHOLD JUDGMENT UNTIL THE CONCLUSION OF A TRIAL, NOR DO THEY LIMIT THEIR CONCLUSIONS TO AN ANALYSIS OF ADMITTED EVIDENCE

Despite judicial instructions to the contrary, jurors do not confine their analysis of a case to the evidence admitted in court, nor do they withhold their judgments, as instructed, until the conclusion of a case. It is a cardinal belief in the legal system that jurors must withhold decisions until a case is submitted for judgment. Entire procedural systems, for example allowing defendants to have final closing arguments (presumably more influential by coming last), hinge on this belief. The very selection of jurors to eliminate those whose life experiences might bias their analysis is in line with the idea that jurors must come to the process a blank slate, free from any preconceived notions of legal culpability. The judicial system is predicated on the notion that fundamental fairness can be achieved only by jurors who draw no conclusions until they reach the jury room.

However, studies indicate that this ideal is not achieved: jurors’ personal characteristics (e.g. sex, race, education, socioeconomic status, etc.) effect their analysis as do the

---

** Professor & Chair, Dept. of Business Law, California State University, Northridge. J.D., Boston University School of Law, 1981; B.A. University of California, Santa Cruz, 1977.

1 Dennis J. Devine, et al., Jury Decision Making: 45 Years of Empirical Research on Deliberating Groups, 7 PSYCH. PUB. POL. AND L 622, Table 2 at 632-638, (2001) (summarizing the research of J.P. Lipton, Racism in the Jury Box: The Hispanic Defendant, 5 HISPANIC J. OF BEHAV. SCI. 275 (1983). (Empirical research finding that men were more lenient on a predominantly female jury, and that women were harsher on a predominantly male jury.)

2 Id.

3 In a study of complex litigation, researchers found a positive correlation between a juror’s ability to understand scientific evidence and the juror’s occupation and education. Neil Vidmar, The Performance of the American Civil Jury: An Empirical Perspective, 40 ARIZ. L. REV. 849,
characteristics of the parties (for example a party’s age, whether a victim was randomly chosen, whether there are multiple plaintiffs, etc.). The identity of a defendant has an effect on the outcome, with corporations more likely to be held liable than an individual defendant. Awards against wealthy defendants were higher when they were engaged in commercial activity rather than personal business. Differences in juror and defendant prestige scores increased the likelihood of convictions. In addition to juror, participant and case characteristics, a juror’s perceptive and decision-making style has an effect on the jury outcome, with “authoritarian” jurors more likely to vote to convict in criminal trials. Similarly “cynical” individuals make for more conviction-oriented results. “Egalitarian” jurors, by contrast, were more likely to find a criminal defendant non-guilty. Jurors’ emotional responses have an impact on decisions.


4 Devine, supra n.1, summarizing the work of F. Adler, *Socioeconomic Factors Influencing Jury Verdicts*, N.Y. UNIV. R. OF LAW & SOCIAL CHANGE 3 (1973). (“Juries that convicted had higher average prestige scores (based on socioeconomic status). Discrepancy between defendant prestige and average juror prestige positively related to conviction regardless of whether defendant or jury higher.”)


6 Id. at 357.

7 Irwin A. Horowitz, et al., *Effects of Trial Complexity on Decision Making*, 81 J. OF APPLIED PSYCHOLOGY 757 (1996) (An increase in the number of plaintiffs reduces the likelihood that a defendant will be found blameworthy.)

8 Vidmar, supra n.3 at 872.

9 Vidmar, supra n.3 at 872.

10 Devine, supra n.4.


12 Devine supra n.1 at 677.

13

again, had an effect on outcome, with men more likely to find defendants guilty,\textsuperscript{15} to be effected by expert testimony,\textsuperscript{16} to give greater value to eyewitness testimony\textsuperscript{17} and to be persuaded in the jury deliberation process.\textsuperscript{18} Jurors with prior jury service were more likely both to vote for conviction\textsuperscript{19} and to be able to withhold judgment until the conclusion of a case.\textsuperscript{20} A myriad of other factors, beyond the scope of this paper, have been exhaustively researched and established in empirical studies.\textsuperscript{21}

All of the factors, predictably enough, demonstrate that outcomes are effected long before jury deliberations begin. There is a high likelihood that a jury’s decision will correlate to its initial position.\textsuperscript{22} Indeed, studies show that most jurors have reached a conclusion during the case presentation.\textsuperscript{23} Ninety percent of jury verdicts are the same as the jury’s first vote.\textsuperscript{24}

\textsuperscript{15} Jacqueline Pope & Robert Meyer, \textit{An Attributional Analysis of Jurors’ Judgments in a Criminal Case: A Preliminary Investigation}, 563 Social Behavior and Personality 563, 570 (1999) (Men were more likely than women to find a defendant guilty.) \textit{But see} S. Nagel and L. Weitzman, \textit{Sex and the Unbiased Jury}, 56 JUDICATURE 108 (1972) (Gender composition of jury had no effect on verdicts, but male-dominated juries gave higher monetary awards to male plaintiffs and predominantly female juries gave higher awards to female plaintiffs.)


