LEGAL DELIBERATION
a study in the philosophy of law

by
Gregory R. Hagen
B.A., The University of British Columbia, 1987

A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF
THE REQUIREMENTS FOR THE DEGREE OF
MASTER OF ARTS

in
THE FACULTY OF GRADUATE STUDIES
Department of Philosophy

We accept this thesis as conforming
to the required standard

THE UNIVERSITY OF BRITISH COLUMBIA

AUGUST 1990
© Gregory R. Hagen, 1990
Authorization

In presenting this thesis in partial fulfillment of the requirements for an advanced degree at the University of British Columbia, I agree that the library shall make it freely available for reference and study. I further agree that permission for extensive copying of this thesis for scholarly purposes may be granted by the head of my department or by his or her representatives. It is understood that copying or publication of this thesis for financial gain shall not be allowed without my written permission.

Department of Philosophy
The University of British Columbia
1866 Wesbrook Mall
Canada V6T 1W5
Abstract

This thesis examines deliberation in legal proceedings. Legal deliberation is conceived of as the procedures by which a judge, jury, or other rational deliberating agents arrive at a verdict. Legal deliberation involves deliberation about laws and about facts. This thesis is concerned chiefly with deliberation about facts and how value considerations impinge on deliberation about facts. In legal proceedings there are a number of principles that are generally accepted, although their application varies according to whether the procedure is criminal, civil, administrative or other. These principles include: an accused must be proved to have committed an act according to a given standard; a person is presumed innocent until proven guilty; a proposition may be presumed by making an inference from a basic proposition which has been proved; only relevant evidence may be admitted; only reliable evidence is accepted; evidence may be accepted on the basis of judicial notice; and unreliable evidence may be admitted if corroborated. Some less familiar principles are that proved propositions are consistent; all the elements of a case need to be proved in order for the case to be proved; a proposition at issue is not proved unless it is based upon complete evidence; and that the degree of persuasion that a deliberator has towards a proposition at issue be equal to the objective probability of that proposition.

Although these principles are generally accepted the interpretation of these principles is unsettled. This thesis attempts to give an interpretation of these principles which justifies them. All interpretations have in common, I hold, that a rational agent has principles for modifying his deliberative state given new evidence. The deliberative state of an agent consists of a set of elements \(< B, D, S, K, +>\) where \(B\) is the agent's degree of belief over a set of propositions \(S\); \(K\) is a subset of \(S\) — the full or accepted beliefs; \(D\) is the agent's degree of desirability over propositions in \(S\); \(+\) is a dynamical principle of deliberation which determines how values for \(B(S)\), and \(D(S)\) change over time. The desires of the agent are taken to be a reflection of the values inherent in legal principles.

A traditional principle is that in order to convict someone it must be proved that the accused committed the alleged act. There is little agreement, however, about what is involved in proving that a person has done something. There are two main theories which are used in law. One theory, the Austinian theory, takes an action to be a bodily movement that is voluntary. A second, wider view is that an action includes the bodily movement, consequences, relevant circumstances and voluntariness, and perhaps other elements, such as omissions, and things that happen to one, but not a mental event. I argue that an action is a part of a sequence of causally
related events. It is that part of the sequence with the properties that are represented to the agent in his causally efficacious mental state.

The interpretation of “prove” in the last cited principle is also unsettled. All views hold that, in some sense, a proved proposition be sufficiently probable. There are five views of probability that I canvass: the logical view, the subjective view, the relative frequency view, the chance view and the epistemic view. I argue that the epistemic view is particularly suited to legal reasoning. On this view probability is conceived as a mind independent logical relation between evidence admitted and the conclusion reached on the basis of that evidence. Probability also reflects the underlying chance of single events and so applies to individual actions.

The traditional practices have been interpreted as the dynamics of deliberative states. There are two plausible models of these dynamics: Bayesian and non-Bayesian. On Bayesian theories all changes of belief are by Bayes’ theorem or generalizations thereof. On a non-Bayesian view beliefs are changed by accepting new beliefs, conjoining them with the old beliefs, and modifying the old beliefs on the basis of the new ones. As an interpretation of legal deliberation the Bayesian view has a number of disadvantages. Among other difficulties I survey, on the Bayesian view one can not consider a case proved if all the elements of a case are proved, and one cannot regard a proved proposition at issue as true. Hence I reject the Bayesian theory.

The principle that a person is proved to have committed an act if it is sufficiently probable that he committed such an act gives rise to a difficulty. Ultimately the problem amounts to how a theory of deliberation can meet three principles of legal reasoning: the deliberating agent’s beliefs are consistent, the agent believes a proposition A if the probability of A is sufficiently high, and if the agent believes A and believes B then he believes \( A \land B \). I show how this problem is resolved by requiring probability to be resilient.

A person is proved to have committed an act if the probability of having committed that act reaches an appropriate standard of proof. But what is the standard that is at issue here? If the judge is a utilitarian, for instance his desire function must meet the constraint that it equals the average desires of all other agents. In the final chapter I argue that a utilitarian rationale for standards of proof violates a person’s right not to be convicted if innocent. This is due to the fact that a person can be convicted by a utilitarian deliberator even though it is more probable than not that he did not commit the alleged offence.
Table of Contents

1. Introduction.

1.0 The Justification of Principles of Deliberation in Law. 1
1.1 An Outline of Evidence Law. 3
1.2 Theories of Probability and Inference. 7
1.3 Action. 12
1.4 The Rational Deliberator. 15

2. Proof of Action.

2.0 Proof of Action. 19
2.1 Causal Theories of Action. 21
2.2 The Narrow theory: the Classical Austinian View of Action. 26
2.3 The Wide Theory of Action. 30
2.4 The Coval-Smith Theory of Action in Law. 34
2.5 Law and Descriptions of Events. 38
2.6 Conclusion. 39

3. Interpretations of Probability in Law.

3.0 Introduction. 40
3.1 The Uses of Probability in Law. 42
3.2 The Logical Interpretation of Probability. 46
3.3 The Subjective Interpretation of Probability. 49
3.4 The Empirical Interpretation of Probability. 56
3.5 The Need for Proof of Individual Conduct: the Chance Theory 57
3.6 The Epistemic Interpretation of Probability. 60
3.7 Conclusion. 62
<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>4. Legal Deliberation.</td>
<td></td>
</tr>
<tr>
<td>4.0 Legal Deliberation.</td>
<td>64</td>
</tr>
<tr>
<td>4.1 A Model of Deliberation: The Rational Deliberative Agent.</td>
<td>65</td>
</tr>
<tr>
<td>4.2 Probability and Proof.</td>
<td>69</td>
</tr>
<tr>
<td>4.3 Rationality, Deductive Closure and Consistency.</td>
<td>71</td>
</tr>
<tr>
<td>4.4 Presumptions, Judicial Notice and Admission of Evidence.</td>
<td>73</td>
</tr>
<tr>
<td>4.5 Completeness of Evidence.</td>
<td>79</td>
</tr>
<tr>
<td>4.6 Presumptive Inference.</td>
<td>82</td>
</tr>
<tr>
<td>4.7 Conjunction.</td>
<td>85</td>
</tr>
<tr>
<td>4.8 Conclusion.</td>
<td>87</td>
</tr>
<tr>
<td>5. The Gate-crasher Paradox.</td>
<td></td>
</tr>
<tr>
<td>5.0 Standards of Proof the Gate-crasher Paradox.</td>
<td>88</td>
</tr>
<tr>
<td>5.1 What are the Standards of Proof?</td>
<td>90</td>
</tr>
<tr>
<td>5.2 The Gate-crasher Paradox.</td>
<td>91</td>
</tr>
<tr>
<td>5.3 The Lottery Paradox.</td>
<td>92</td>
</tr>
<tr>
<td>5.4 Attempts to Save the Purely Probabilistic Standard of Proof.</td>
<td>95</td>
</tr>
<tr>
<td>5.5 The Weight of Evidence.</td>
<td>99</td>
</tr>
<tr>
<td>5.6 Resiliency, Chance, and Probabilistic Standards of Proof.</td>
<td>102</td>
</tr>
<tr>
<td>5.7 Conclusions.</td>
<td>103</td>
</tr>
<tr>
<td>6.0 Introduction.</td>
<td>104</td>
</tr>
<tr>
<td>6.1 Utilitarian Standards of Proof.</td>
<td>105</td>
</tr>
<tr>
<td>6.2 The Standard of Proof in the Case of Multiple Defendants.</td>
<td>110</td>
</tr>
<tr>
<td>6.3 Rights Based Standards of Proof.</td>
<td>111</td>
</tr>
<tr>
<td>6.4 Conclusion.</td>
<td>112</td>
</tr>
<tr>
<td>7. Conclusions.</td>
<td>114</td>
</tr>
<tr>
<td>References.</td>
<td>116</td>
</tr>
</tbody>
</table>
List of Tables

<table>
<thead>
<tr>
<th>Table</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Table 1</td>
<td>47</td>
</tr>
<tr>
<td>Table 2</td>
<td>52</td>
</tr>
<tr>
<td>Table 3</td>
<td>76</td>
</tr>
<tr>
<td>Table 4</td>
<td>77</td>
</tr>
</tbody>
</table>
## List of Figures

<table>
<thead>
<tr>
<th>Figure</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Figure 1</td>
<td>13</td>
</tr>
<tr>
<td>Figure 2</td>
<td>36</td>
</tr>
<tr>
<td>Figure 3</td>
<td>36</td>
</tr>
</tbody>
</table>
Acknowledgments

Thanks are due to my parents for support, to Sam Coval for his comments and supervision, to Andrew Irvine for his expeditious and thoughtful reading, and to Steve Savitt for trying to help me write better and do better philosophy. I also thank the Philosophy Department for giving me a Teaching Assistantship for two years.
Introduction

1.0 The Justification of Legal Deliberation.

Our legal system operates on the assumption that there are rational principles which govern deliberation and proof in legal proceedings. It is assumed, for instance, that in order for an accused to be successfully prosecuted for an offence that the Crown must be able to prove that the accused committed the offence. It is also assumed that the judge, jury or other decision maker engage in some form of rational deliberation which involves weighing the evidence that is presented, while at the same time considering the rights of individuals and the judgement's effects on the community. What are these principles governing deliberation and to what extent are they justified? This is the fundamental philosophical question that needs to be answered. Without an answer to this question the everyday procedures in our courts, tribunals and other settings lack any appropriate justification.

On my view what is needed is a reconstruction of legal practice. At a superficial level there is agreement on the procedures and principles utilized in law. But these traditional procedures in law are capable of a wide variety of interpretations. For instance all legal philosophers think that a standard of proof must be met in order for a proposition at issue to be considered proved. One traditional view is to regard a proposition which is proved as true. If a person is found guilty beyond a reasonable doubt then it is considered to be true that he committed the action he was charged with. A different view, associated with Bayesians, is that propositions at issue are not considered true or false. Rather a subjective attitude of confidence is applied to the proposition. Considering a proposition proved is only to passively acknowledge that a certain degree of confidence has been met. Both of these interpretations are meant to explain what is held in common: that a standard of proof must be met in order for a person to be convicted.

On this and other principles of legal deliberation there is a superficial level of agreement that might be taken to be the core of legal practice. What needs to be done is to arrive at justifiable interpretations of these practices. This is to engage in a rational reconstruction of legal deliberation.

There is a principle that says that in order to find one liable it must be proved that he committed a particular act. There is a need for an interpretation of “action”. I argue that an action is a sequence of causally linked events such that the first event is a mental state. The
action is that part of the sequence of events which is represented to the agent in his causally efficacious mental state. This view allows one to make important ontological distinctions that are required for law. The causal view distinguishes between actions, events, things that happen to one, states of a person, omissions, and ascriptions of responsibility.

The principle that it needs to be proved that an accused has committed an act in order to be found liable raises the question of what proof is. Nearly all theorists agree that the proposition at issue must have a sufficient probability, but there is little agreement about what probability is. A number of theories have been advocated over many years: the subjective theory, the logical theory, the relative frequency theory, the objective chance theory, and the epistemic theory. I argue that legal reasoning requires probabilities to apply to individuals, rather than to classes of individuals as in an insurance calculation, and that probability be interpreted as an objective logical relation which is relative to current knowledge. Thus I opt for the epistemic theory of probability which defines the probability of A being guilty as the logical relation which holds between the evidence of A’s guilt, and the conclusion that A is guilty. The evidence is understood to be a belief about the objective chance that A is guilty. The objective chance is determined more or less by the disposition of a person to act in a certain way in a given circumstance.

All sides can agree, I take it, that principles of deliberation are principles that a rational agent would use to adjust his mental states upon receipt of new evidence. But there are two main approaches to the dynamics of deliberation: Bayesian and non-Bayesian. The Bayesian approach takes changes of belief to be governed by "conditioning" through Bayes’ theorem or its generalizations. Non-Bayesian approaches take the point of view that changes of belief arise through addition of probable beliefs to our set of beliefs and then modifying the rest of our beliefs on the basis of these new beliefs. The rule of conditioning is not plausible in many cases of legal reasoning.

Furthermore Bayesians don’t regard propositions which are proved as true, while non-Bayesian hold that the outcome of deliberation is an acceptance of a proposition as true. On the Bayesian view some principles of deductive logic are justified in general, but not in every case. So the Bayesian does not hold the conjunction principle, that if A is proved and B is proved then (A & B) are proved. For the Bayesian this principle often holds but not in general. For the non-Bayesian it would be irrational not to hold the conjunction principle because if A is true and B is true so is (A & B). But, of course Bayesian’s don’t regard a proved proposition as true.
In the fifth chapter I follow up a question raised in the chapter on legal deliberation. This is L.J. Cohen's paradox of the gate-crasher. While many would agree with the view that a proposition at issue is proved if it has a high probability this purely probabilistic definition of the standard of proof runs into difficulty. Ultimately the paradox amounts to the assertion that probabilistic standards of proof conflict with a basic principle of proof, the conjunction principle, which states that if A is proved and B is proved then (A & B) is proved. I demonstrate that given probabilities based upon chances, there is no conflict between the legal principles.

A basic principle of legal reasoning is that proof must reach certain standards which are imposed for ethical reasons. In the sixth chapter I argue against a utilitarian justification of standards of proof. On a utilitarian view the deliberator maximizes the social welfare of society calculated in terms of the average utility level of persons in society. There are two mistakes that can be made in arriving at a verdict. One mistake is to find a person liable when he is not liable. A second mistake is to not find a person liable when he is liable. On the utilitarian view the value of these mistakes can be assessed in terms of the effect on the average utility in society. Given these assumptions the preponderance of the evidence standard and the proof beyond a reasonable doubt standard can be derived. But it also follows, I argue, that in the case of multiple defendants there is a wide range of occasions in which individual rights are violated because utilitarianism implies that a defendant should be found liable even though it is improbable that he committed the required act.

In the present chapter I give some background information which may be helpful. I give an overview of evidence law, as well as some background on theories of probability and inference, action, deliberation and the ethics of proof.

1.1 An Outline of Evidence Law.

The law of evidence can be divided into two branches: rules of proof and rules of admissibility. Rules of proof govern the logical structure of proof in legal proceedings. The rules of admissibility are rules for excluding evidence based upon legal considerations such as protection of the innocent, solicitor-client privilege, and the interests of the public. In practice these two branches are often not separated and questions of proof and admissibility are blended together. This blending together of proof and legal considerations is what I call deliberation. In
what follows I give an overview which is a highly idealized picture of an ill defined area of law and one which is not necessarily reflective of every jurisdiction.

The law of proof consists of a number of principles. Some of these principles are familiar enough. The most important rule is that, in general, a proposition at issue is considered proved only if it is beyond a reasonable doubt in criminal cases and more probable than not in civil cases. *(Miller v. Minister of Pensions)* In the U.S. the standard of clear and convincing evidence is sometimes cited. *(McCormick, 1984, 959)* These standards are allowed to change given the nature of the action for which one is accused. *(Bater v. Bater, Smith v. Smith)* It is considered a greater harm to find an innocent person guilty of a serious offence than of a minor one. The challenge here is to specify the nature of the interpretation of probability and the level of proof that should be required.

In distinction to the standard of proof there is an ultimate burden of proof that is placed upon one of the disputants. *(Thayer, 1898, 355)* This is the burden that exists for the party who will lose the case if the standard of proof is not met. The ultimate burden is atemporal in the sense that it extends throughout the proceeding. Secondly there is an evidential burden which may shift back and forth given the admission of different evidence. *(Thayer, 1898, 355)* This evidential burden lays upon the party who will lose the case at any particular time during the proceeding if the proceeding then stops. So it is possible for one party to have both the persuasive burden throughout the proceeding and yet have, at some point, admitted enough evidence such that the evidential burden is on the other party to answer the evidence.

The presumption of innocence is an important principle. Yet there is some dispute over what this principle is. *(Morton and Hutchison, 1988)* In an ordinary sense if one is presumed to be innocent then one is believed for that time to be innocent. But there has been some doubt whether one can be involved in legal proceedings while being presumed innocent. Thus it is often advocated that presuming innocence really is just an elliptical statement of who has the burden of proof. *(Morton and Hutchison, 1988)* One is presumed innocent if and only if the opposing party has the ultimate burden of proof. But these two principles are logically distinct. One could be presumed innocent at the outset of the case and have the persuasive burden of proof. In this sense the burden of proof would be easily met since innocence is presumed at the outset of the proceedings. Which view of presumptions is the best interpretation of legal practice?
Evidence is made available in legal proceedings either by judicial notice or admission of evidence. The general principle for admission of evidence is that all relevant evidence is admissible, subject to exceptions due to legal reasons. (Thayer, 1898, 264-9) Relevance is defined in terms of probability. (Stephen, 1948, 4) Evidence is relevant to a proposition at issue if it either raises or lowers the probability of that proposition. However, sometimes a certain degree of relevance may be required, in the sense that the evidence must confirm or disconfirm the proposition to a certain degree. (Wigmore, 1983, sec. 8) In this way evidence which is relevant only trivially is not admissible. However the Supreme Court of Canada apparently believes that any piece of evidence which is only very slightly relevant should be admissible. (Morris v. R) Judicial notice is a principle which allows the deliberator to take common knowledge into consideration. (R. v. Evgenia Chandris, The) There is some question as to what types of evidence should be judicially noticed and whether notice should be open to rebuttal on the basis of further evidence. (Zuckerman, 1989)

There are a number of rules of exclusion of evidence which properly belong to the law of proof and not admissibility. These are rules which specify that certain types of evidence are inherently unreliable or that their reliability is indetermined. Unreliability of evidence can be given a probabilistic formulation. The testimony E that H occurred is reliable if and only if the probability of E given H is sufficiently high. So, for example, the opinion rule specifies that opinions are generally unreliable. (Wigmore, 1978, sec. 1917) Hearsay is evidence of testimony made by a person not called as a witness, to a witness, which is offered by the witness as evidence of facts asserted in the testimony. The reliability of hearsay evidence can not always be established due to the inability to examine the witness. Hence hearsay is often considered inadmissible. (Sheppard, 1989, 421)

However, evidence that is considered unreliable for criminal proceedings may be considered reliable enough for non-criminal proceedings. Like relevance, reliability can be a matter of degree. Some witnesses are not considered competent to testify because their testimony is unreliable. Yet the testimony of witnesses may be considered reliable through corroboration of such testimony by others.

The presumption of innocence differs from other presumptions in law. The presumption of innocence is an epistemic attitude of assuming or taking for granted the innocence of the accused. Presumptions are inferences from a basic fact to a presumed fact. The law often divides presumptions into presumptions of law and presumptions of fact. Presumptions of fact are supposed to be inferences that are based upon common knowledge. Presumptions of law are
presumptions which must be drawn according to law. Presumptions in law appear to take a number of different forms. Two prominent forms are straightforward deductive arguments based upon inferences from background generalizations, and direct statistical inferences from statistical generalizations. In the first case one might infer that A is dead, from the fact that A has been missing for seven years, and the presumption that if A is missing for seven years, then A is dead. In the second case, it might be inferred that the probability that a specific letter reached its destination is 90% because 90% of all letters reach their destination.

The principles of admissibility are principles for excluding evidence due to legal considerations. There are a number of principles for excluding evidence due to legal considerations. These principles vary between criminal and civil proceedings. In criminal proceedings the rules of admissibility are quite strict and entrenched. In civil proceedings, the exclusion of hearsay, opinion, and privileges is often relaxed compared to criminal proceedings. (Sheppard, 1989, 212-231)

Some communications are considered privileged and such communication is inadmissible. (Zuckerman, 1989) Solicitors and their clients communications are considered privileged information. Marital communications are also considered privileged. Sometimes evidence is considered inadmissible due to public interest, or due to national security. Evidence, such as illegally obtained evidence may be inadmissible, if it brings the administration of justice into disrepute. (Sheppard, 1989) Evidence of character and the behaviour of the defendant in similar situations may be inadmissible because of the prejudice to the defendant. The prejudice to the accused is that the deliberator may be unable to correctly assess the support the evidence gives to the proposition at issue.

In addition to these commonly known principles and procedures there are a number of other principles which are not always explicitly stated. The first requirement is one of consistency. Propositions which are believed by the judge or other decision maker must be consistent. This principle holds both in the case of the interpretation of laws and proof of facts.

Another principle is that of conjunction. (Cohen, 1977) If an element A of a case is proved and an element B of a case is proved, then (A & B) is proved. So if A is considered to be the mens rea and B the actus reus, it follows that if each of these two elements are proved then both are proved. The elements of contention about the actus reus and mens rea may require the issues to be split up even further. If this is so, the conjunction principle holds that for all these elements A, B, C ... N, if A is proved, B is proved, C is proved, ... N is proved, then (A & B & C & ... N) is
proved. On a strict reading of the statement in *R v. Graham* the principle only states that it is necessary that all the elements of a case be proved in order for the case to be proved. But a natural reading is that it is also a sufficient condition that if all the elements of a case are proved then the case as a whole is proved.

Deductive closure is the principle that if an element of a case A is proved, and A implies B then B is proved. (Hempel, 1962; Kyburg, 1974; Levi, 1980) This principle is apparently related to the principle that Wigmore called "inference upon inference". According to this principle, for a conclusion in a chain of inferences to be probable to a given degree, each inference must be established with the same degree of probability. (Tillers, 1983, sec. 41)

There is a principle which is hardly ever explicitly stated. It is the assumption that the deliberator's degree of persuasion in the proposition at issue conforms to the objective probability of the proposition being true. This principle was stated as far back as Gilbert's *Treatise on Evidence* in 1754. This is the idea that there is some rational basis to the judgement and degree of persuasion of deliberators. This rule might be considered a basic principle of rationality.

Another requirement that is usually assumed and sometimes stated is that that decisions and judgements in legal proceedings are based upon the complete evidence. (Cohen, 1988) There are several degrees of completeness that can be distinguished. Verdicts can be based upon all the relevant evidence that is available to the proceedings. *(Philips v. Ford Motor Company)* This is strong completeness of evidence. But usually not all relevant evidence is admissible, and so the weaker degree of completeness of evidence requires that verdicts are based upon the evidence that is known to the deliberator. An intermediate requirement is that the evidence which is known to the deliberator possesses sufficient weight. (Cohen, 1977, 1986, Davidson and Pargetter, 1987)

1.2 Theories of Probability and Inference.

A central question in legal philosophy involves deciding what interpretation of probability is the most appropriate for legal reasoning. Probability is a central concept in legal reasoning, and always has been since the birth of evidence law in western law. Probability is used in the formulation of standards of proof, presumptive inference involving causal and statistical
reasoning, the definition of relevance and reliability of evidence, and corroboration of testimony. Here I offer a summary and background to the discussion in the remaining chapters.

Most interpretations of probability assume the following mathematical structure. Probabilities are defined as functions over sets such as the Boolean algebra of the propositional calculus. The set of propositions consists of atomic propositions and molecular propositions which are constructed from the connectives of negation, disjunction, material implication, conjunction, and biconditional. A set of propositions is said to be closed under the Boolean connectives if it contains all Boolean compounds of propositions it contains. A set of propositions is a Boolean algebra iff it is closed under all Boolean connectives.

The standard purely mathematical definition of probability is due to Kolmogoroff. (Fine, 1973) A probability $P$ on a Boolean algebra is a function $P$ which assigns to each proposition $H$ in the Boolean Algebra a real number, $P(H)$ such that:

1. $P(H)$ is greater than or equal to zero
2. $P(H)$ equals 1, where $H$ is a necessary proposition
3. If $(H \& E)$ is a contradiction, then $P(H \vee E) = P(H) + P(E)$

Conditional probability, the probability of $H$ given $E$, is defined as follows:

4. $P(H \mid E) = \frac{P(H \& E)}{P(E)}$

From the axioms and the definition of conditional probability the multiplication principle for conjunction follows:

5. $P(H \& E) = P(H) P(E \mid H)$

The multiplicative principle of conjunction is of central concern in legal reasoning. The principle of conjunction states that if element A, such as the *actus reus*, is proved, and element B, such as the *mens rea*, is proved then the action constituted by A and B is proved. The concern is that admitting evidence by way of a probabilistic standard conforming to the multiplicative principle will not conform to the conjunction principle.
Bayes' theorem is the all important theorem for many philosophers and legal theorists as an inverse inference from evidence to hypothesis. (see Tillers, 1988) From the multiplication axiom a simple form of Bayes' theorem follows immediately:

(6) Bayes' theorem: \( P(H \mid E) = \frac{P(E \mid H) \, P(H)}{P(E)} \)

In general, if one and only one of the hypotheses \( H_1, H_2, H_3 \ldots H_n \) are true, then there is a more general form of Bayes' theorem:

(7) Bayes' theorem: \( P(H_i \mid E) = \frac{P(H_i)}{\sum P(H_j) \, P(E \mid H_j)} \) for all \( i = 1-n \).

Other theorists hold that inverse inferences from evidence to conclusion are made on the basis of an application of Bernoulli's theorem.

(8) Bernoulli's theorem: (roughly) the relative frequency of the occurrence of an event in \( n \) probabilistically independent repetitions of an experiment tends to its probability as \( n \) increases without limit.

The import of this theorem is that one can arrive at a suitable estimate of the probability of a property in a population within the required level of error by choosing a suitable sample size of information.

These axioms and theorems are part of the uninterpreted calculus of probabilities. The theorems proved from the axioms are truths about mathematical objects and as yet, have no connection to any non-mathematical reality. Since the topic of interest here is with legal reasoning an appropriate interpretation of probability is needed. There are a number of possibilities:

Logical: \( P(A \mid B) \) is the degree of entailment between A and B.

On the logical interpretation, probability is a measure of the degree of entailment between evidence admitted and the propositions at issue. Thus it is conceived to be a generalization of deductive logic. In deductive logic \( A \) implies \( B \) iff \( B \) is true in every possible world (model) \( A \) is true in. On the logical interpretation the probability of \( A \) given \( B \) is the number of possible worlds in which \( A \) and \( B \) are true, divided by the number of possible worlds in which \( B \) is true.
Unfortunately the elaboration of this view by Carnap (1950) succumbed to the criticism of arbitrariness.

Subjective: P(A | B) is the degree of belief an agent has in A on the supposition of the truth of B.

The subjective theory, currently popular with philosophers, some statisticians, and a growing number of legal theorists holds that to say that a proposition at issue regarding an accused's guilt or liability is highly probability is to say that one has a high degree of belief in that proposition. The main argument for the use of subjective probabilities comes from coherence and dynamic coherence arguments. According to this argument, a rational deliberating agent must conform his degrees of belief to the standard probability calculus because otherwise he will be a sure loser to a cunning bettor. This argument, although powerful and interesting, has the serious flaw that it fails to apply in situations in which there is no cunning bettor.

Frequency: P(x ∈ B) is the relative frequency of As among Bs, where A is the reference class.

On the limiting frequency conception of probability to say that the probability of x being B is Z is to say that the limit of the ratio of As among Bs is Z where A is the reference class for x being B. A problem with the limiting relative frequency view is that it is not possible to single out a non-arbitrary limiting relative frequency. On the finite frequency interpretation to say that the probability of x being B is Z is to say that A is a member of a finite class B where the proportion of As among Bs is Z, where A is the reference class of x being B. Notice that on this interpretation the probability that x is an A is just an elliptical way of saying something about a class ratio and not anything about x. But in law it is required that probabilities apply to unique individuals.

Chance: P(A | B) is the propensity or chance of A occurring given that B occurs.

On the chance interpretation of probability chance is a physical property of a thing and its environment. According to this view, the mistake involved in the relative frequency interpretation is that the relative frequency view confuses what probability is -- a dispositional property of an object -- with the justification for asserting the probability. On the chance view the relative frequency can be considered justification for asserting the chance of an event. Chance is regarded as a physical property of an event and its system. In the case of law, the chance that A committed act D, could be taken to be the disposition that A would do D in a specified situation.
In order to assert the chance of an event having property A, it is plausible to require that if conditions of the environment are changed, the probability will remain invariant. In this manner chances support counterfactual conditionalization. A measure of invariance under counterfactual conditionalization is the resiliency of probability:

Resiliency: the resiliency of \( P(A) \) being \( a = 1 - \max |P(A) - P(A|Q)| \) for all Q.

Epistemic: \( P(x \in B) \) is the proportion of A's among B's, where A is the reference calls for x being B, relative to one's set of rational beliefs K.

The epistemic theory has been advanced by Kyburg (1974, 1982). This is a theory which is a careful blend of the previously mentioned theories. According to this view probabilities are objective logical measures of the relation between a hypothesis and a conclusion relative to one's belief state and based upon proportions. So the probability that A committed act D based upon evidence E, is an objective judgement. The evidence, E, is regarded as evidence of the chance (disposition) that A had to commit act D in the given situation. Kyburg's theory is appropriate for legal reasoning, I argue, because it is able to provide a good interpretation of traditional legal practice.

In addition to the fundamental question of the interpretation of probabilities in legal reasoning, there is the question of how beliefs are changed in a rational deliberator on the basis of new evidence. The Bayesian school relies on Bayes' theorem and its generalizations for making inferences. The main point of interest here is that Bayes' theorem is thought to provide an appropriate way of expanding the deliberator's set of beliefs. Bayes' theorem is taken to be that rule of expansion. So if evidence E is offered in evidence in favour of a proposition at issue H, Bayes' theorem is used in computing the probability of H, in light of the admission of E.

(9) Conditionalization: \( P(H) = P(H|E) = P(E|H) P(H) / P(E) \)

On this view the deliberator has a prior probability that attaches to the proposition at issue H and simply revises the probability in light of the evidence based upon Bayes' theorem. So the judge, for instance, already has an opinion about the accused and modifies this probability on the basis of the Bayes' theorem. This formula has the problem, though, that evidence admitted must be regarded as certain since \( P(E|E) = 1 \). Thus \( P(E) \) cannot be lowered by from 1 by
repeated applications of Bayes' theorem. It does not allow evidence, such as eyewitness testimony, to be questioned through examination after it has been admitted.

But the appropriate generalization as suggested by Jeffrey (1965) is:

\[ P(H) = P(G) P(H | G) + P(B) P(H | B) \]

One can understand this as a generalization of Bayes' theorem by noting that if the probability of the evidence of B, or G, is 1, then the equation reduces to Bayes' theorem above.

If the deliberating agent is conceived to learn by updating his entire set of partial beliefs \( S = H_i \), for \( i = 1 - n \), then the appropriate generalization of this rule is:

\[ P(H_i | E) = \frac{P(E | H_i) P(H_i)}{\sum P(E | H_j) P(H_j)} \text{ for all } i = 1-n. \]

The main argument in favour of conditionalization by Bayes' theorem is the dynamic dutch book argument due to Lewis and reported by Teller (1976). This argument is that a rational agent will not be put in a position such that if he made a bet he could lose for certain. But if an agent does not use conditionalization when new evidence is received a conditional bet can be made against him such that he will lose for certain. Thus a rational agent will conditionalize.

There is also a dutch book argument for probability kinematics. (Skyrms, 1990)

I offer a contrary view to Bayesian deliberation. A rule of expansion which is more in line with traditional legal deliberation is to revise a set of beliefs by a rule of expansion which conjoins a proposition A to one's set of beliefs K, and adds all the deductive consequences of A to K. (Levi, 1980) The Bayesian rule of expansion cannot satisfy this rule of expansion. I argue in favour of this rule of expansion based upon logical considerations.