\textsuperscript{17} \textit{Id.} See also Pope & Meyer, \textit{supra} n.15 at 570 (Men view eyewitnesses as much more credible than do women jurors.)

\textsuperscript{18} C.J. Mills & W. E. Bohannon, \textit{Juror Characteristics: To What Extent are They Related to Jury Verdicts?} 64 JUDICATURE 23 (1980) (Male jurors more likely to report that they changed their verdict during deliberations.)

\textsuperscript{19} \textit{See} Devine, \textit{supra} n.1 at 678.


\textsuperscript{21} For a comprehensive review of the literature, see Devine, \textit{supra} n.1.

\textsuperscript{22} Dean Pavlos, \textit{Legal Analysis or Social Psychology?: An Evaluation of the Supreme Court’s Use of Social Psychology in Setting Standards Regarding the Composition of Criminal Juries}, 23 RUTGERS L. REC. 7 (1999).

\textsuperscript{23} Pope & Meyer, \textit{supra} n.15 at 569 (In a study studying mock jurors based on the juror’s attributional style, researchers found that regardless of style, all “reported that they were certain of the defendant’s guilt or innocence by the middle of the trial presentation.”) \textit{See also} Litchman, \textit{supra} n.20.
These findings – that jurors are effected by matters beyond the admitted evidence and that they often reach conclusions prior to the culmination of the case - are typically analyzed as illustrations of failure of the legal system. That is, if jurors’ decisions vary depending on the characteristics of the case, of the trial participants, and of the juror’s own profile and if decisions are largely made before the deliberation process is scheduled to begin, then a legal system predicated on unbiased jurors withholding decisions has not been achieved. But is this aspiration realistic, or even desirable? Proponents of the “story” model of jury behavior hold that decision making is constructive process, and that jurors do not and cannot consider all of the data bearing on a case. Proponents of this model posit that jurors process evidence in the larger context of their own experience and develop alternative interpretations (or “stories”) for the events in dispute, selecting the version that is the most logical given the juror’s framework of experience. Another theory is based on statistical analysis of jury behavior, using models for analyzing the jury decision-making process. The statistical analysis approach uses the assumption that processing information is a statistical process, and that reasoning skills are key to a juror’s success. Thus, a juror’s ability to process statistical information and their ability to understand, analyze and reach logical conclusions based upon factual information determines the juror’s

26 Vidmar supra n.3. See also Devine, supra n.1 at 625, “(T)he most widely adopted approach to juror decision making is the ‘story’ model, wherein jurors attempt to assemble the evidence into a coherent whole that is consistent with the facts of the case and makes sense given their existing knowledge.” discussing R. HASTIE (ed.) INSIDE THE JUROR: THE PSYCHOLOGY OF JUROR DECISION MAKING (1993).
success. Unfortunately, studies find most jurors deficient in these skills, at least in part.\textsuperscript{27} Both the story model and the statistical processing theories discard the notion that jurors are human data receptors, simply receiving information, discarding inadmissible evidence and withholding judgment until the conclusion of a case.\textsuperscript{28} Rather, it is more useful to think of perception itself is a constructive process. We argue that both the story model and the statistical analysis theory are useful models in analyzing juror behavior, but that they must be used in the context of understanding how people process information. Using this approach, we reach the conclusion that it is neither possible nor desirable for jurors to follow instructions.

To accurately assess the importance of evidence a juror must first understand the meaning of the speech act in which that evidence is communicated. The next two sections will examine this process. Section III of the paper will examine the perceptual and cognitive constraints on the determination of meaning. Importantly, perception and cognition are statistical processes, in fact all that we perceive is the statistical structure of the world. The determination of meaning, then, is also a statistical process, Section IV will examine the difficulties involved in consciously determining the meaning of statistical evidence.

III. PERCEPTION AND MEANING

There is a growing understanding that although on the surface it seems that the perception and the understanding of events, speech acts and objects in our environment are easy problems;