1.3 Action.

A person can be found liable only if it is proved that he committed a particular act. Hence there needs to be some conception of what an action is. On causal views of action (the only type I will consider) to do something is to cause something to occur. An action is taken to be a part of the causal sequence. The main question is what part of that sequence is to be regarded as the action. A second important question is what the criterion for non-standard, eg. unintentional,
careless, and involuntary actions, is. The various views can be illustrated by means of a simple diagram. The following diagram represents a sequence of events beginning with M, a mental event causing B a bodily movement, which in turn causes further events.

\[ M \rightarrow B \rightarrow E_1 \rightarrow E_2 \rightarrow E_3 \rightarrow E_4 \rightarrow E_n \]

Figure 1

Davidson (1963) takes the action to be B, the bodily movement. So if I move my finger, and at the same time, flip the switch, and turn on the light, the action is moving my finger. My turning on the light is a description of moving my finger in terms of the consequences which are not part of the action. According to Davis (1970) and Hornsby (1980) the action is M, specifically a volition for Davis and a try for Hornsby. The bodily movement and the consequences are not part of the action. On Thomson (1977), Dretske (1988) and Costa's (1987) view the action is the entire sequence beginning with M and continuing along the causal chain.

The main problem with most causal views is they do not allow for a preferred description of one's action. In the shooting case, A crooks his finger, pulls the trigger, fires the gun and kills B. These are all cases of what A did. None of these descriptions are to be preferred to another on Davidson's view. For legal purposes, of course, one of these descriptions needs to be preferred to the others.

Legal commentators have not done a good job of dealing with issues about action. According to the traditional view of action in law, an event, such as A hitting B, consists of two components: a physical component, the \textit{actus reus} and a mental component, the \textit{mens rea}. Suppose that A hits B. On the Austinian view, the \textit{actus reus}, the "action", is thought to be the bodily movement of moving the arm with the property of volitariness. The \textit{mens rea} is taken to be the mental attitude that A has when A's arm is moving. This partitioning of the event of hitting into the \textit{actus reus} of hitting and the mental element of hitting is serious confusion.

On William's (1953) view the \textit{actus reus} is everything involved in a crime except the mental element. Williams includes in the \textit{actus reus} the actions of others, omissions, things that
happen to one, states of a person, mere behaviour, and legal relations. This definition of action plays no useful purpose as far as I can see. It merely confuses policies with factual descriptions of events of agency. As I have said, the causal view is able to distinguish all of the above elements in a coherent and useful manner.

There is another error that the narrow and wide views of action make. On both the traditional view and William's view mental elements, such as intention, carelessness, recklessness, accident and so on are understood to be positive mental attitudes toward the actus reus. So on the standard account, if A accidently hits B, then accidentalness is understood to be a property of A running down B. So on this account even if A did not mean to run down B, he is thought to have run down B because that is the actus reus. But this actus reus has the additional property of accidentalness.

On my view the action is that part of the sequence that is represented to the agent in his causally efficacious mental state. This definition is neutral with respect to which event is being singled out as the action. The action may be the bodily movement or a sequence of events. But it is required that that part of the sequence be such that it is represented by the agent in his causally efficacious mental state.

A R's iff
i) B is a set of events a, b, c, d, e ... n such that R
ii) R includes the fact that a causes b, b causes c, c causes d ... causes n
iii) a is a mental representation of the events B such that R

The first clause takes R to be the relation between events a, b, c, d...n. This relation will include temporal, spatial, and causal relations. The second clause specifies that the relation R includes a causing b, b causing c causing .... n. The third clause specifies that a represents the events as having a certain relation R. Linguistically "R" is the description of the action.

At a simple level it seems that one only need to take into consideration the attitudes of belief and desire in order to define the representational states for this model of action. On this simple model of an agent higher order mental states are complex desire and belief states. When a complex enough mental state such as an intention is formed an action ensues.
1.4 The Rational Deliberator: Bayesian vs. Non-Bayesian.

The approach I will use in arriving at a theory of legal deliberation is to define the ideal reasoning of an agent. I offer principles which an ideal deliberator would use in legal reasoning. The model of legal deliberation is defined in terms of a deliberational state of an agent. Roughly, a rational deliberating agent is one who has a set of beliefs and desires and a rule for changing these beliefs and desires given new evidence. The deliberative state of an agent consists of a set of elements \(< B, D, S, K, +>\) where \(B\) is the deliberator's degree of belief over a set of propositions \(S\) and \(B(A) = P(A)\) for all propositions \(A\); \(K\) is a subset of \(S\) — the full or accepted beliefs; \(D\) is a degree of desire over propositions; \(+\) is a dynamical rule for changing the values for \(B\), and \(D\) in response to new information.

The most important aspect of the agent's state is his set of beliefs \(K\). On my view \(K\) is the set of rational beliefs, and it is rational beliefs that are taken to be those beliefs which are proved. The dynamical rule must account for three different important changes in the agent's set of beliefs. First it must have a rule for expansion of belief. This occurs when a new belief is added to \(K\). This can occur, for instance when evidence is admitted. Second, there must be a rule for contraction. This occurs when a belief is subtracted from \(K\). This may occur, for instance, when some claim is successfully rebutted, or some presumption defeated. Third, there is the case where beliefs are revised. Some beliefs are added to \(K\) and some are subtracted from \(K\). This may occur, for instance, when testimony of a witness is powerful enough to rebut earlier evidence.

There are a number of different possible rules for the dynamics of epistemic states. As I explained above, Bayesian base their rule of expansion on Bayes' theorem or probability kinematics. An alternative rule of expansion, which I will argue in favour of, is the function \(+\) which is the function of conjoining \(A\) to \(K\) and forming the deductive closure of all propositions in \(K\). This implies the principles of conjunction and deductive closure. The operation of expanding one's epistemic state \(K\) by \(A\) is denoted by \(K^+ A\). The operation of contracting a propositions from \(K\) will be denoted \(K^{++} A\). It is assumed here that revisions can be analyzed in terms of expansions and contracting propositions and their deductive consequences.

On the view I have offered \(B\) and \(D\) are mental states directed toward propositions. The degree of belief in a proposition \(A\) is equal to the epistemic probability of \(A\), and the desire function \(D\), really a set of desire functions, is equal to Jeffrey's desire function as I define below.
On Jeffrey's theory desirabilities and probabilities attach to propositions. One could alternatively think of desirabilities and degrees of belief as attaching to events as described by propositions. I give Jeffrey's axioms as background, and draw attention to two theorems which will be needed in chapter six. One theorem gives the desirability of an act (in terms of the propositions describing the act) in terms of consequences and states of the world. A second theorem is that the expected desirability of a proposition is equal to the desirability of that proposition.

Jeffrey's basic axioms are:

(12) Desirability: \( P(X \lor Y) \cdot D(X \lor Y) = P(X) \cdot D(X) + P(Y) \cdot D(Y) \)

Assume that \( D \), the set of Boolean propositions, is a complete, atom free, Boolean algebra. Also, > (preference) and - (indifference) are relations in \( D \) such that:

(13) transitivity: for all \( X, Y, Z \) in \( D \), if \( X > Y \) and \( Y > Z \), then \( X > Z \).

(14) trichotomy: for all \( X, Y \) in \( D \), exactly one of the following holds: \( X > Y \), \( Y > X \), \( X = Y \).

(15) Averaging: for all \( X \) and \( Y \) in \( D \), if \( X \) and \( Y \) are disjoint, then: (a) if \( X > Y \), then \( X > (X \lor Y) > Y \) and (b) if \( X \not= Y \), then \( X > (X \lor Y) > Y \).

(16) Impartiality: for all \( X, Y, Z \) in \( D \), if \( X, Y, \) and \( Z \) are pairwise disjoint, \( X > Y \) and \( Z \) and \( (X \lor Z) = (Y \lor Z) \), then for all \( W \) disjoint form \( X \) and \( Y \), \( (X \lor W) > (Y \lor W) \).

(17) Continuity: for all \( X_1, X_2, X_3, \ldots \) and \( Z \) in \( D \), if \( (X_i) \) \( i=1 \ldots n \), is an increasing [or decreasing] sequence in \( D \), \( X = (X_1 \lor X_2 \lor \ldots X_n) \) [or \( X = (X_1 \land X_2 \land \ldots X_n) \)] and \( Y > X > Z \), then \( Y > X \), \( i > Z \) for all \( i \) larger than some number.

Then there exists a probability assignment \( P \) and a desirability assignment \( D \) on \( D \) such that for all \( X \) and \( Y \) in \( D \), \( X > Y \) if and only if \( CED(X) > CED(Y) \), where \( CED \) is calculated in terms of \( P \) and \( D \).
And if,

(i) for all $X$ and $Y$ in $\mathcal{D}$, $X > Y$ if and only if $CED(X) > CED(Y)$ where $CED$ is calculated in terms of $P$ and $D$,

(ii) for all $X$ and $Y$ in $\mathcal{D}$, $X > Y$ if and only if $CED(X) > CED(Y)$ and $CED$ is calculated in terms of $P'$ and $D'$,

This theorem defines expected desirability:

$$D(A) = \sum P(S_i | A) D(S_i & A)$$ for all $i=1$ - $n$, where $S$ represents a state, and $A$ a basic act.

In addition to the postulates above, there is an additional postulate that allows there to be a theory about what an agent rationally does. I adopt the view that a rational agent acts such that he maximizes his desirability function.

1.5 The Ethics of Proof.

It is axiomatic that proof must conform to certain standards that the law has. Kaplan (1968) proposed that the choice of standards of proof should be made within the framework of decision theory. The deliberator is to attach a probability and a level of desire to each possible verdict and maximize his expected utility. However this view does not specify what the desire function of the deliberating agent is supposed to represent. It may represent the deliberator's personal desires, the desires of the community expressed through him, a utilitarian based desire or a desire to secure people's fundamental rights.

In the last chapter I explore the interpretation where a judge's or other deliberator's desires are utilitarian. On a utilitarian view the deliberator maximizes the social welfare of society calculated in terms of the average utility level of persons in that society. There are two mistakes that can be made in arriving at a verdict. One mistake is to find a person liable when he is not liable. A second mistake is to not find a person liable when he is liable. On the utilitarian view the value of these mistakes can be assessed in terms of the effect on the average utility in society.

Given these assumptions the preponderance of the evidence standard and the proof beyond a reasonable doubt standard can be derived. In civil proceedings the rule, in general, is to find
liable the defendant with the highest probability of having committed the alleged act. This
standard implies where there are one or two defendants that the probability is greater than .5.
But it also follows, that in the case of multiple defendants there are a wide range of occasions
in which individual rights are violated because utilitarianism implies finding a defendant
liable even though it is improbable that he committed the required act.
Proof of Action

2.0 Proof of Action.

Legal deliberation is concerned, in part, with determining what actions an accused has performed or failed to perform. It is up to the Crown and individuals to prove that certain actions – killing, robbing, defrauding, assaulting and other events or omissions – have occurred. These alleged actions are the facts at issue in legal proceedings. In order to prove that certain alleged actions have taken place there is a need both for a logic of how propositions at issue are proved, and for a conception of what actions are. In this chapter I tackle the second requirement of advancing a conception of what an action is. But, as will become clear, there is no generally accepted theory of action in legal theory or practice. There is no generally agreed upon answer to what doing something is.

This overriding problem, then, is the problem of the nature of action, that is, the problem of determining the nature of action. The lack of agreement of the nature of action creates a lacuna in legal theory. There is a lack of agreement on how to describe correctly the events caused by persons and ascribe actions to persons. The need to give a non-arbitrary, that is, a preferred description of a person’s events in law is due to both practical and theoretical reasons. The practical reason is that the rules of liability are framed in terms of such preferred descriptions. For instance, suppose the straight rule of responsibility — that persons are responsible for all and only their (intentional) actions — is used in assessing liability. Suppose A fires a gun, a bullet enters B and B dies. Then one must be able to determine whether A killed B, A fired a gun at B, A crooked his finger, A inadvertently killed B ... and so on. The theoretical reason is that descriptions of events give a causal explanation of events and this causal explanation needs to be a unique explanation. I shall suggest a view as to the nature of action in law which solves this problem of providing a preferred description.

In the second section I will review the nature of action by looking at causal theories of action. These theories hold that doing is a type of causing. There is a substantial amount of agreement about the ingredients in a causal theory but disagreement over precisely how these ingredients are combined. There is some disagreement over the exact nature of the mental event that initiates the causal process: volitions, intentions, agency, primary reasons, desires, or beliefs. There is also disagreement over what part of the causal chain is the action: the mental antecedent, bodily movement, or further effects. As I suggested, the chief criterion for
evaluating causal theories will be the ability of the theory to provide a non-arbitrary, that is, a preferred description of the events that a person causes.

I will argue that what a person does is part of a causal sequence of events, such that the first event is a mental state of an agent and the causal sequence has the properties that were represented in the agent's mental state initiating the sequence. This view of the nature of action is vague enough to incorporate several causal views of action. That is, it does not single out the part of the sequence that is the action, and it does not single out the nature of the initial mental state. What is important in this formulation is that the action is that part of the sequence that is represented to the agent, and that the initial mental state is a representation. With this formulation of the nature of action the ability to give a preferred description of an event of agency is solved.

I have said there is no agreed upon theory of action in law. But the law does unreflectively waiver between a wide and a narrow view of action. On the narrow Austinian view actions are willed bodily movements. The wide view is that both bodily movements and consequences of actions are included in actions. But the wide view is sometimes extended in an unintuitive way, as by Williams (1953) and adopted by Smith and Hogan (1983) to include omissions, happenings, the actions of others, and legal relations.

In the third section I examine the influential Austinian (1873) account of action in law. On the traditional Austinian account of action to prove that an action has occurred one does not have to prove the existence of a mental element such as intention, recklessness, or inadvertence. On this view in order to prove that an action has occurred there is no need to prove that any mental event has occurred. This is a counterintuitive position because mental events are usually thought to be included in a person's action.

In order to augment the Austinian definition of action the law requires that one must prove a physical component and a mental component, that is, an actus reus and a mens rea. It is then held that the possible mental attitudes are intention, negligence, recklessness; and that the act is a voluntary bodily movement. A defect of this position, to be addressed in the fourth section is that the mental elements are held to be positive properties attributed to the action. When one is accused of careless driving, say, the carelessness is thought to be a property of the driving, whereas it is really a property of one's bodily movements.
In the fourth section I look at attempts to widen the Austinian theory by, among others, Salmond (1957), Williams (1953) and Hart (1949). Salmond's view is quite moderate. He simply asks for a view of action which allows the consequences and circumstances to be a part of the action. I agree with this as far as it goes, but argue that actions must include a mental element as well. Hart's theory of action emphasizes that ascribing actions to persons is a way of ascribing responsibility. This is correct but threatens to confuse factual and value judgements. But Williams goes to extremes by including states of affairs like being found addicted, omissions, the actions of others, legal relations, and elements of legal responsibility. This conflates the basic categories in the ontology of action.

In the fifth section I examine the recent theory of Coval and Smith (1986). Coval and Smith follow the line of reasoning established by J.L Austin (1956) and Hart (1968) that adjusters such as involuntariness attributed to an event show that it is not an action. They go on forcefully to conclude that the properties of action can be uncovered from the conversion of the adjusters. The acceptance of this thesis solves a number of problems which I outline regarding, attempted actions, strict liability actions, incorrect descriptions of events, and the mental/physical dichotomy.

One example of the anti-adjustor view is the problem that results in describing events. This is the view that action adjusters such as involuntariness, attempt, carelessness and so on positively attribute properties to actions. Thus, for example, if it is said that one has negligently caused the death of another by running him down, the action is said to be running him down (on the Austin line) and the negligence is a positive attribute of that causal chain. But on the adjustor view to say that A carelessly ran over B is not to ascribe the running over of B to A, rather it is to deny that A intentionally ran over B. Hence running over B was not A's action.