\textsuperscript{27} Vidmar, \textit{supra} n.3 at 851.
in fact, the discovery of meaning is a complex process. At the most basic level, by themselves, individual perceptions are radically underspecified—for example, there is not enough information provided by the two-dimensional image of an object projected onto the retina, by itself, for the brain to determine the three-dimensional shape of that object.\footnote{See Zhiyong Yang & Dale Purves. \textit{A Statistical Explanation of Visual Space}, 6 \textit{Nature Neuroscience} 632-640 (2003) and S. Pinker, \textit{How the Mind Works} (1997). See also Donald D. Hoffman, \textit{Visual Intelligence: How We Create What We See} (1998) and Roger Shepard et al, \textit{Special Issue on the Work of Roger Shepard}, 24\textit{Behavioral and Brain Sciences}, 579-791. (2001) for more extensive discussions of the topic.} The brain is able to form three-dimensional representations of objects in the world because it makes use of built-in information and constraints to produce various perceptions and to allow for an unambiguous determination of what is out there most of the time. These built-in mechanisms are in part the product of the evolutionary imprint of the statistical structure of the physical world on the brain. Various optical illusions can be used to illustrate the workings and failings of some of these mechanisms. At a level more relevant to our purposes is the more general problem which over the last several decades has been addressed by the philosopher John Searle and others: the problem of the understanding of intentionality.\footnote{John R. Searle, \textit{The Rediscovery of the Mind} (1992); John R. Searle, \textit{The Construction of Social Reality} (1995); John R. Searle, \textit{Mind, Language and Society} (1998); John R. Searle, \textit{Rationality in Action} (2001) and Ray Jackendoff, \textit{Foundations of Language: Brain, Meaning, Grammar, Evolution} (2002).} Intentionality, used here in the philosophical context, refers to the fact that “…mental states are directed at or about, or of, or refer to or aim at, states of affairs in the world.”\footnote{See John R. Searle, \textit{The Construction of Social Reality} (1995), p. 25.} Individually, the meaning of a given intentional state is radically underdetermined. Searle calls the phylogenetic and ontogenetic evolutionary mechanisms that, among other properties, facilitate identification and understanding of intentional states background. “Intentional phenomena such as meanings, understandings,
interpretations, beliefs, desires, and experiences only function within a set of Background capacities that are not themselves intentional.”32 The meaning of any representation cannot be determined in isolation, but only within the context of a set of background capacities.33 The problem of underdetermination occurs with all representative states for human as well as nonhuman animals, but for the purposes of understanding the problem facing jurors the focus here will be on the underdetermined nature of language. Furthermore, “[t]he simplest argument for the thesis of Background is that the literal meaning of any sentence can only determine its truth conditions or other conditions of satisfaction against a Background…. Think, for example, of the occurrence of the word ‘cut,’ in sentences such as ‘Sally cut the cake’ or ‘Bill cut the grass’ or ‘The tailor cut the cloth’…”34. That Sally is not going to run over the cake with a lawn mower or that Bill is not going to slice the grass with a knife is not obvious from the sentences themselves, yet we understand these statements because of our background capacities.35 The meaning of these statements by themselves is underdetermined, but, we know what they mean because we carry in our brains significant built-in capacities and knowledge about the world.

Searle also makes the point that the need for background in understanding is not the result of ambiguity, because “…progressive efforts at precision are not sufficient to remove the need for Background.”36

Suppose I go into a restaurant and order a meal. Suppose I say, speaking literally, “Bring me a steak and fried potatoes.” Even though the utterance is meant and understood literally, the number of possible misinterpretations is strictly limitless. I take it for granted that they will not deliver the meal to my house, or to my place of

33 Id. at 177.
35 Id.
36 See SEARLE, supra n.4, at 180.
work. I take it for granted that the steak will not be encased in concrete, or petrified. It will not be stuffed into my pockets or spread over my head. But none of the assumptions was made explicit in the literal utterance. The temptation is to think that I could make them fully explicit by simply adding them as further restrictions, making my original order more precise. But that is also a mistake. First, it is a mistake because there is no limit to the number of additions I would have to make to the original order to block possible misinterpretations, and second, each of the additions is itself subject to different interpretations.37

The following survey from psychologists Amos Tversky and Daniel Kahneman illustrates the problem of indeterminacy in the discovery of meaning using the phenomenon that they call “framing.”38 Framing is a term psychologists use to describe a phenomenon well known to any trial attorney: the way a problem is presented can skew a juror’s perception.

In this study, subjects were presented with a prologue and then asked to evaluate two programs:

Prologue:
Imagine that the U. S. is preparing for the outbreak of an unusual Asian disease, which is expected to kill 600 people. Two alternative programs to combat the disease have been proposed. Assume that the exact scientific estimate of the consequences of the programs is as follows:

Problem 1
If Program A is adopted, 200 people will be saved.

If Program B is adopted, there is a 1/3 probability that 600 people will be saved and a 2/3 probability that no people will be saved.

A sample of 152 students was asked to state which program they would favor. Seventy-two percent preferred program A and 28 percent preferred program B.

37 Id.
A second group of 155 students was given the same prologue and was also asked to evaluate two programs:

**Problem 2**

If Program C is adopted, 400 people will die.

If Program D is adopted, there is $\frac{1}{3}$ probability that nobody will die and $\frac{2}{3}$ probability that 600 people will die.

When asked which of these two programs they would favor, 22 percent favored program C and 78 percent preferred program D.

In fact, the number of people who live and the number who die in A and C (400) are identical, and the expected deaths are the same in both B and D. “The only difference between them is that the outcomes are described in problem 1 by the number of lives saved, and in problem 2 by the number of lives lost. The change [in the descriptive language used] is accompanied by a pronounced shift from risk aversion to risk taking. We have observed this reversal in several groups of respondents, including university faculty and physicians.”39 Thus, the same information is presented in each of the two problems, but presented differently.

This result confounds adherents to a rational choice model. Rational choice requires that different descriptions of a set of options that provide the same information should not result in different preference orderings over the options. Furthermore, effective communication requires that individuals grasp the meaning of a particular concept in virtually an infinite number of sentences that could be formed involving that concept. The necessity of built-in background in order to understand the meaning of individual utterances explains the reason for the problems of framing. In fact, framing is a potential problem in all social interaction and it is mitigated in most cases by the existence of sufficient background for meaning to be determined. Individuals
bring to each act of communication a great deal of knowledge about the physical and social environment, much of this is a product of growing up and living in a particular environment. Importantly, meaning is not contained in words, “…rather, meaning is constructed. The construction process relies on environmental signals (words and gestures) and conceptual knowledge and processes. Signals, words, gestures, and expressions do not mean; they are prompts for the construction of meaning.”40 The discovery of meaning, then, is not simply an act of perception, it is rather a process of construction. As we will discover below, part of the process of the discovery of meaning involves assessing the statistical implications of data. Further, asking jurors to withhold from making a judgment until all the evidence is in is equivalent to asking them, to defer understanding the meaning of the evidence until it is all in; this is not possible given the nature of human perception and cognitive mechanisms.

IV. THE STATISTICS AND MEANING

Failure to accurately construct meaning is a particular problem when individuals confront novel situations where they may have insufficient background for that construction to take place, a problem with technical, medical and scientific evidence. Further, in adversarial trial situations lawyers may purposefully frame their statements to sway jurors. Assessing statistical evidence is a serious problem in both criminal and civil trials. Individuals seem to be particularly poor at understanding and combining statistical information at a conscious level. This applies to experts as well as non-experts. The problem addressed in this section is often referred to as the base rate

39 Id.
fallacy and it is clearly illustrated by the often cited study in which sixty students, faculty and staff (physicians) at Harvard Medical School were asked to solve the following problem:

Determine the probability that a patient who tested positive on a test for a disease actually has the disease. The patient has no other symptoms or known risk factors, the prevalence of the disease in the population at large is 1/1000 and the test has a 5 percent false positive rate, that is, 5 percent of those who do not have the disease, test positive, alternatively, 95 percent of those who do not have the disease test negative.

The answer given by the largest number of respondents was 95 percent, the average answer was 56 percent and the correct answer is approximately 2 percent, which was given by 11 of the respondents.\textsuperscript{41} The correct answer is derived by using Bayes’s rule.\textsuperscript{42}

\textsuperscript{41} Ward Casscells, Arno Schoenberger & Thomas B. Graboys, \textit{Interpretation by Physicians of Clinical Laboratory Results}. 299 NEW ENGLAND JOURNAL OF MEDICINE 99 (1978).

\textsuperscript{42} See R. Duncan Luce & Howard Raiffa, \textit{Games and Decisions} (1957). Bayes’s rule provides the optimal way of combining statistical information. It is given by:

\[
P(A|B) = \frac{P(A) P(B|A)}{P(B)}
\]

In the medical example A is event that the patient has the disease, B is the event that he tests positive so that P(A|B) is the probability that he has the disease, given that he tested positive. P(A) is the probability that he has the disease, before we know the test result—given that we have no other additional information. The probability that he tests positive whether he has the disease or not, is given by P(B), and P(B|A) is the probability that the individual will test positive given that he has the disease, in the example above, it was assumed, for simplicity, that P(B|A) = 1, meaning that if a patient has the disease he will always test positive that is, there will never be a false negative. [The assumption that P(B|A) = 1, has no effect on the basic implications of the discussion.] Bayes’s rule allows for the optimal combining of the information available prior to the test, with the new information provided by the test result. The probability that the person has the disease prior to taking the test, P(A), is combined with the new information, that the test result was positive. The value, P(A), is called the prior probability and in the example P(A) = .001, P(A) is also called the base rate. Bayes’s rule allows us to optimally combine the prior probability, P(A), with the new information. Applying Bayes’s rule gives us the term on the left side of the equation, P(A|B), which is the posterior probability, the probability that the individual has the disease, given that he tested positive. The complicating factor here is that the test is not 100 percent reliable, 5 percent of the time a person that does not have the disease will test positive. The base rate fallacy occurs because too much weight is placed the relatively low 5 percent error rate and not enough weight is given to implications of the low prevalence of the disease in the general population, given this error rate.
How realistic is it to assume that jurors, who on average will have less education and training than Harvard Medical School faculty, staff and students will correctly analyze statistical information? This problem has led some researchers\textsuperscript{43} to argue that statistical evidence should be presented in frequency form rather than in terms of probabilities. They present evidence that even without training or instruction medical and law students are much better at making correct statistical inferences when given data in frequency from rather than in terms of probabilities.\textsuperscript{44} Converting the probabilities to frequencies greatly simplifies the problem and helps reduce the likelihood that individuals will commit base rate facility. In the example above, if one thousand people are given the test on average one person will test positive because they have the disease, but an additional fifty people will 50 people will test positive, and not have the disease. In frequency form, on average 1/1000 will have the disease and about 50/1000 more will test positive, but not have the disease. Of the 51 individuals who test positive, on average only one will actually have the disease which is a frequency of 1/51 meaning that slightly less than 2 percent of the individuals who test positive will actually have the disease.\textsuperscript{45} A false positive rate of 5 percent seems to indicate that the test is fairly accurate, but given to low prevalence of the disease in the population—the low base rate—the number of false positives will swamp the number of true positives in any significantly sized random sample of test results examined. Even if the false positive rate were reduced to 1 percent the likelihood of having the disease given a positive test result increases to only 10 percent. And if the false positive rate is reduced .1