2.1 Causal Theories of Action.

In recent years there has been a burgeoning of interest by philosophers interested in the nature of action. In this section I try, within the context of legal examples, to outline the basic ontology of action by distinguishing between events, actions, omissions, things that happen to one, and states of a person. Secondly, I review the main views on what part of the causal chain ensuing from the antecedent mental state is the action.
There is an intuitive distinction between things that happen to someone and things that one does. The basis of the distinction is that in the case of something happening to you, the (primary) cause is external to you; whereas if you do something then the (primary) cause is internal to you. (Dretske, 1989, Ch. 1) The law does recognizes such a distinction but sometimes wrongly classifies it as involuntary action. If A is carried by B onto C's land then this is something that happens to A and not something A does. A has not trespassed on C's land. (Smith v. Stone) But it is not correct to say that this is because the action of A was involuntary, as is sometimes said (eg. Linden, 1982, 32; Fleming, 1977, 3) but rather because this is not A's behaviour at all. A is in no sense the cause of his being on C's land.

If A swings B's arm which strikes C, B will not be found liable. (Weaver v. Ward) This is because B did not strike C. This is because the cause of the hand hitting C's face was not internal to B, but external to B. In fact, A hit C because the cause of the slap to the face is internal to A. A could be found liable for hitting C. This can lead to interesting results. In the trespass case, where A carries B onto C's land, A may have permission to be on C's land but B may not have permission. Perhaps A might be considered to be trespassing in virtue of carrying B, who is not welcome, onto the land.

There is another distinction that merits notice. Notice that there are some things we do such as shivering, coughing, trembling, stammering and so on, which are different than running, writing and other actions. (Thalberg, 1972, ch 1) The former type of behaviour does not require any thought process or mental representation whatsoever to occur. These events simply happen and might be called, simply, "behaviour". (Dreske,1989, ch. 1) From henceforth "action" will be distinguished from "behaviour", and "doing" will be distinguished from "behaving".

The distinction between behaviour and involuntary action is sometimes blurred in law. Suppose that A strikes B with a rock for no apparent reason as in R v Rabey. The evidence shows that A was in an unconscious state when he struck B with the rock. It might be thought that this is simply an unconscious involuntary action and "hence" still an action. This idea is often put forth. For instance Zalev Co. Ct. J. stated that "simply put, automatism is unconscious involuntary action, that is, doing something without knowledge of it and without memory afterward of having done it." (Mewett and Manning, 1982, 243, my italics) But what is the action? In cases of involuntariness due to a gun at one's head the action is saving one's life. But there is nothing A can be said to have done simpliciter. All that can be said is that A's body moved. This case is one of mere behaviour only and no action in any full fledged sense has occurred.
The causal theory also is able to distinguish between actions and omissions. When someone acts there is a causal sequence of events beginning with a mental event and extending through one's bodily movement and consequences of the movement. There may be cases where through an agent's failure to act a certain event occurs, although there is no causal relation between the mental state of the agent and the subsequent event. In *R. v. Lewis*, a parent, who was a Christian Scientist, withheld medical treatment from his son. The son, who was very ill, subsequently died. There was no causal connection between the bodily movements or mental states of the parent and the death of the child. It was the disease of the child that killed the child, not the inaction of the parent. The parent omitted to do anything for the health of the child. However the law is able to hold people responsible for *not* exerting a causal influence over events.

The most influential picture of a causal theory of action is due to Davidson. Davidson holds that the causally efficacious mental antecedent to an action are primary reasons. (1963, 4) Primary reasons consist of beliefs and pro attitudes such as desiring. I want to concentrate on what action is according to Davidson, that is, whether action is bodily movement, mental events, neural events, volitions, consequences, or something else. I suggest that Davidson's view that bodily movement is action leads him into trouble because he is unable to give a preferred description of a given action. The ability to give a preferred description of an event is a necessary minimum for a useful theory of action in law.

The requirement of a preferred description has an explanatory purpose. The action description is simply the manner in which the causal explanation for a sequence of events is given. Suppose that A is charged with murder. The description of an event as a murder is an explanation of the event in terms of the properties of the initiating representational state of an agent (he intentionally caused the death) and the identical properties of the death event (it was intentionally caused by A). The existence of a preferred description is due to the fact that there is an essential relation of identity between the action and the representation state of the agent who caused the event described. (S.C. Coval, pers. comm.)

One reading of Davidson is that action is bodily movement which is caused by one's primary reasons. This is an event which is intentional under some description. Thus if I move my finger, flip the switch, turn on the light, and alert the burglar, the action is the movement of my finger. There is only one action but the action of moving my finger can be redescribed in terms of further consequences. This is referred to by Feinberg as the accordion effect, because one can
squeeze more or less of the consequences into the description of moving my finger. (1970, 134) On my interpretation the accordion effect is that each of these descriptions of the event in the above example is true. The problem, then, is to reconcile the truth of multiple descriptions with existence of one action. Which description of the event is the preferred description of the event and not a redescription in terms of causes or effects of the event?

The view that there is only one action involved in the preceding example motivates the idea that different descriptions are redescriptions of the same event. The ability to give a preferred description might be thought to be solved by this redescription manoeuvre. But it appears, after greater attention, that Davidson's theory does not allow that any description of the event is the preferred description of the event. So he is unable to give a non-arbitrary, preferred description of an event which is considered the action.

Recall that one reading of Davidson was that bodily movement is action. Accordingly, the description in terms of bodily movement should be thought of as the preferred description of his event of agency. But Davidson's treatment of the accordion effect would seem to exclude the bodily movement from necessarily being the action. Internal neural stimulation occurring prior to the movement of my finger is intentional under the description of turning on the light. Hence these neural events are actions as well. This seems to lead to the idea that there is nothing that counts as the action. Each differing description of an event is the same event under a different description. Thus Davidson's theory does not allow for a preferred description of events of agency.

One can be led to a different view of action by examining the case of a paralyzed person. If a person is paralyzed and tries to raise his arm there is no bodily movement. So the only candidate for an action is the mental state or a proximately caused neuronal state. One might want to identify the action with the proximately caused event of one's causally efficacious mental state. On this view the action is the immediately spatio-temporally related event to the initial mental event. One might be impressed with this line of reasoning and adopt the view that all actions are, strictly speaking, attempts. The action is not even the proximately caused event, it is the mental event itself. Thus, on this view, a number of descriptions of an event are true, but the description in terms of the mental event is preferred because it is the only necessary event for action.

But a theorist who takes another part of the causal sequence to be an action, such as a bodily movement could reply to this line of argumentation by denying that attempts or tryings are
actions at all. Trying to wiggle one's toe when paralyzed is not an action. If the trying is successful the wiggling of a toe results, but the trying is not itself the wiggling of the toe. An (unsuccessful) attempt might also be thought of as an adjuster as I argue below. On this view to describe a sequence of events as an attempt is to say that the intended action failed. To describe the paralyzed person's mental events as an (unsuccessful) attempt is to describe why these events were not actions.

According to the view that actions are attempts the killing of Abe Lincoln was, strictly speaking, the attempt to kill Abe followed by Abe's death. But then the killing of Abe is just a description of the event, the initial mental state, in terms of its consequences, that is, pulling the trigger, firing the gun, hitting Abe, and the death of Abe. In essence one argues that, A did not kill Abe, he only attempted to kill Abe and the rest of the events just happened. "Attempting to kill Abe" is thought as the preferred description because the attempt seems to be the only necessary part of a successful action. But there is no reason to consider the mental state the only necessary part of the causal sequence. The death of Abe might equally well be considered a necessary component of killing Abe.

The causal process view of action, associated with Thalberg (1972, 1977), Thomson (1977), Dretske (1988), and Costa (1989), emphasizes that an action is a causal process, not a single event, but a sequence of events. The explanation of the accordion effect on Davidson's view is that we can redescribe our bodily movements in terms of the consequences of those actions. But, as was seen, this redescription ploy leads to problems. Davidson has no basis for singling out one description as the preferred description. One can also describe an event in terms of the antecedent, neuronal events, or antecedent mental event, as in attempts. This allows for a preferred description but at the expense of saying things like the death of Abe Lincoln isn't a necessary part of his killing.

On the causal process view, the solution to the problem of the nature of action is to consider actions not as a single event but as a causal sequence of events. One then identifies action with the entire causal sequence. There is no trouble, prima facie, with gaining a preferred description of the event because the description in terms of the entire sequence is the preferred description. A redescription is simply describing a part of the sequence. If the action is A killing B, then the accordion effect does not hold. A pulling the trigger is not the entire sequence. One could redescribe the event in terms of pulling the trigger, but it is an incomplete and misleading description of what occurred. To sat that A pulled the trigger, in the above case is thus not a preferred description on the causal process view.
But the identification of the action with a causal process is not sufficient to give a preferred description of the event because there is no non-arbitrary boundary to the process. So the causal process view is stuck with the problem of singling out the part of the causal sequence that is the action. On my view the proper part of the causal sequence to pick out is that part of the sequence which is represented in the causally efficacious mental state of the agent. So if the agent A is representing to himself as killing B, that is he has the representation of himself crooking his finger, pulling the trigger, the bullet firing, B being struck by a bullet and consequently dying, and this mental representation initiates this chain of events, then A killed B.

2.2 The Narrow Theory: the Classical Austinian View of Action.

In this section I examine the classical Austinian view that an action is a willed muscular movement. This view has the inability, like the Davidsonian view, to deal with picking out a preferred description of an event. In Austin's sense an act is a muscular movement that is willed. Austin distinguished between the objects of the will, which are actions and objects of intention which may be consequences of the act but are not part of the act. Thus Austin said: (1873, 1961, 120)

Most of the names which seem to be the names of acts, are names of acts, coupled with their consequences. For example, if I kill you with a gun or pistol, I shoot you: And the long train of incidents which are denoted by that brief expression, are considered (or spoken of) as if they constituted an act, perpetrated by me. In truth, the only parts of the train which my act or acts, are the muscular motions by which I raised the weapon; point it at your head or body, and pull the trigger. These I will. The contact of the flint with the steel; the ignition of the powder, the flight of the ball towards your body, the wound and subsequent death, with the numberless incidents included in these, are consequences of the act which I will. I will not those consequences, although I may intend them.

This theory appears to be a causal theory and takes the initial mental state to be a willing of events. This is certainly how it is usually interpreted in law. Austin (1873, 1961, 118-19), however, thought that there was no such thing as a will distinct from simply a desire. Austin thought that willing was nothing more than desiring. Austin's conception is not perfectly free from ambiguity. A willed or voluntary muscular movement could be a process of willing followed by a muscular movement. But this has not been the favoured interpretation. A second interpretation is that the voluntariness is a property of the muscular movement. Thus the
muscular movement is an action if it has the property of being willed. This is how action is often interpreted in law. Here again, though, there is another ambiguity of "being willed". Is a muscular movement willed as a result of being caused by the will or is a muscular movement willed as a result of being the intentional object of the will? Austin was unclear on this point.

There is, in Austin's example, a long train of events beginning with mental events causing muscular contractions, a firing gun, an impacting shell, and a death, and even further consequences presumably. He notes that we call the sequence from the mental event to the death event. However Austin holds that the act is the "muscular motion by which I raise the weapon". (1873, 1961, 119) Thus Austin's view is much like Davidson's in holding that the action is a physical movement of the body. But Austin has to be able to give a preferred description of the event of agency as well. And it would appear, much like Davidson's case, that choosing any one event in the causal sequence and calling it the action, or redescribing the event in terms of that one event in the sequence is arbitrary.

Austin's identification of action with the willed muscular movement suffers from the same problems that Davidson's model does. Recall the accordion effect, that one can squeeze into the description of an event various events. Thus the above event can be described as crooking of the finger, or pulling the trigger, or killing. According to the accordion effect, each one of these descriptions is true. Davidson would argue that there are not four actions, moving my finger, pulling a trigger, shooting a gun, and killing someone. There is only one action, and four different descriptions of the event. But one of these descriptions needs to picked out as primary, or preferred. As I argued above, this can't be done on the theories advocated so far.

Austin tries the ploy of denying the accordion effect (in a way). He says in his example, to suppose I kill you, but he doesn't really mean it, for Austin holds that only one description is the "true" description; this is the description "crooking his finger". The description of the event as a killing is false. Austin has the problem of accounting for the standard view that the accordion effect holds. Another problem with this is that, like Hornsby's view, Austin is committed to denying that a death is a necessary part of killing. The only true description of A killing B is that A crooked his finger. Thus it is inconsequential whether B dies in order to say that A killed B.

Let's apply Austin's and Davidson's doctrines to an example. Suppose that A is threatened with death by B to participate in a robbery as in Dunbar v. The King. A then assists in robbing a bank in order to avoid being killed. Has A participated in the robbery or merely saved his life?
According to Davidson there is no one true description. One can describe the event in terms of the muscular movement only, or the movement plus the consequences. But this will not do for legal purposes (or for any other). For then all one can say is that A is criminally liable under one description and not under another description. But some such description must be decided upon because A is either held liable or not. On Davidson's view there is no method of picking a preferred description. Austin's answer is not much better. Austin apparently denies that accordion effect. The true description of A's action is that he moved his body in a certain sort of way. It is not even true, on Austin's theory, that A robbed the bank. A merely crooked his finger.

Hart has argued against Austin's particular version of this view on the basis that we very rarely will our muscles to contract. (1968, 101) There might be special cases in which we will our muscles to contract as in the case of working out in the gymnasium. But in general no such thought occurs in ordinary action. Hart is making the important point that what in fact we do must have some connection with the properties of the mental states that initiate the action. The action, I would say, must have the same properties as was represented in the initiating mental state of the agent. So Hart is correct that the action could not be muscle contraction because we rarely want, desire, will, muscle contraction when we are doing something. The same criticism holds for Davidson's account. We may want some muscular movement in general when we are doing something. But we certainly do not will some precise muscular movement.

Hart makes the point that when we do something like tie our shoelaces, we do not have a clear and distinct mental representation of tying our shoelaces. Thus the entire action may not include every aspect of the muscular movement but only those aspects that are represented to us in our mental state. If I mow the lawn the mowing of the lawn may not include the crushing of a bug on the lawn despite its being part of the causal sequence initiated by my mental state. It is not part of the action because it was not part of the causal sequence that was represented to my causally efficacious mental state.

Although Austin's definition of action is constantly referred to, it is clear that in legal practice it is rejected. Austin's definition of action is in keeping with the tradition in law, up until the twelfth century (Smith and Hogan, 1982, 47), that a mental event was not a part of an action. This contradicts an intuitively plausible criterion of responsibility, that one is responsible only for one's (intentional) actions. One can be proved guilty on the basis of a bodily movement, regardless of ones mental attitude; regardless, that is, of whether the antecedent attitude was one of intention, mistake, carelessness or so on. So the law adopted the doctrine that an act does not make a man guilty, unless his mind is also guilty. Hence two elements are
usually necessary for proof, a mental element and a physical element. So in practice if A is accused of killing B, the criterion that is used for proving this action is not the Austinian one.

Rather, Austin's view was melded into a strange mental/physical dichotomy view of action that exists today in law. The physical part of the action is considered to be a willed bodily movement following Austin. Apparently, being willed is taken to be a property of the bodily movement. This makes sense only, as far as I can see, if the willing is a relational property that the bodily movement has to the actual prior mental antecedent of willing. But other mental states such as intention and negligence are not considered to be relational properties of the act but part of mens rea. Although this doctrine holds generally in law there are cases when the law holds that mental states are part of the action. If one carries an offensive weapon, then part of the act is considered to be the intent to offend. (Smith and Hogan, 1982, 397)

This mental/physical dichotomy has the further problem, which will be discussed in detail in section four, that the wrong actions are attributed to persons. The mental/physical dichotomy has coupled with it the view that a person's mental attitude, which is the mental element, is directed toward the physical element which is his action. Thus if A is charged with negligently running over B, the mental element of A is said to be negligence and this mental attitude is said to be directed toward A's action of running over B. (Smith and Hogan, 47) Thus apparently in order to convict A one is supposed to prove that A had a negligent state of mind and that A ran over B. But the point is that A did not run over B, he only carelessly ran B down. If he had run over B it would be murder. A's carelessness is directed toward the running over of B, but this is not A's action. A's action was driving.

Perhaps the best reason for rejecting the Austinian view is the spatio-temporal problems that plague it. (Hornsby, 1980) Suppose that Bennett tells his broker in British Columbia to sell his shares, and his broker then forwards the order to the Toronto Stock Exchange where they are sold. (R. v Bennett, Doman and Bennett) Where has the act been committed? Have two actions been committed, namely selling shares in B.C. and selling shares in Ontario? On Austin's view, the act is the bodily movement, the bodily movement associated with placing the order, with the property that it was willed. But it leads to the strange property that Bennett has traded his shares before the shares were traded since he talked on the phone before the shares were traded.