\textsuperscript{43} Ulrich Hofferage, Samuel Lindsey, Ralph Hertwig, & Gerd Gigerenzer, \textit{Communicating Statistical Information}, 290 SCIENCE 2261 (2000).
\textsuperscript{44} Id.
\textsuperscript{45} This is a slight simplification, in fact on average if one individual has the disease then 999 individuals will not be sick and there will be .05\times999 false positives, which is slightly less than 50.
percent, the likelihood of having the disease, given a positive on the test will still only be 50 percent.

Bayes’s rule should be applied in all cases where evidence is considered by the jury. All forensic tests, including drug and alcohol tests, finger print matching, ballistic and fiber tests and even DNA matching are subject to errors which can confound the results.\textsuperscript{46} Optimal decision-making requires that jurors use Bayes’s rule to update their assessment of guilt or innocence when considering each new piece of evidence. Individuals, not surprisingly, are not very good formal Bayesians because they do not have the proper background to consciously construct the meaning of the evidence presented. This background can be acquired by taking classes in statistics and probability, however, this is not likely to be a solution to the statistical problems confronting jurors. On the other hand, as we have seen, presenting evidence in frequency form rather than as probabilities simplifies the problem by converting it to a format that many individuals, in fact, have sufficient background to work with, allowing them to do a much better job at arriving at the correct Bayesian solution. Here again, we have a framing problem: Presenting the same information in two different forms, as frequencies or as probabilities, changes the answer many individuals give to a problem. As we will see below individuals are much better Bayesians dealing with problems that they frequently encounter in their daily activities. Further, there is evidence that assessing the reliability, honesty and integrity of individuals is a human trait that is a product of natural selection. It is a trait that was necessary because it allowed individuals to cooperate in a prestate social environment. Much of this cooperation involved nonsimultaneous exchange in the absence of state enforced contract law and property rights. This topic will be briefly covered in the next section.

\textsuperscript{46} \textit{Id.}, Hofferage, et al. report a false positive rate in laboratory DNA test of about .003.
V. THE EVOLUTION OF COOPERATION

Anthropologist Christopher Boehm argues that the dominant political feature of recent hunter-gatherer societies, and presumably that of Paleolithic hunter-gatherer (H-G) societies as well, was egalitarianism.47 These H-Gs lived in small bands of about twenty-five individuals in which cooperation in general and the sharing of large game in particular was a necessity for survival. The high level of cooperation that made human survival possible was a product of the evolution of large brains, in our hominid ancestors and the evolution of language.48

The large brain and language necessary to support the high level of human social and economic cooperation where extremely costly. “When the body is resting, the nervous system consumes about 20 percent of the body’s oxygen supply, which is the lion’s share, considering that the brain accounts for only about 2 percent of the body’s mass”49 “[B]ig brains are expensive organs, requiring a lot of energy to maintain—22 times as much as an equivalent amount of muscle requires when at rest”50 “The fact that an organism has a large brain means that it really must need it very badly, otherwise the forces of natural selection will inexorably favor individuals with smaller brains simply because they are cheaper to produce. Animals that have to spend a lot more time feeding to provide fuel for their big brains (or mothers that have to do even more to provide for the infant’s brain as well as their own) expose themselves to proportionately greater risks of predator attack, not to mention starvation when things get tough in famine years”51

49 PATRICIA SMITH CHURCHLAND, NEUROPHILOSOPHY 36-37 (1986).
Given the cost of heat dissipation and other constraints, energy production, for a given body size, is relatively fixed, in order for more energy to be made available for a larger brain, other organs must use less. Since the size of the heart, lungs, kidneys or liver cannot easily be reduced; the cheapest means of freeing up energy was the evolution of a smaller gut. A smaller gut will use less energy, however, it will also absorb less energy from ingested food. To maintain energy input with a smaller gut, humans must eat foods that are higher in nutrient content or that contain nutrients that are more easily absorbed. The brain-gut substitution is sometimes called the expensive tissue hypothesis. Significantly increased prenatal and postnatal nutritional contribution by the mother, was also necessary for the development of our large brain. “…[This] maternal investment hypothesis proposes that most of the extra energy comes early in life—from Mom, through the placenta during pregnancy and through breast milk between birth and age 4, when the human brain reaches 85% of its adult size”

The increased energy requirements of a larger brain forced our hominid ancestors to turn to meat with its high energy and nutrition content. It is in the context an increased importance of meat consumption that meat-sharing takes on its importance in human biological and cultural evolution. Meat-sharing is a sophisticated form of cooperative activity that required significant cognitive capacity, as it entailed complex mental accounts and forward-looking actuarial intelligence that far exceeds the abilities of chimpanzees our closest cousins; further, meat-sharing was necessary for such cognitive evolution to take place. The problem is that meat-sharing and other forms of H-G cooperation involved the non-simultaneous exchange of

\footnotesize
\begin{itemize}
\item 52 Id. at 125.
\item 53 Id.
\item 55 See Dunbar, *supra* n.51 at 127 and Mithen, *supra* n.50 at 103.
\end{itemize}
goods and services in the absence of state enforced contracts, where the potential for cheating is always a problem. Significant increases in cognitive capacity where necessary for the evolution of informal the norms and social structures that allowed these societies to survive.

Anthropologist Robert Kelly points out that “[s]tudents new to anthropology…are often disappointed to learn that these acts of sharing come no more naturally to hunter-gatherers than to members of industrial societies…. Sharing…strains relations between people” 57

Though other food resources were shared in some limited circumstances, the meat of large game is always shared. 58 Variations in the daily returns to hunting, where food storage is not an alternative, resulted in meat-sharing being an efficient way of reducing the variance in access to meat facing any one family. Hunters confront diminishing marginal value with the large quantities of rapidly perishable meat yielded by large game. Meat-sharing can be considered a form of insurance allowing hunters who were successful on a given hunt to trade lower-valued surplus meat for claims to high-valued meat in the future when their hunting was not successful. Kelly reports that a band size of 25 would contain seven to eight full-time foragers and hunters, which he calculates was sufficient to minimize the variance associated with the uncertainty of food acquisition.

Finally, a group of “…25 persons (with a mix of men and women) is also the minimum number that could withstand short-term fluctuations in fertility, mortality, and sex ratio for any length of time. That is, groups smaller than 25 persons have a low probability of being reproductively viable.” 59 The capacity of a foraging area also constrains group size. A further factor important in the maintenance of H-G egalitarianism, is that a significantly larger group

57 Id. at 164-165.
58 Id. at 165.
59 Id. at 211.
would, by sharply increasing the cost of maintaining that egalitarian social arrangement, forcing the group to become hierarchical, with a dominant individual coordinating the group’s activity. All these factors imply that about 25 persons was the optimal band size.

Hunter-gatherer egalitarianism and nonsimultaneous exchange involving meat-sharing and other forms of cooperation was maintained by that constant monitoring of the behavior of others and by the development of what Boehm calls actuarial intelligence. Actuarial intelligence is “…the intuitive human capacity, seen abundantly in hunter-gatherers, to think stochastically and to understand rather complex systems on an intuitive but statistically valid, predictive basis. Regardless of what drove human brains to be so large, one product was the generalized capacity to understand and manipulate complex systems of various types.”

Actuarial intelligence, allows individuals to compute the long-term costs and benefits social interaction and to maintain the complex forward-looking assessments of other individuals necessary for the social system to function.

Social control involves far more than an outraged group’s suddenly deciding to employ dramatic sanctions. In any small group anywhere, people keep track of one another’s behavior and try to read underlying motives. Types of deviance that all human groups watch for, gossip about, and react to, include murder within the group, heavily self-interested verbal deception, theft, and stinginess or failure to cooperate when this is appropriate. On the positive side, foragers talk about generosity, cooperativeness, honesty, and other prosocial behaviors that involve goodwill. In effect the band keeps a dossier on every individual, noting positive and negative points.

Voluntary cooperation has been studied from both a theoretical and an experimental perspective using the concept of image scoring. In the group each individual has an image score

---

60 BOEHM, supra n.48 at 183.
61 BOEHM, supra n.48 at 73.
that measures the degree to which that individual has cooperated in the past with others and that is known by all players, though not perfectly. The image score is a measure of the individual’s trustworthiness, reliability, reputation and status and it is subject to continual reassessment and reevaluation by others. Cooperation in this setting occurs in the form of repeated nonsimultaneous exchange, in the absence of formal contract enforcement, where the meat sharing of hunter-gatherer bands is a classic example. Nowak and Sigmund find that “[c]ooperation wins in a computer simulation of indirect reciprocity,”64 where the cooperation “depends crucially on the ability of a player to estimate the image score of the opponent.”65 The evolution of hunter-gatherer type cooperation requires that individuals have sufficient cognitive capacity to recognize other members of their group, distinguish between cheaters and cooperators (i.e., assess the general trustworthiness and reliability of others), and maintain appropriate mental accounting that reflects the social reliability of others as well as the individual’s obligations to each of the others in the group and theirs to him, in the form of image scores that are a product of human actuarial intelligence. Humans have evolved to maintain just this sort of mental accounting apparatus which are used to maintain social control by identify those individual who violate the social contract who then face group sanctioning. Furthermore individuals maintain and update these mental accounts during their ongoing daily activities using nonconscious Bayesian mechanisms. They also use gossip to acquire information about others and to further refine the image scores that they maintain on the individuals in their group.66 These evaluations of individuals are in some ways similar to the types of assessments of