Similarly consider the following treatment of the spatio-temporal spread of action. President Garfield was shot in Washington D.C. in 1882 and carried into New Jersey where he
subsequently died. \textit{(U.S.}\textit{v Guiteau)} Where did the killing take place? It was held that the court in Washington D.C. had jurisdiction. But why is that? If it had jurisdiction on the Austinian thesis that the bodily movement was in D.C. then Guiteau's proper defence is that he was being tried for an act preceding Garfield's death. Secondly, Guiteau could have argued that the only act that that occurred in D.C was his crooking his finger into Garfield which, causing a certain amount of immediate harm, is only assault and battery.

On the view I have advocated an action is that part of a sequence that is represented in the agent's causally efficacious mental state. The temporal problem is solved by noting that the event must have the properties that were represented by the agent's mental state. Thus if the agent represents the phone call to his broker as initiating a sequence resulting in the transfer of shares, the phone call only has the property of causing the transfer if transfer is later completed. On the causal process view the temporal problem is solved simply because the action is the part of the sequence which includes the phoning, and the effects of phoning as represented by the agent. Thus the shares are in the process of being transferred during the process and finally transferred once the process is complete.

The spatial problem is resolved differently by these two theories. If one now identifies trading shares with talking on the telephone (represented as causing the further effect of share transfer) then the trading has taken place in B.C. while the effects of trading have taken place outside of B.C. On the causal process view, the trading takes place across provincial boundaries. Guiteau killed Garfield in Washington when killing is pulling the trigger (represented by Guiteau as causing Garfield's death). On the causal process view Garfield is killed in both Washington and New Jersey.

2.3 Wide Theories of Action.

Wide theories of action include more in the \textit{actus reus} than did Austin. Wide theories of action have arisen for a number of reasons. First, the inability of the traditional Austinian account to provide for an adequate solution to the preferred description of the event of agency given the accordion effect has lead to widening of the notion of action in law, to include consequences and circumstances surrounding actions. As was noted above, this manoeuvre does not solve the problem. Second, Hart noticed that ascriptions of actions, often perform a function of ascribing responsibility. This led to the idea that ascribing actions includes ascribing responsibility. Finally, William's (1953) notes that there are many things that need to be
proved in legal proceedings, the actions of others, legal obligations, omissions. He argues that these elements constitute part of action.

The tendency to enlarge the concept of *actus reus* beyond an Austinian sense has resulted in much confusion. Perhaps the worst tendency of accepting a wider view of action is the ability of legislators to deem things which are not actions to be actions. An example of this is to "transfer intent" from one action to another action. To take an example of transferred intent, it is thought that if A attempts to hit B, and B ducks, and A hits C instead the intention of A to hit B is transferred to A's hitting C. Thus A is considered to hit C intentionally. *(Carries v. Thompson)*

Of course, there is no such thing as transferring intent. It is simply a legal fiction. In this way facts, such as whether A hit C intentionally are confused with policy, such as holding A strictly liable for unintentionally hitting C. The wider definition of action, which extends "action" far beyond an ordinary notion of "action", provides some rationale for this dangerous trend of the application of legal fictions.

According to the causal process view the most adequate view of action was to include the whole causal process of events beginning with and represented by the causally efficacious mental state of the agent as an action. This was in order to solve the accordion effect, and the arbitrariness of descriptions of events which resulted therefrom. The problem with this view was to give a non-arbitrary boundary to the effects considered to be part of the action. Salmond has suggested something like the causal process view in law. However Salmond does not, apparently, include the causal chain from the mental event to the bodily movement as constituting part of the action: (1917, 402)

This limitation of identifying an act with a muscular contraction, however, seems no less admissible in law than contrary to the common usage of speech. We habitually and rightly include all material and relevant circumstances and consequences under the name of the act. The act of murder is the shooting or poisoning of his victim, not merely the muscular contractions by which this result is effected.

Many legal theorists have accepted Salmond's recommendation, at least in part. Some, however have gone even further. Williams' wide view of action starts from the premise that the *actus reus* is defined as all the elements of a crime, tort or contract, except the mental element. (1953, 16) This differs from the Austinian act in many ways. First it is held to include states of affairs, omission, other peoples actions, legal relations, and even the jurisdiction. Quite obviously this changes the notion of an act far beyond its ordinary meaning. It is not clear whether Williams thinks that he describing what an action is or proposing some new object to
take its place. But I see no useful place for Williams' "action". What is required is an extension of what type of events for which person can be held liable, not an extension of the domain of action.

An example, according to Williams, of an action including another person’s action is rape. The action of A raping B includes, it is said, the non-consent of B, a speech act of B’s. But this makes B’s non-consent something that A does. But the speech act of B cannot be said to be A’s act because B causes the speech act of non-consent. William’s view, then, denies the basic part of the causal view of action that an action of A’s is a set of events caused by internal states of A.

The wide view of action also motivates the prominent idea that being found in a certain state is an action. These are the so-called “status offences”. The extension of action to include states of affairs, or mere events, or things that happen to one is to account for the existence of status offences. Suppose, as in Robinson v. California that one is prosecuted for being an addict. It is clear that a distinction can be drawn, on the view I advocate, between being in a state and doing something. One can be in a state without having caused that state, or intentionally caused that state. But, it is argued by Williams, that since a crime requires an action, “action” must be redefined so as to include states of affairs. Otherwise one cannot be prosecuted for status offences.

There is no reason to include states of affairs as actions in law. First, it questionable whether status offences really exist. It is doubtful that the law holds people responsible for things that happen to them. The conviction of Robinson, in Robinson v. California for being an addict was found unconstitutional because punishment for being in a certain state was held to be cruel and unusual punishment. Even supporters of states of affairs being part of the action allow that “Even 'state of affairs' offences ought to require proof that D either caused the state of affairs or failed to terminate it when it was possible to do so.” (Smith and Hogan, 42) In other words, convicting a person ought to require that the person did something, or that someone he is responsible for did something. Even if one could be convicted for being in a certain state, there is no need to call this state an action. Rather the solution is to just admit that people will be found liable for things which they may not have caused or intended to happen.

Omissions are also considered to be part of actions by some. In law these events are often referred to as “acts of omission”. On causal views of action omissions are not actions. Actions require that a sequence of events is cause by an agent’s internal states. But omissions are the opposite of actions, they are inactions. An omission of A is an event which is not caused by the
internal states of A. But there is no reason to enlarge action to include omissions. If a person does not come to the aid of another injured person, then he has omitted to aid that person. One may be found liable for this omission without regarding it as an action. Thus there is no need to extend action to include omissions.

A further element of the widening concept of action was due to Hart. Hart has emphasized that to ascribe an action to someone is to ascribe responsibility to him. (Hart, 1968) Although Hart (1968) has abandoned this view it has a kernel of truth in it which should be explored. The kernel of truth is that ascriptions of action sometimes do double duty as both an ascription of action and by implicature an ascription of responsibility. To accuse A of stealing money from B, for example, both ascribes the act of taking money from B to A, and also carries the implicature that the money does not belong to A. On a straight rule of responsibility, that one is responsible for all and only his intentional actions, an action ascription will imply an ascription of responsibility. But the straight rule is not always the rule that is used in law or morality. In case of strict liability, one is responsible for unintentional actions. Hence, an ascription of responsibility does not always follow from an ascription of action.

Action ascriptions and ascriptions of liability are systematically blurred in law. In the preceding case there are two elements which are blurred together. The factual element is that A took the money. The legal element is that if A took money which was not his, then he is guilty of theft. Yet the law often fails to make these important distinctions. But for clarity and understanding of what the facts at issue are it is necessary to separate the factual matters from the legal matters. Omissions, intentional action by way of transferred intent, and having something happen to you are not actions. Neither is an ascription of an action necessarily an ascription of responsibility.

It has been found that there is little need to extend action to include actions of others, transferred intent, omissions, mere behaviour, things that happen to a person or states of a person. In each of these cases the law only needs to recognize that people can be held liable for events other than their actions.
2.4 The Coval-Smith Theory of Action.

Coval and Smith have argued that there is a single view of action operating in the law and that they have identified it. They have offered a quite novel approach to action in law. Coval and Smith (1986, Ch. 1) have argued that action concepts are primary concepts which enable us to understand well-known and recurrent features of the world. There also exists a secondary set of concepts that mark deviations from the normal recurring conditions of an action. Following J.L. Austin, one could say "no modification [of action] without aberration". (1961, 12) Or as Hart has put it, to say that an action is involuntary is to say that "... some radical defect is present, and some vital component of normal action is absent ...". (1968, 104) According to this view, the concepts of accident, inadvertent, mistake, careless, unintentional and involuntary, qualify the description of something as an action. In this way they provide reasons for considering something not to be an action.

Then the argument proceeds as follows. If the adjustors mark the aberrations of action, then normally the negation of those qualities must be present in an action. Thus an action (unadjusted) is voluntary, intentional, done with care, the consequences are foreseen, the causes and conditions leading to the final goal are recognized. To say that someone has done something, then, is to say that all these abilities were in effect when the event was caused.

The Coval-Smith view (henceforth CS) has some interesting repercussions in law because, as I shall argue, many influential legal theorists hold an anti-adjustor view. The trouble begins when the ordinary doctrine of requiring a physical and mental element in crime, tort and contract is unfortunately coupled with an anti-adjustor view of the mental and physical elements. Succinctly put, the mental attitudes of negligence, and recklessness are taken to be positive attitudes with respect to the actus reus. (eg. Smith and Hogan, 47) This contradicts the adjustor view which holds that negligence and recklessness as well as mistake, inadvertence and involuntariness are negative attitudes with respect to the action. They indicate why the object of the mental attitude is not an action. Thus I think that the CS view is right about action, and even right about unselﬁcious uses of it in law. But many theorists, and judges do consciously hold an anti-adjustor view.

A good example of the anti-adjustor view is given by attempts. I think a case can be made for attempt being an adjustor. Although attempt is sometimes used in the context where attempts can be successful, in legal contexts accusations of attempt are used when persons are unsuccessful. When someone has been charged with attempted murder, I think, generally "unsuccessful
attempt" is meant. Thus, in this sense, to say that A attempted to kill B is to give a reason why A's action failed. The reason why the action failed might be because of carelessness, mistake or otherwise, but it needn't.

Some believe that the notion of attempt in law does not imply unsuccessful attempt. Smith and Hogan say, for instance, if A is convicted of attempted murder of B and then B subsequently dies as a result of injuries inflicted by A, A can be convicted of murder without the attempted murder conviction being invalidated. (Smith and Hogan, 266) This seems to be the English view of the matter as stated by the 1967 Criminal law act (sec 6.4): "... where a person is charged on an indictment with attempting to commit an offence or with an assault or other act preliminary to an offence, then ... he may be convicted of the offence charged notwithstanding that he is shown to be guilty of the completed offence." But on the adjustor viewed the attempt is a negative attribute.

An example of the application of the view that attempt is not an adjustor is the case of Webley v. Buxton. A was sitting on a motorcycle pushing it along with his feet about eight feet across a pavement. He was charged with attempting to take a conveyance for his own use without the consent of the owner. The judges satisfied themselves that A had successfully taken the motorcycle, but nevertheless convicted A of attempting to take a conveyance. If attempt is viewed as an adjustor, however, this judgement is wrong. To say that A attempting to take the motorcycle means that he was not successful.

This same anti-adjustor type of reasoning is implicit in cases such as Partington v. Williams where a person was found not guilty of attempted theft because there was no money in the wallet he had taken. The crux of this reasoning, I think, is that there can be no real attempts to do what is physically impossible because one can not have a mental state directed toward something which is physically impossible. A's attitude of attempt is taken to be directed, on this view, toward the taking of the money in the purse. But since A's mental state could not be directed toward taking the money in the purse, since there was no money in the purse, he couldn't be guilty of an attempt.

On the adjustor view however, attempt is taken to be synonymous with not being successful. In the purse example, to say that A attempted to steal the purse implies that he was unsuccessful. Hence his action was not taking the money in the purse. His action was simply grabbing the purse and looking for money. He did however, attempt to steal the money in the purse.
One can get a better idea of the difference between the adjustor and anti-adjustor view by diagramming them. Take an example of attempting arson, but failing because the match won't light. On the anti-adjustor view attempts are modelled as follows:

\[
M \rightarrow B \rightarrow E_1 \rightarrow E_2 \rightarrow E_3 \rightarrow \ldots \rightarrow E_n
\]

Figure 2

where  
B, striking the match is the attempt to burn down the house, 
E₁, the ignition of the house, 
E₂, house engulfed in flames, 
E₃, the house is burnt down, are the intended consequences of B 
E_n, further consequences

It is easy to see how on this view every action necessitates an attempt because the action of attempting to burn down the house precedes the burning down of the house.

\[
E_n^* \\
E_2^* \\
E_1^* \\
M \rightarrow B \rightarrow E_1 \rightarrow E_2 \rightarrow E_3 \rightarrow \ldots \rightarrow E_n
\]

Figure 3

Here the action is the sequence (M, B). The attempted action was (M, B, E, E₁ ... Eₙ). But of course (E₁, E₂, E₃, ... Eₙ) did not occur. Hence the characterization of attempt indicates why it was not the action. (M, B, E₁*, E₂*, ... Eₙ*) is an unintentional action. These events did occur.
The characterization of unintentional refers to the fact that the events were not represented by the agent in his causally efficacious mental state.

Another instance of the anti-adjustor thesis emerging in legal thought is with the problem of voluntariness of actions and strict liability offences. There has been thought to be a problem with proving strict liability offences, where a strict liability offence is one where at least some of the normal mental elements need not be proved. Turner has argued that voluntariness must be considered part of the mental element and not the physical element. (Smith and Hogan, 159) Turner gives the following argument against the Austinian conception of action. Suppose one is attempting to prove a strict liability offence. These offences are thought to allow one to be held liable for an involuntary act. But if the very nature of an act is that it is voluntary, as in the Austinian conception, it is never possible to prove that someone involuntarily acted because by definition acting is voluntary. Thus no one can ever be held liable for an involuntary act.

This argument is side stepped by Smith and Hogan, who differ as to the definition of "strict liability." Strict liability is not, they say, the lack of a requirement to prove a mental element. Rather it relaxes the requirement to prove a certain mental element. But this does not touch Turner's argument. What if the law wanted to hold someone absolutely liable as in the case of R. v Sault St. Marie. In this case a municipality was held liable for the behaviour of a contractor who disposed of waste into the water supply. No defence was allowed. Thus in this type of case there is no need to prove a mental element. But how could the municipality have been found to have polluted the water if there was no actus reus because of no voluntariness? Turner's argument seems to go through for this example.

But again, the problem is that involuntariness is directed toward the action of polluting the water, but this was not the action of the municipality. The fact that the act was involuntary signals that the city did not pollute the water in a full sense of polluting. What they did was, quite obviously, hire a contractor who, unknown to them, polluted the water. This is the actus reus, not the municipality polluting the water.

In conclusion the wide view if action leads to a series of confusions and solves no problems regarding the nature of action.
2.5 Law and Descriptions of Events.

I have argued for a causal view, that what an agent does is a causal sequence of events, such that the first event is a mental state of an agent and the causal sequence has the properties that were represented in the agent's mental state initiating the sequence. When we say that someone has done something, what we are doing is ascribing to a person that part of the sequence. The following definition can be put forward for action:

A R's iff

i) B is a set of events a, b, c, d, e ... n such that R

ii) R includes the fact that a causes b, b causes c, c causes d ... causes n

iii) a is a mental representation of the events B such that R

The first clause takes R to be the relation between events a, b, c, d ... n. This relation will include temporal, spatial, and causal relations. The second clause specifies that the relation R includes a causing b, b causing c, causing .... n. The third clauses specifies that A represents the events as having a certain relation R. Linguistically "R" is the description of the action.

The model of action proposed can be related to the model of a deliberating agent that is proposed. On a simple model of an agent, an agent has a set of partial beliefs and an intensity of desire that range over propositions. His partial beliefs satisfy some sort of probability calculus and his desires some sort of desirability calculus. The degree of belief of some of the partial beliefs is high enough that they are beliefs simpliciter. For the purpose of modeling action the representational state of the agent can be taken to be the agent's set of beliefs and his set of desires.

Suppose again, that A crooks his finger, pulls a trigger, fires a gun, and the bullet enters person B and B dies. Has A killed B? The answer is that if the mental event which began the sequence described was a representation of the elements of the sequence described, then A killed B. According to the causal view I have given, the action is the set of events that satisfy the properties that are represented to the agent in his causally efficacious mental state.

Suppose that, in the example above, A was hunting and he mistakenly thought B was a moose. The mental state that initiated the action consisted of partial beliefs and desires towards propositions. A's belief that A was a moose was false. A believed he was crooking his finger, pulling the trigger, firing the gun and that he was killing B. But he thought B was a moose. Thus the killing of a person was not part of his mental representation. So the action was
not a killing of a person B. Certainly it was not a killing of a moose either, since no moose was killed. The sequence that satisfies the definition of action is the shooting of something A believed to be a moose. This is what A did.

What did A do by mistake? This question may be taken to ask what A's mental state was in relation to the death of person B. A's mental state was one which included the belief that B was a moose, which was mistaken. And so A killed the person B, by mistake. This point makes clear that mistakeness is not a positive property of killing the person B. It is an adjustor. On the causal process view, A did not kill B. To say that A killed B by mistake, is to imply in fact that A did not kill B.

2.6 Conclusion.

I have argued that legal deliberation is concerned with persons' actions. Proof of action is a matter of fact and not a matter of ethical or legal considerations. Failure to recognize this distinction creates problems for proof procedures. The central problem is that there is no coherent theory of action imbedded in law. This leads to the problem of a non-arbitrary, preferred description of someone's event. A preferred description is needed for it is through descriptions of events that liability is imposed upon persons.

The causal theory of action advocated by Davidson (1963) fails to provide a preferred description of actions. Hornsby's theory (1980) defines actions as tryings. This allows for preferred descriptions but denies the plausible view that death is a necessary part of killing. Austin's view holds that the one true description of action is in terms of a person's bodily movement. This suffers form the same problem as Hornsby's account. Williams (1953) and others have adopted a wide notion of action which includes relational properties of mental states, but not the mental states themselves, all legally relevant facts to the case, the actions of others, omissions, mere behaviour and assessments of responsibility, fictitious mental states and so on. This view of action distorts the proof process and confuses facts and law.

Coval and Smith's theory of action recognizes that some properties of actions are not positive properties. This view of action ascriptions is consistent with the causal view. On the causal view an action is a part of a causal sequence initiated by a mental state of an agent. The action is that part of the sequence whose properties are identical with the representational content of the agents initiating mental state.
Interpretations of Probability in Legal Reasoning

3.0 Introduction.

In this chapter I examine the nature of probability as used in law. There are five distinct uses of probability in law. The uses are found in the expression of the standards of proof, in the requirement that admissible evidence be reliable and relevant, in presumptive inferences and in confirmation of evidence in proceedings. Unfortunately, as one might expect, there is no agreed upon view of which interpretation of probability is or should be used in law. And it is important to realize that the different theories which have been advanced lead to different conclusions regarding the guilt or liability of a disputant. (See Kyburg 1974) The main difficulty in assessing theories of probability is the extent to which the probabilities assigned to propositions at issue in legal proceedings are arbitrary, and hence the extent to which verdicts based upon these assignments are arbitrary. The ability to provide a non-arbitrary assignment of probability will be the main criterion in evaluating theories of probability in law.

To the extent that there is any self-conscious understanding of probability in law there are two main theories that are advanced currently. First, the subjective theory, currently popular with some philosophers, some statisticians, and a growing number of legal theorists, holds that to say that the probability of a proposition at issue is Z is to say that one's degree of belief in that proposition is Z. In the third section I examine the subjective theory. The main argument for the use of subjective probabilities comes from coherence and dynamic coherence arguments. According to these argument, a rational deliberating agent must conform his degrees of belief to the standard probability calculus because otherwise he will be a sure loser to a cunning bettor. This argument, although powerful and interesting, has the serious flaw that it fails to apply in situations in which there is no cunning bettor.

But, the subjective interpretation would fail even if the coherence arguments succeeded because the constraint on belief imposed by the probability calculus is not sufficient to rule out non-arbitrary assignments of probability. A final consideration weighing against the subjective theory is that degrees of belief do not express the "rational connection" between evidence and propositions at issue that is a natural interpretation of legal reasoning.
The second popular view considered here is that probability is an infinite or finite relative frequency. On the limiting frequency conception of probability, to say that the probability of x being B is Z is to say that the limit of the ratio of As among Bs is Z, where A is the reference class for x being B. I examine the frequency views in the fourth section. A problem with the limiting relative frequency view is that it is not possible to single out a non-arbitrary limiting relative frequency.

On the finite frequency conception of probability, to say that the probability of x being B is Z is to say that the ratio of As among Bs is Z, where A is the reference class for x being B. Notice that on this interpretation the probability that x is B is just an elliptical way of saying something about a class ratio and not anything about x. But in law it is required that probabilities apply to unique individuals. The problem of application to a unique individual leads to the chance interpretation of probability.

The chance interpretation is an empirical interpretation that holds that chance is an empirical phenomenon distinct from relative frequencies, not just an elliptical property of a person in terms of class ratios, but a property that applies to unique events or individuals. The chance of a tossed coin turning up heads is a physical property of the coin and experimental set-up to produce a limiting relative frequency of .5 heads/total tosses. An individual may likewise be such that in a given environment and given his disposition, a given relative frequency of types of actions may will occur. This ability to apply to individuals is of some importance when one is dealing with the rights of individuals as individuals and not just as a member of a group.

A fourth interpretation of probability in law, and one that has the support of Leibniz, Wigmore (1913), Thayer (1898), and Cohen (1977), is the logical interpretation of probability. On this view probability is a measure of the degree of entailment between evidence admitted and the propositions at issue. This, according to the logical interpretation, probability is a generalization of implication. Legal reasoning can then be interpreted as inductive reasoning. Unfortunately the elaboration of this view by Carnap (1950) succumbed to the criticism of arbitrariness.

A final theory that is considered is that of Kyburg (1974). This a theory which is a careful blend of the previously mentioned theories. According to this view probabilities are objective logical measures of the relation between a hypothesis and a conclusion relative to one's belief state and based upon chances. Kyburg's theory is appropriate for legal reasoning, I argue,
because it is able to subsume traditional principles of legal reasoning and meet the objections of traditional theories.

3.1 The Uses of Probability in Law.

There are five distinct uses of probability in legal proceedings. Probability is used in the expression of the standards of proof, the relevance of evidence, the reliability of evidence, presumptive inferences, and in confirmation of evidence. In each case the question arises as to what interpretation of probability best suits the practice.

The standards of proof stipulate the necessary and sufficient conditions for a proposition at issue to be considered proved. Standards of proof are often given a purely probabilistic interpretation. (See Miller v. D.P.P. and Kaye, 1988) On this view proof beyond a reasonable doubt requires that the proposition at issue have a sufficiently high probability. Proof on a balance of the probabilities or preponderance of the evidence requires that the proposition at issue be more probable than not. Sometimes, in the U.S. a third standard is cited. This is the requirement that the evidence be clear and convincing. On a probabilistic analysis this latter standard is taken to require an intermediate degree of probability between proof beyond a reasonable doubt and proof on a balance of the probabilities.

It is also often thought that the standards of proof are not constant but vary in degree depending upon competing values, such as the type of crime. (Bater v. Bater.) Many empirical studies assume that a probabilistic analysis can be given of the standards of proof and attempt to measure the degree of probability that deliberating agents require in order to satisfy these standards. (Simon and Mahan, 1971) The evidence indicates that the standards of probability actually imposed not only vary depending upon the severity of the crime but vary from depending upon whether the deliberator is a judge, jury or student for instance.

A second use of probability in law concerns relevance. Relevance has long been thought to be definable in terms of probability. Stephen defined a fact A to be relevant to another fact B when it makes the existence or non-existence of fact B probable. (Stephen, 1948, 4) This idea from Stephen is often cited down to the present day. In Cloutier v. R it was said: (Cross, 1974, 16)
For one fact to be relevant to another, there must be a connection or nexus between the two which makes it possible to infer the existence of one from the existence of the other. One fact is not relevant to another, if it does not have any real probative value with respect to the latter.

A suggested interpretation of the standard definition of relevance can be made in terms of probability, as in Carnap (1950). This definition of relevance stipulates that a proposition A is relevant to proposition B iff the probability of B given A is not equal to zero. Carnap's definition is clearer because the elements of the definition — probability, propositions and the logic of these entities — are better understood than "connection" or "nexus", "real probative value" and "facts".

The standard of relevant evidence is usually thought to be completely categorical. In other words, if the proposition that is being proved (disproved) is more (less) probable given the evidence, then the evidence is relevant to the proposition. But one can also measure degrees of relevance by measuring the change in probability that results from admitting evidence. The degree of relevance is just Carnap's degree of confirmation which is equal to $P(H \mid E) - P(E)$. This might be what Wigmore had in mind, when, in a mysterious passage he said: (1940, s. 28)

[The court will require] something more, a generally higher degree of probative value for all evidence to be submitted to the jury than would be asked in ordinary reasoning. The judge, in his efforts to prevent the jury from being satisfied by matters of slight value, capable of being exaggerated by prejudice and hasty reasoning, has constantly seen fit to exclude matter which does not rise to a clearly sufficient degree of value. In other words, legal relevancy denotes, first of all something of a minimum of probative value. Each single piece of evidence must have a plus value.

An alternative interpretation of relevance is that given by Cross. Cross defined relevance as a matter of balancing probative value against ethical considerations, such as prejudice to the accused, confusion of issue, and wasting of time. (Cross, 1985, 54) So for example, in Cloutier v. R, where the accused was charged with importing marijuana, evidence that the accused was a user of marijuana was held irrelevant because of the prejudice to the accused. The jury might take this evidence to be more relevant than it really is. There is some truth in this view but Cross's way of putting the point is misleading. As was said in discussing Wigmore, the degree of relevance that is required is determined by value considerations, but the definition of relevance is strictly in terms of probability.
A third use of probability concerns the reliability of evidence. Both direct and circumstantial evidence can be more or less reliable. Some forms of evidence, such as opinion evidence, may be considered inherently unreliable. Similarly the problem with admitting hearsay evidence is that its reliability can not be determined. (Sheppard, 1989, sec. 8-9) One form of reliability of evidence is the credibility of witnesses. In the case of the testimony of witnesses the law defines the witnesses' credibility as the degree of trustworthiness of the testimony. (Raymond v. Bosanquet) If "trustworthiness" is defined in terms of "probability", then the natural interpretation is that reliability or credibility of testimony \( E \) of a witness as to an event \( H \) is the probability of \( E \) given \( H \). The credibility will depend upon factors such as the witnesses' intelligence, observational powers, memory character and so on and may be discovered by examination of the witness.

It should be noticed that if evidence is admitted via Bayes' theorem, that is, via, \( P(H|E) = \frac{P(E|H) P(H)}{P(E)} \), then the relevance of \( E \) is a function of the prior probability of the hypothesis, \( P(H) \), the probability of the evidence, \( P(E) \), and the reliability of the evidence, \( P(E|H) \). Obviously, just as relevance of evidence admits of degree so does reliability of evidence. And just as the degree of relevance is determined by ethical considerations, so is the degree of reliability.

A fourth use of probability in law concerns the use of presumptive inferences. Presumptive inferences are supposedly inferences from a basic proposition to a "presumed" proposition. Many presumptive inferences have a hidden probabilistic element to them. These inferences include information in the form of dispositions, habits, reliability, custom, capacity and similar fact evidence. In R v. Shrimpton the accused's character was taken to indicate that there was little disposition to commit a robbery. And, hence it was argued, the accused did not commit the robbery. This disposition could be given a probabilistic definition as the probability that a person would A given trait B. The inference, then, takes the form of drawing a conclusion about the accused individually on the basis of the level of probability. Or take another example. Everyone knows the post office is not very reliable in its delivery of letters. The reliability of the post office justifies an inference about the sending and receipt of individual letters. (Stephenson v. Dandy)

Commonly, two forms of statistical inferences are distinguished, direct inference and inverse inference. (Seidenfeld, 1977; Kyburg, 1980; Levi, 1977) Direct inference is thought to be relatively unproblematic. Most people will agree that some situations have a chance element to them. These chances, whatever they may be, are thought to conform to the mathematical
probability calculus. So, for example, a coin is flipped and it is known that in general, the bias of the coin for landing heads is .5. A principle of direct inference, is a principle which allows one to infer the probability for an arbitrary or random toss of the coin. Or to return to the post office example, if we know the probability that letters will be delivered on time is 75%, then we can, through a direct inference, calculate that the probability that an arbitrary individual letter was delivered is 75%.

On the more problematic inverse inference the bias of the coin is not known. Rather there is a record of tosses of the coin, and the question is what is the bias of the coin. This type of inference is also made in legal proceedings. Any time sampling information is offered as evidence of some parameter of the entire population, inverse inference in being used. Bayesians, generally, think it is possible to duplicate the force of the direct inference by a unique point-like value for the bias of the coin, through the application of Bayes' theorem. Non-Bayesians object to this view. For example, on Kyburg's view, in the present situation only an interval valued probability can be established as the bias of the coin.

A fifth use of probability in legal proceedings concerns confirmation of propositions. Often when evidence is considered unreliable, or when a very high standard of reliability is imposed upon evidence, as in sexual cases and evidence from accomplices, secondary evidence is needed to confirm the primary evidence. Various definitions have been offered of corroboration. It is often required that evidence offered to confirm a proposition is independent testimony which "connects" the accused with an act. It was said by Lord Hailsham: (Zuckerman, 1989, 168)

The word "corroboration" by itself means no more than evidence tending to confirm other evidence. In my opinion, evidence which is (a) admissible and, (b) relevant to the evidence requiring corroboration, and if believed, confirming it in the required particulars, is capable of being corroboration of that evidence and, when believed, is in fact such corroboration.

There is another situation in which evidence confirms a proposition at issue. In legal proceedings this is often referred to as the convergence of evidence. In this situation two or more propositions each confirm the same proposition. Cohen has claimed (1977) that corroboration and convergence have the same logical structure. But Eggleston (1983) claims that corroboration and convergence do not have the same structure. Both situations conform to Carnap's notion of increase in degree of confirmation. Thus the tendency to confirm can be interpreted by way of Carnap's notion of increase of confirmation = P(H|E) - P(H).
3.2 The Logical Interpretation of Probability.

On the logical interpretation of probability, probability is a measure of partial entailment between evidence and a conclusion. Leibniz, both a lawyer and philosopher, long ago offered this interpretation of probability. According to Leibniz "The whole of judicial procedure is nothing but a kind of logic applied to questions of law." (Hacking, 1975, 86) This view is also implicit in the writings of Thayer (1898) and Wigmore (1913). Wigmore (1913) clearly has the logical interpretation in mind in the following quotation: (1913, 27)

Thus, throughout the whole realm of evidence circumstantial and testimonial, the theory of inductive inference, as practically applied, is that the evidentiary fact has probative value only so far as the desired conclusion based upon it is a more probable or natural inference .... But the general spirit and mode of reasoning of the courts substantially illustrates the dictates of scientific logic.

Unfortunately, despite writing a large book on the principles of judicial proof, Wigmore was not able to provide anything more advanced than this statement. It was not until the work of Keynes (1921) and Carnap (1950) that this interpretation was worked out in any detail. Cohen (1977) was the first philosopher to offer a detailed form of the logical interpretation explicitly designed for legal reasoning. According to Keynes (1921, 7) propositions are not probable in themselves but only relative to evidence, and this relation is a logical relation. The logical interpretation is a far more natural interpretation for law than any of the other interpretations because there is supposed to be some "rational connection" between evidence admitted and the conclusions that are drawn. On the logical view of probability this relation is one of partial entailment.