64 See supra n.32 at 573.
65 See Nowak & Sigmund, supra n.62 at 575.
66 See DUNBAR, supra n.51 and BOEHM, supra n.48.
individuals that juries must perform. However these evaluations are an ongoing part of daily life and do not require that individuals remember significant amounts of information before making their assessments or reassessments of individuals.

Jury instructions require that they do not form any posterior probabilities until all the evidence has been heard. This seemingly presents the juror with a monumental cognitive task. When new information is presented sequentially, Bayesian updating can be performed as each new item of information is processed, alternatively, the information can be stored in memory and then processed all at one time during jury deliberation. The memory and processing requirements are significantly greater for the latter approach than for the former.

VI. MENTAL ACCOUNTING

Each individual maintains an image score or mental account for each of the other individuals with whom she associates. These accounts are a product of evolution and their development was necessary for the evolution of cooperation in the form of nonsimultaneous exchange and for prosociality in general. Given the capacity limitations of the human memory systems and the fact the accounts are continuously and permanently maintained they, as suggested by Nowack and Sigmond, very likely rely on relatively simple Bayesian sequential updating mechanisms. A juror in a criminal trial will maintain an image score that reflects the juror’s belief about the likely guilt or innocence of the defendant as well as separate scores on others involved in the trial. Each new item of evidence will result in the score being updated using nonconscious Bayesian processes. Though individuals do a poor job using conscious Bayesian updating increasingly neuro and cognitive scientists are finding that using a Bayesian approach to perception and cognition produces “…remarkable successes in generating models
that predict existing empirical findings.…"67 This naturally leads to the question: If the outcomes of perceptual and cognitive systems can accurately be predicated by models assuming Bayesian optimality, why do individuals do such a poor job when they try to consciously solve updating problems? First of all, we have seen that when the data is presented in frequency form individuals are fairly good Bayesian updaters. Secondly, the perceptual and cognitive systems, that perform this updating using nonconscious lower level mechanisms, are a product of biological evolution. Third, they work best in situations that are familiar to the individual, Koehler (1996) argues that in ecologically valid situations, that is, real world situations that individuals encounter regularly, individuals are much more accurate in their use of Bayes’s rule in statistical analysis.68 Fourthly, although doctors might do a poor job when first confronted with a new medical test, their understanding of the subtleties of combining base rate information with the results of imperfect tests improves significantly with experience with that test.69 Further, as we have seen individuals have evolved to be proficient at the type of social assessment required of jurors in both criminal and civil trials. Jurors will automatically maintain an image score on defendants, updating that score as new evidence is introduced. Finally, that score is likely to closer to Bayesian optimality than a conscious assessment process performed after all the evidence has been heard. A further problem with deferring judgment is the limitations of human memory.

69 Id.
VII. WORKING MEMORY

Although there are several types and ways of categorizing memory, for our purposes here, we need only consider short- and long-term memory. “An essential criterion of distinction between the two is their temporal persistence—in other words, the length of time during which a memory is retained and retrievable. That length of time may be seconds to minutes in short-term memory and up to years in long-term memory. Information capacity is another distinction between the two: The capacity of short-term memory is limited, whereas that of long-term memory supposedly is not.”70 Short-term or working memory is memory that currently active and available for conscious introspection and processing. We are not conscious of the contents of long-term memory, rather it must be retrieved into working memory to be available for conscious introspection. Memory is not perfect, long-term memory malfunction leads to a juror forgetting some vital piece of evidence and the limited capacity of short term memory creates additional problems.

Working memory has two components, the first being a form of short-term memory, the contents of which are derived from both long-term memory—things we have previously learned—and new information provided by the senses. The contents of the short-term memory are maintained “on line” and available for processing. The second component of working memory involves the processing of the contents in short-term storage, this processing of the contents in short-term storage is facilitated by the prefrontal cortex. However, it is more correct to say the prefrontal cortex facilitates the highest level of processing, the most abstract, and that involving plans for the most distant future. In the brain, unlike a computer, all areas, from the

highest to the lowest, are involved in the processing of information and facilitate the generation of plans of action.