The logical interpretation of probability, as developed by Carnap, was supposed to satisfy a number of criteria (Spielman, 1982). First, probability is a metalinguistic relation defined over a formal language that is rich enough for scientific inquiry. This would require a language such as set theory. This would, I add, also be sufficient for legal inquiry. Keynes' idea that probabilities are relative to evidence is satisfied by requiring that every sentence in the formal language is assigned a probability relative to every potential set of sentences of that language. Second, the truth values of the probability statements are logically true or false rather than empirically true or false, and can be determined by calculation. Third, probabilities are the rational degrees of confidence that one should have relative to the evidence. Fourth, and important for scientific and legal inquiry, the probability function is not arbitrary in any way.
Carnap defined probability as a metalinguistic relation over a formal language, rather than over possible worlds. Carnap worked with simple languages of first order predicate calculus, with finite numbers of predicates and individual constants. Given this language there are a number of descriptions of possible states of the world. There are state descriptions which are conjunctions of every atomic sentence or its negation but not both. One can also distinguish structure descriptions which consist of a disjunction of isomorphic state descriptions, that is, each of which can be obtained from one of the other by permuting the individual constants in a state description. An example of a simple language with two individuals and two predicates is given below.

<table>
<thead>
<tr>
<th>State descriptions</th>
<th>Structure descriptions</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Fa &amp; Ga &amp; Fb &amp; Gb</td>
<td>Fa &amp; Ga &amp; Fb &amp; Gb</td>
</tr>
<tr>
<td>2. Fa &amp; Ga &amp; Fb &amp; -Gb</td>
<td>(Fa &amp; Ga &amp; Fb &amp; -Gb) or (Fa &amp; Ga &amp; Fb &amp; Gb)</td>
</tr>
<tr>
<td>3. Fa &amp; Ga &amp; -Fb &amp; Gb</td>
<td>(Fa &amp; Ga &amp; -Fb &amp; Gb) or (-Fa &amp; Ga &amp; Fb &amp; Gb)</td>
</tr>
<tr>
<td>4. Fa &amp; Ga &amp; -Fb &amp; -Gb</td>
<td>(Fa &amp; Ga &amp; -Fb &amp; -Gb) or (-Fa &amp; -Ga &amp; Fb &amp; Gb)</td>
</tr>
<tr>
<td>5. Fa &amp; -Ga &amp; Fb &amp; Gb</td>
<td>Fa &amp; -Ga &amp; Fb &amp; -Gb</td>
</tr>
<tr>
<td>6. Fa &amp; -Ga &amp; Fb &amp; -Gb</td>
<td>(Fa &amp; -Ga &amp; Fb &amp; Gb) or (-Fa &amp; -Ga &amp; Fb &amp; Gb)</td>
</tr>
<tr>
<td>7. Fa &amp; -Ga &amp; -Fb &amp; Gb</td>
<td>(Fa &amp; -Ga &amp; -Fb &amp; Gb) or (-Fa &amp; -Ga &amp; Fb &amp; -Gb)</td>
</tr>
<tr>
<td>8. Fa &amp; -Ga &amp; -Fb &amp; -Gb</td>
<td>-Fa &amp; -Ga &amp; -Fb &amp; Gb</td>
</tr>
<tr>
<td>9. -Fa &amp; Ga &amp; Fb &amp; Gb</td>
<td>(-Fa &amp; Ga &amp; -Fb &amp; Gb) or (-Fa &amp; -Ga &amp; Fb &amp; Gb)</td>
</tr>
<tr>
<td>10. -Fa &amp; Ga &amp; Fb &amp; -Gb</td>
<td>-Fa &amp; -Ga &amp; -Fb &amp; -Gb</td>
</tr>
<tr>
<td>11. -Fa &amp; Ga &amp; -Fb &amp; Gb</td>
<td></td>
</tr>
<tr>
<td>12. -Fa &amp; Ga &amp; -Fb &amp; -Gb</td>
<td></td>
</tr>
<tr>
<td>13. -Fa &amp; -Ga &amp; Fb &amp; Gb</td>
<td></td>
</tr>
<tr>
<td>14. -Fa &amp; -Ga &amp; Fb &amp; -Gb</td>
<td></td>
</tr>
<tr>
<td>15. -Fa &amp; -Ga &amp; -Fb &amp; Gb</td>
<td></td>
</tr>
<tr>
<td>16. -Fa &amp; -Ga &amp; -Fb &amp; -Gb</td>
<td></td>
</tr>
</tbody>
</table>

Table 1

Each state description (or structure description) is then assigned a measure, a positive real number between 0 and 1, such that the sum of all the state descriptions (or structure descriptions) measures are 1. In our example each structure description is assigned 1/10 because there are 10 structure descriptions. It is also stipulated that the value of the measure for any two contradictory sentences, that the measure of the disjunction is the sum of the measure function for each sentence. The measure function for each sentence is considered the appropriate probability of the sentence. Then a measure function m, the extent to which evidence E confirms H, is defined as the ratio:
\[ c^*(H,E) = \frac{m(E \& H)}{m(E)} \]

For example, relative to the situation as described by the limited language above \( c^*( Fa \& Ga \& -Fb \& Gb) = 1/20 \) and \( c^*(Fa, Fb) = 3/5 \).

It is already obvious that an infinite number of different measure functions will satisfy the constraint that the sum of the measure functions for each state description (or structure description) equals 1. So the arbitrariness of the confirmation function infects the system very soon. A natural function is \( c \), which assigns the same measure to each state description. Another measure, as used above, is \( c^* \) which assigns to each structure description the same measure. These two different definitions yield two different sets of assignments of probability to all the sentences in the language. It is plain to see that there are an infinite number of possible assignments of measures to sentences because each state description or structure description can take on any real number between 0 and 1. But Carnap's system is also arbitrary because the measure functions depend upon the number of state descriptions or structure descriptions which depends upon the descriptive capacity of the language.

It is possible to ignore the metalinguistic aspect of Carnap's theory, and define logical probability as a relation over possible worlds rather than their descriptions. One can think of a state description as describing a possible world relative to the description potential of that language. The crux of the logical interpretation, then, is that if proposition A is entailed by B then A is true in every possible world in which B is true. If A is neither entailed by B nor inconsistent with B, then B is not true in every possible world in which A is true, but only some of these possible worlds in which B is true. One can obtain a measure of degree of entailment by defining the degree of entailment (logical probability) of A given B as equal to the number of possible worlds in which A and B are true divided by the number of worlds in which B is true.

If this program could be carried out, it would be adequate for legal as well as scientific reasoning of a non-deductive nature. Carnap assumed that logical probabilities could be measured by single real numbers relative to evidence \( w \) in language \( L \) satisfying the standard probability calculus. But Carnap and many of his followers gave up on this program because they were not able to satisfy the condition that probability be non-arbitrary. The constraints on rationality were not sufficient to generate a unique confirmation function, but rather, as has been seen, they generate an infinite set of \( c \) functions.
The problem of arbitrariness and other problems led many philosophers to dismiss the logical interpretation of probability. All that one can do is commit oneself to a particular c function, and then the truth values of all sentences are determined. But the rationality involved cannot settle disputes between rational agents who adopt different c functions. Thus one is lead to a form of subjectivism in which one is free to choose one's c function. Any attempt to construct a form of the logical interpretation must solve the problem of choosing a non-arbitrary c function.

3.3 The Subjective Interpretation of Probability.

Contrary to the writings of Wigmore, Thayer and Cohen, some legal theorists hold that legal proof is not an objective matter. Zuckerman, for example, simply states without argumentation that "the evaluation of evidence in courts is fundamentally subjective" and identifies the probability of a proposition with one's degree of confidence in it. (Zuckerman, 1989, 107) Kaye holds that the subjective definition of probability provides "a meaningful interpretation of the probabilities involved in the legal process." (Kaye, 1979, 45) Of course it may be a meaningful but poor interpretation. Immediately the proposal to identify probabilities with degrees of belief looks suspect since the relation between the truth of the proposition at issue and the judge's degree of confidence in it are unrelated. But it is important to examine this view sympathetically since it is becoming quite popular in philosophy and law. (Tillers (Ed), 1988)

David Kaye, following in a tradition since Bentham (1827), argues that a subjective probability is a measure of a judge's willingness to bet on a proposition. (Kaye, 1979) There are a number of objections to this view, but they all boil down to the problem that, not unsurprisingly, the probabilities are too arbitrary. The first question that arises is to what extent the subjective interpretation of probability permits disagreement to exist in legal proceedings. Unless proceedings do not conform to any rational methods of inquiry (which I suppose some may agree with), it would appear that there is real disagreement between competing parties over issues of assessment of evidence. Different attitudes are adopted towards whether a proposition has been proved beyond a reasonable doubt or on a preponderance of the evidence. The participants are not just expressing their own degree of persuasion but what they believe is the rational degree of persuasion that they should have regarding the issues. But, as far as we have seen, the subjective theory would not seem to allow for this.
Another way of putting the criticism is that the verdict of the judge based upon a subjective probability has no connection with the truth of the proposition at issue. The assignment of a probability must in some way indicate what the truth of the proposition is. This would require that based upon the evidence a judge would be constrained to adopt some non-arbitrary attitude toward the proposition at issue. But in the case of the subjective interpretation of probability it would appear that the judge is completely free to adopt any degree of belief in the proposition at issue. As a consequence the judge does not even have to consider the evidence in question because his degree of belief is independent of the evidence. In Bentham's time, and from most lawyers who express such a viewpoint, no adequate answer to these charges has been forthcoming.

An answer to the question of how to relate degrees of belief to truth was started by Ramsey (1926) primarily, and later built upon by De Finetti (1937), Savage (1972), and particularly Lewis. (Teller, 1976). On this view a rational agent's degrees of belief are not arbitrary at all. So the interpretation is slightly different than the pure subjectivism above. Probabilities, on this view are degrees of belief held by rational individuals. According to this view a rational person must conform his degrees of beliefs to the probability calculus. The argument is that a rational person is one who would not place his bets on event such that he will lose for sure. Surely, one would think, that this is a minimal requirement. But given this stipulation one can show that a person's degrees of belief must conform to the probability calculus. So probabilities are degrees of belief that an agent must hold in order not to be a sure loser.

As Ramsey (1926, 1978, 84) explained:

These are the laws of probability, which we have proved to be necessarily true of any consistent set of degrees of belief .... If anyone's mental condition violated these laws, his choice would depend upon the precise form in which the options were offered him, which would be absurd. He could have a book made against him by a cunning bettor and would stand to lose in any event.

But those who have wanted to interpret standards of proof in a subjective manner, such as Kaye, have failed to show how this argument applies to legal proceedings. Indeed they have simply assumed that the betting analogy holds for legal proceedings. But it is difficult to see how it holds. Kaye says: "an involuntary bookmaker logic is particularly apt in the context of trials." (1979, 47) But quite the opposite is true. There is no involuntary bookmaker. Nothing could be more inapt. And the supposition that events conspire in a manner as a cunning bookmaker is simply incorrect. Nature does not conspire against us.
The Dutch book argument should be spelled out a bit more. In essence, a rational person would not allow himself to be suckered into a betting situation such as a toss of a coin where the opponent says “heads I win, tails, you lose”. As a simple example, say you offer odds of 9 to 1 on a proposition P and 1 to 1 on not P. This means that your degree of belief in P is $9/9+1 = .9$. The bookie sets the stake at $1.00 on both P and not P. If P is true, the bookie wins $.90 on his bet on P and loses $.50 on his bet on not P. If P is false then bookie loses $.10 on his bet on P and wins $.50 on his bet on not P. Hence the bookie wins $.40 from you regardless of the outcome of the bet. Thus it seems if you do not want to have a cunning bookie make a book against you, you had better not give the type of odds mentioned above. In general this shows one should not offer degrees of confidence that do not add up to one.

Adapted to the legal setting the argument is best thought of as an insurance model of legal process. The standard of proof is said to be met if the probability of the proposition at issue is sufficiently high. The probability is a measure of the judge’s degree of belief on that proposition as evidenced by his betting. The judge might be thought to be literally in a position where he must make a bet on the liability of the accused. But this bet, it should be noted is not meant to imply anything regarding the (full) belief of the judge in the proposition at issue. A bet of more than even odds is not meant to imply, for the subjectivist, that the judge (fully) believes that the proposition is true.

The more convinced the judge is, as Bentham said, that a proposition is true, the higher the odds he will be willing to bet on that proposition. (Postema, 1983) The degree of belief of the judge can be identified with the judge’s least betting odds. So if the lowest odds a judge is willing to give for P are a:b then his degree of belief is $a/a+b$. Next it is accepted that a rational judge would not accept a bet for which a “cunning bettor” could guarantee himself a win. The judge is held to accept the metaphor that nature is a cunning bookie against whom he is playing. The above example shows, for instance, that a judge should not have degrees of confidence which do not add up to one. If the judge gives favourable odds on both one proposition and the negation of that proposition he will lose for sure. The argument can be extended for each axiom of the Kolmogoroff calculus.

Who is the cunning bettor? In the original argument there is a bookie who sets the stake and the bets. This is certainly understandable reasoning at the race track. On the legal model the cunning bettor might be thought of as society or perhaps the accused. But the idea that society is conspiring against the judge is false and silly. What about the view that the accused is a
cunning bookie in some sense? The accused leads his life so as to maximize his own gain, and in
doing so tries to make shrewd bets on whether he will get caught. Let's examine how this
argument might go.

The basic idea is that in order not to be outdone by a cunning accused, the judge's judgement
about the hypothesis H, that A is guilty, must be constrained by Bayes' theorem. This depends
upon the judge's ability to make conditional bets. The judge makes a bet on H, that A is guilty
which is to be called off if E is false. Suppose the judge gives odds of m:n, so his degree of belief
is m/n+m=a. The stake is set at 1 for convenience. Now it can be established that the
propositions must conform to Bayes' theorem:

Bayes' theorem: \( P(H \mid E) = P(H) \frac{P(E \mid H)}{P(E)} \)

Suppose that the the judge makes a conditional bet on H given that E is true, to be called off if
E is false. Let the judge's partial belief \( P(H \mid E) = a \), \( P(H) = b \), \( P(E) = c \), \( P(E \mid H) = d \). Then if \( a \neq bc/d \) a dutch book can be made against the judge. ("+" indicates the bet in favour of H and "-" against H.)

<table>
<thead>
<tr>
<th></th>
<th>H(\mid E)</th>
<th>H</th>
<th>E</th>
<th>E(\mid H)</th>
</tr>
</thead>
<tbody>
<tr>
<td>H E</td>
<td>+</td>
<td>-</td>
<td>+</td>
<td>-</td>
</tr>
<tr>
<td>T T</td>
<td>1-a, -(1-a)</td>
<td>1-b, -(1-b)</td>
<td>1-c, -(1-c)</td>
<td>1-d, -(1-d)</td>
</tr>
<tr>
<td>T F</td>
<td>0, 0</td>
<td>1-b, -(1-b)</td>
<td>1-c, -(1-c)</td>
<td>-d, d</td>
</tr>
<tr>
<td>F T</td>
<td>-a, a</td>
<td>-b, b</td>
<td>0, 0</td>
<td>1-d, -(1-d)</td>
</tr>
<tr>
<td>F F</td>
<td>0, 0</td>
<td>-b, b</td>
<td>0, 0</td>
<td>-d, d</td>
</tr>
</tbody>
</table>

Table 2

The stakes are thought to be set by the cunning accused. He can guarantee himself a win by
setting the stakes as follows. If \( ad \neq bc \), then either \( ad>bc \) or \( ad<bc \). In the first case \( ad-bc \) is positive. Then the accused bets against E at stake a, for H at stake c, against \( H \mid E \) and for \( E \mid H \). In each case the outcome is \( ad-bc \) and so is positive. In the second case \( bc-ad \) is positive. The accused bets against H at stake c, for E at stake a, for \( H \mid E \) and against \( E \mid H \). The outcome is \( bc-ad \), and so is positive. So by betting in this manner the accused can guarantee a benefit.
The interpretation of Bayes' theorem is that a judge's hypothesis $H$ must be related through Bayes' theorem to the evidence $E$, and the probability of the evidence, given the hypothesis. This shows how the judge must consider the evidence if he is to be coherent. One of the problems is what role the called off bet has. If the judge can make a conditional bet to be called off if the evidence is false, then there must be some independent method of determining the truth of the evidence. But if this is so why bother making a bet?

But even if this framework can be made sense of there is a further gap in the argument because the requirement that a judge's degrees of belief conform to the probability calculus holds only at a particular time. A judge must evaluate the evidence in light of Bayes' theorem and the probability of evidence. But this theorem only requires that at a particular time Bayes' theorem hold. But as the judge responds to the new evidence that is admitted there is nothing to prevent him from altering the probabilities of $E$ and $E|H$ so as to make $P(H|E)$ whatever he wants.