The working memory system maintains memory “online,” the contents of which are used to organize temporally sequenced behavior. That is, working memory facilitates purposeful conscious behavior, including “abstract reasoning, complex problem solving, and planning for the future…” There is a cost to maintaining information online for rapid processing, however, and that is that the amount of information that can be maintained for ready availability is extremely limited. In a seminal article Miller reviewed the evidence regarding the limited capacity of short-term memory. He concluded that individuals could maintain seven plus or minus two items in short-term memory. In a review and evaluation of more recent evidence Cowen (2000) argued for the even smaller capacity of four plus or minus two. Whatever the exact capacity constraint, the limitations of working memory suggests that jurors will have trouble processing a significant amount of evidence at any one time. Given the constraints on working memory, jurors will be unable to consider all the evidence at one time and will have to chunk it sequentially, either as it is heard; or after it is “all in” by recall from long-term memory. Waiting places an increased burden on long-term memory since it requires that all the details must be remembered, whereas updating sequentially requires remembering only the most recent evidence and the current image score.

71 Susan M. Courtney, Laurent Petit, José Ma Maisog, Leslie G.Ungerleider,; and James V.Haxby, An Area Specialized for Spatial Working Memory in Human Frontal Cortex, 279 SCIENCE 1347-1351 (1998).
72 George A. Miller, The Magical Number Seven, Plus or Minus Two: Some limitations on our Capacity For Processing Information, 63 PSYCHOLOGICAL REVIEW 81 (1956).
VIII. THE PRODUCTION OF MEANING

The meaning of our perceptions, speech acts and all other intentional states must be constructed in the brains of the perceiver. The mechanisms that allow for the construction of meaning are products of both phylogenetic and ontogenetic processes. Their substrate is a product of natural selection, so that the ontogenetic is built upon a foundation of phylogenetic mechanisms. The construction of meaning from our perceptions and intentional states is for the most part an automatic and ongoing process that occurs within milliseconds of the initial perception of the objects, events, speech acts, desires, etc. that form the content those intentional states. The mechanisms of meaning are not only automatic they are involuntary processes. The assignment of meaning is not something over which we have voluntary control. If these mechanisms did not determine meaning rapidly, automatically and involuntarily, our ancestors could not have responded rapidly and appropriately to perceived threats and opportunities in the environment which would have significantly reduced the likelihood of their survival.

Higher social mammals, including humans that participate in cooperative activities, have as a product of natural selection, mechanisms that that allow them to keep tract of, assess and update the reliability and trustworthiness of members of their group. The evolution of language and large brains allows humans to maintain complex forward-looking image scores as part of a complex mental accounting system used to evaluate and update their assessment of others when new information is available. These mechanisms contribute to the construction of the meaning of the actions of others. In other words, the mental accounts participate in the construction of the meaning of the social interactions we observe directly and hear about form others. Like all mechanisms of meaning the updating of the mental accounts we maintain on other individuals after the receipt of new information is for the most part at automatic and nonvoluntary.
We have argued that jurors make use of these mechanisms of meaning in their assessment of witness, defendants, lawyers and others involved in a trial. Further, that these assessments are unavoidable, that is the jurors cannot follow the instructions. Asking them to do so would be like asking someone to watch a movie in their native language, but refrain for determining the meaning of what the actors say until the movie is over. Even if jurors could refrain from the assessment of meaning until jury deliberations the constraints on both long-term and working-memory would make the task impossible. The primary aid available for keeping track of a large amount of detailed information is writing. Writing as an external analog of both long-term and working memory, is often not permitted in the courtroom.

It seems unavoidable that jurors will bring to the deliberation process an image score that reflects their belief in the guilt or innocence, or degree of liability or negligence of the defendant. However the consultative group process of deliberation is similar to the gossip sessions of our hunter-gatherer ancestor in which they shared information about others, and altered their image scores when the information warranted, hopefully jurors using a similar process arrive at the “facts” of the case. In other words, forming an opinion about guilt or innocence during the course of the trial may be unavoidable, jurors instead should be encouraged to go into the deliberations with an open mind, and a willingness to change their opinion if the argument and facts presented during deliberation suggests that their initial opinion was mistaken.

IX. CONCLUSION

There are biological, physical and social constraints that prevent jurors from following the standard instructions, to refrain from forming opinions about a defendant’s guilt or liability until all of the evidence has been presented. In part this is a product of the complexity of the problem. First, there is the difficulty of accurately perceiving and understanding the meaning of
the evidence presented. Second is the problem of combining the evidence in an optimal manner. Third, there are severe memory constraints that make an “after all the evidence has been heard” combinatorial process difficult if not impossible.

Both biological and cultural evolution has equipped individuals with the ability to evaluate information about others and use that evaluation to assess the reliability and trustworthiness of others. Jurors will automatically evaluate defendants, in fact we argue that not do so is practically impossible. However, if evidence is presented in an easily understandable form, a process in which jurors update their assessment of guilt or innocence while the trial is in progress will facilitate a more accurate assessment of the facts given the limitations in human cognitive capacity.