What is needed on the subjectivist view, then, is a way of changing probabilities over time to reflect the consideration of new evidence. As Ramsey (1926, 1978, 94) argued:

Since an observation changes (in degree at least) my opinion about the fact observed, some of my degrees of my belief after the observation are necessarily inconsistent with those I had before. We have therefore to explain exactly how the observations should modify my degree of belief; obviously if $p$ is the fact observed, my degree of belief in $q$ after the observation should be equal to my degree of belief in $p$ before, or by the multiplication law to the quotient of my degree of belief in $pq$ by my degree of belief in $p$. When my degrees of belief change in this way we can say that they have been changed consistently by my observation.

Ramsey is describing what has been called conditionalization:

admission by conditionalization: $P(H) = P(H|E) = P(H&E)/P(E)$

The use of conditionalization in legal reasoning has been asserted by Kaye (1979, 1980). One reason in favour of conditionalization is that it guarantees that the new probability values will be conform to the probability calculus. In addition a dynamic Dutch book argument is used to justify the use of conditionalization. (Lewis, in Teller, 1973, Skyrms, 1990) Suppose that you
have revised your probabilities according to the conditionalization rule above. Suppose that $P(H)$ has been revised to $a$ and $P(H\mid E)$ was $b$. Then if $b$ is not equal to $a$, a dutch book can be made via a temporally conditional bet. Before the truth of $E$ is known the bookie places two bets that take place only when $E$ is verified. He places his first bet in favour of $H$ if $a$ is less than $b$ and against $H$ if $a$ is greater than $b$. His second bet is placed oppositely on $H$ given $E$, to be called off if $E$ is not verified. If $E$ comes true the bookie will win given his different bets. But if $E$ does not come true the bet is called off. For example, let the stake be $1. If a is less than $b$, and $H$ is true he is paid $1 - a$ on the bet for $H$, and $- (1 - H)$ on the bet against $H$ given $E$. The net benefit is $b - a$, which is positive. If $H$ is false he also receives $b - a$.

But this rule of conditionalization is open to the problem that observations are considered certain. Notice that the $P(E\mid E) = 1$. So that when a witness testifies to some fact, say that $A$ was seen robbing a bank, then this evidence enters into the probability calculus and cannot be changed given contradictory testimony. An alternative rule which meets this criticism is probability kinematics due to Jeffrey (1965). Since $P(H) = P(H \& E) or P(H \& \neg E)$, the following formula holds.

Admission by probability kinematics: $P(H) = P(E) P(H\mid E) + P(\neg E) P(H \mid \neg E)$

There is an analogous dutch book argument in support of probability kinematics. (Skyrms, 1990) Notwithstanding dutch book considerations, there are still remaining problems with the Bayesian calculus. First, it is assumed that the judge adopts a point valued degree of belief. But as Tribe and others have argued, this degree of precision is out of place in legal contexts. Some subjectivists, like Good (1959), have replied to the above criticism by noting that a subjectivist need not have exact probabilities but instead may work with inexact probabilities. So this objection may be overcome. In fact it can be shown in an analogous way to the coherence arguments given above, that indeterminate probabilities also must conform to a generalized probability calculus. (Kyburg, 1974)

Second, there is the problem of justifying the prior probability that the judge adopts. There seems to be no reason why a judge should not adopt any probability as a prior probability. On the subjective view the prior probability itself was due to the principle of conditionalization and so met the coherence requirement. So that if one traces the prior probabilities into the past
one should arrive at some initial correct prior probability. On what basis is this probability justified? The subjective theory cannot answer this question.

The problem of justifying the prior probability has been addressed by De Finetti (1937). It is possible to prove that if events are exchangeable, in the sense that their order of occurrence does not affect their probability, then as more and more observations are made through collection of evidence, the subjective probability of different observers converges to the same value regardless of the prior probability. For the purpose of legal proceedings the idea is that as more and more evidence is offered in court, regardless of the prior probabilities held by the judge, jurors or other observers, the final probabilities will converge. In this manner the regress argument is supposed to be answered.

It is easy to see in extreme cases where the judge is certain of the guilt of the accused that no amount of evidence can, as a matter of mathematics, change the zero partial belief into a non-zero partial belief. If \( P_{\text{old}}(H) \) is 0 then \( P_{\text{new}}(H) \) must be 0 and similarly for a prior probability of 1. But what prevents a judge from holding that the prior probability has either of these values? Nothing apparently, except the prior probability for that probability and so on. Second, the convergence theorem depends upon the exchangeability of the events. But if exchangeability is no more than the invariance of the probability under change of order of events, then this is simply a feature of one's degree of belief regarding the events. But this exchangeability property is itself completely subjective. It follows then, that the exchangeability of events is something the judge is free to disregard.

The fatal objection to the exchangeability theorem, though, is that it depends upon the assumption that the observers will agree on what observations they have made. This agreements depends, though, on the agreement of observers on what observations have been made, which in turn depends upon the reliability that an observer assigns to his sensations. But this assignment of probability is itself completely subjective. So if one is a skeptic and takes observations to be completely unreliable, then the skeptic's degree of belief will always remain at 0. This follows straightforwardly from Bayes' theorem: \( P(H|E) = P(E|H) P(H) / P(E) = 0 \) where \( P(E|H) \), the reliability, is 0.

My conclusion is to give up the idea that betting odds are the appropriate way to generate verdicts. This method leads to a fundamentally arbitrary assignment of probabilities. A judge must, in the end, have some way to arrive at the correct degree of belief. One answer to this problem is that the judge's degree of belief be constrained by what the objective probability of
guilt is. The burden then, is to develop some method of determining the objective probability. Apparently beliefs can not be constrained by logical principles alone as on the logical theory of probability. But perhaps some empirical properties can constrain rational degrees of belief. This leads to the empirical theory of probability.

3.4 Empirical Interpretation of Probability.

The use of the empirical theory of probability in law has been advocated by Neil Cohen (1985) and adopted without argumentation by Barnes and Connolly (1986), and Barnes, (1983) and Gastwirth (1989). On the empirical theory attributions of probability are understood to be independent of one's state of belief. One can perfectly well believe with utmost certainty that the next toss will come up heads but the probability is still .5 that it will be a head.

John Venn (1866) proposed that probability was a ratio in an infinitely large class. So that in a sequence of trials, say coin tosses, the probability is equal to the ratio of heads to total tosses. The motivation for the infinite set of trials is clear. If there were a finite class the ratio would only accidentally work out to .5. Most likely it would be either below or above .5. By taking the relative frequency over infinite classes the probability will be arbitrarily close to .5. Hence one is left with picking an arbitrarily large set of trials.

If probability is defined in terms of the limit of a relative frequency then the probability appears to be arbitrary because limits of sequences depend upon the order of members of the sequence. For any set of outcomes there may be a number of different possible limits or no limit at all depending upon the ordering. So, in many cases, the probability depends upon the accidental feature of which order the actual trials were performed.

Consider Bertrand Russell's example. (Cohen, 1989) Suppose one wanted to know the probability that a randomly picked number is a prime. If the order of the sequence of numbers picked were the natural order then the limiting relative frequency would be zero. But if the sequence were changed so that the first nine numbers were the first nine primes, and then these were followed by the first non-prime, and then by the next nine primes, and so on indefinitely, it can be shown that the limiting relative frequency is 9/10. Other arrangements of trials are also possible which further change the limiting relative frequency. Thus the probability is not independent of the order of the sequence of trial and is therefore not unique according to the definition of probability as a limiting relative frequency.
It is interesting that even on the limiting relative frequency account of probability one still has to revert to a finite frequency approach to probability because one can only observe a finite sequence of trials. This is clearly the case in dealing with evidence in legal proceedings. On the evidence of a finite sequence of trials, the relative frequency of an outcome O is compatible with any limit x of the relative frequency of outcomes in an encompassing infinite sequence. If a coin is tossed 1000 times and turns up heads, any limiting relative frequency is compatible with these observations. In short, for any finite sequence SF of trials there is an infinite sequence SI of trials of which SF is a subset, for which the frequency of O is x.

But it might be supposed that the finite frequency is good evidence for the existence of a limiting relative frequency. The law of large numbers, a theorem of the probability calculus, states that the relative frequency of the occurrence of an event in n independent repetitions of a trial tends to its probability as n increases without limit. So one can say that given a large number of trials the relative frequency of the finite frequency very probably is the same as the limiting relative frequency. But does the finite frequency really constitute evidence that the probability of this sequence is some number? All that a high probability of a limiting relative frequency applies to is a set of sequences, not the sequence itself. It means that if there were independent sequences of trials, most of the sequences will have a certain limiting relative frequency.

The largest problem for the finite frequency interpretation of probability in legal and some scientific contexts is that relative frequencies do not apply to individual events. On the relative frequency view probabilities apply only to classes. In the case where probabilities apply to an individual by way of a singleton set the probability is always 1. There is just one unique individual A with property P and he is the only member of a class with property P. Hence the probability is 1.

3.5 The Need for Proof of Particular Conduct: the Chance Theory.

It has been held that "the requirement that evidence should focus on the defendant must be taken to be a rule of law relating to proof ...". (Williams, 1980, see also Zuckerman, 1989, Cohen, 1977) Similarly Kaye (1980, 489) says: "... notwithstanding whatever probability theory may teach us, assuring the appearance of fairness precludes imposing liability in the absence of some evidence singling out particular defendants." It does seem morally repugnant to
find a person to be guilty or liable just because he happens to belong to some class with the right ratio of guilty people. This requirement could be interpreted in either a strong or weak sense. In one sense, the evidence must apply to a unique individual and not simply to that individual as a part of a group. Secondly it could mean that the evidence must apply to a unique individual and no others at all.

Nevertheless Williams and Zuckerman are wrong in maintaining that this principle is absolute. In practice the requirement that evidence apply to a single person is not absolute. A and B can be jointly tried and convicted for an offence for which evidence does not single out either of them individually. But it is correct that whoever is accused, the evidence must be able to distinguish between the accused and those who are not accused. In Cooke v. Lewis, for instance, the plaintiff was shot while hunting by one of the two defendants, both of whom fired at different birds at the same time. The evidence is that one of these two hunters shot the third hunter, but there is no specific evidence against either of them individually. The burden of proof is shifted and the defendants have the burden of proving that each did not fire the shot. If neither can prove that he did not shoot the third hunter both are convicted. In the majority of cases, however, the probabilities must apply to individuals

Von Mises, a defender of the relative frequency view of probability, himself recognized that relative frequencies do not apply to individuals. He states: (1964, 11)

When we speak of the "probability of death" the exact meaning of this expression can be defined in the following way only. We must not think of an individual, but of a certain class as a whole, eg., "all insured forty-one year olds living in a given country and not engaged in certain dangerous occupations." .... We can say nothing about the probability of death of an individual even if we know his condition of life and health in detail. The phrase "the probability of death" when it refers to a single individual has no meaning at all for us.

The main problem, then, with finite relative frequency accounts of probability is that probabilities do not apply to single individuals. Only simple intuitive considerations are needed in order to make clearer a distinction between the chance of an event and the relative frequency of an event. Consider a newly minted coin. This coin has never been tossed and will be tossed once and then destroyed. What is the probability of tossing heads on that toss? On the finite frequency account the outcome will be heads or tails and thus the probability, apparently, is either 1 or 0. But the chance of tossing a head is still thought to be .5.

Unlike judgements of relative frequency, the chance that the coin will land heads supports counterfactual judgements. If the coin were tossed 100 times, then the relative frequency of
Interpretations of probability

heads would probably be .5. In possible worlds just like this world except the newly minted coin is tossed 100 times, 50 outcomes would be heads is probably true. Whereas on the relative frequency judgement nothing can be said regarding the counterfactual situation where the coin is tossed one hundred times. Chance, then seems to exhibit a modal property of individual objects, which relative frequency accounts do not have. The coin, whether it is ever tossed or not still has the property that if it were tossed, a certain proportion would likely turn up heads. In the case of the coin, the property appears to have something to do with the construction of the coin and the mechanism for tossing it.

Evidence of chance is through relative frequencies. But what is chance? On one view a chance is an empirical property of an object and its experimental set-up. On this view the disposition of the coin to land heads 50% of the time is a force that the coin, or coin plus tossing mechanism, has. In the case of persons, chances are most naturally interpreted as the dispositions that a person has to do something in a certain situation. The disposition that someone has is most naturally evidenced by relative frequencies of similar events, and by one's mental characteristic such as character, habits, opportunity, and beliefs. As was mentioned earlier, in *R v. Shrimpton* the accused's character was taken to indicate that there was little disposition to commit a robbery. Hence it was argued, the accused did not commit the robbery.

In order to ascribe a chance of tossing heads of .5 to a coin, it is necessary to know that if that coin were tossed it would result in a relative frequency near .5. This implies that if a relative frequency is ascribed to a certain object, say a coin, it should be invariant under counterfactual conditionalization. In other words, if a fair coin is tossed 100 times and comes up heads each time, in order to consider this frequency the chance of turning up heads it should be invariant under an additional 100 tosses. Who would believe that this behaviour of the fair coin is not purely accidental?

This point about chance applies to persons as well as coins. One cannot say that a person has a disposition to rob someone, as in *R v. Shrimpton* unless the relative frequency data is resilient, or sufficiently invariant under different situations. It is not, for example due to the fact that the person was black, and most robberies in that area happen to be by blacks. In order to assert that the probability that A robbed Y because A was black, this probability has to be invariant under many changes, such as locality, circumstances and so on. Only in this way can it reflect the disposition of the person to commit the action.
The notion of resiliency, that is a mark of chance, is related to Skyrms' (1980) and Davidson and Pargetter's (1987) use of a similar form of resiliency in legal reasoning. But these writers assert somewhat different interpretations of resiliency. Skyrms for example, takes resiliency to be a property of degrees of belief, and takes chance not to be an empirical property, but resilient degrees of beliefs. However the subjective theory was found inappropriate for legal reasoning. A probability is resilient if it is invariant under conditionalization (which includes counterfactual conditionalization).

\[
\text{resiliency of probability: } \quad \text{The resiliency of the probability of } A \text{ being } a \text{ is equal to } 1 - \max |P(A) - P(A|B)| \text{ for all } B
\]

It is interesting to cite the connection of resiliency with De Finetti's notion of exchangeability. As Skyrm's (1980) notes, De Finetti's notion of exchangeability is a type of invariance of probability. It is invariance under ordering of events. Exchangeability may be understood as entailed by complete invariance under counterfactual conditionalization. That is, in world B just like our world A except that the order of trials is different in B, the resilience is one. Thus the phenomenon of chance as evidenced by high resiliency, can ground some of the assumptions needed for adopting degrees of belief.

3.6 The Epistemic Interpretation of Probability.

Some of the problems that the empirical theory has still survive when the chance interpretation is adopted. First the problem that there is a lack of statistical evidence to make probability judgements still exists. Most arguments in legal proceedings do not rely on precise fully documented frequency information but rather on the common knowledge of the decision makers. This property of legal reasoning has been made effectively by Tribe (1970) and Cohen (1977). Secondly, even if statistical knowledge was available it is doubtful whether it is reasonable to adopt determinate point like probabilities. This degree of precision seems to be much to optimistic regarding our epistemic abilities. Furthermore, the empirical interpretation fails to make a "connection" between evidence and a conclusion because it is an empirical statement and not a logical relation.
The epistemic theory of probability developed by Kyburg is able to answer these criticisms. (Kyburg, 1974). A rational agent is conceived of as having a set of beliefs which are directed toward sentences. Probability is defined relative to these beliefs and relative to the language in which the sentences are. A rough statement of Kyburg's theory is that the probability of a sentence X in a rational agent's set of beliefs K is in the interval \([p, p^*]\) if and only if:

1. "a \in B \text{ iff } X" is in K
2. "a \in A" is in K
3. "frequency of A's among B's is in \([p, p^*]\)" (in a special case \(p = p^*\)) and this sentence is the best statistical information available. (A is the reference class for "a is a member of B")

This definition requires some explanation. First it should be noticed that probability is syntactic. On Kyburg's theory, probabilities are functions of sentences. This property satisfies the traditional Carnapian requirement that probability be a metalinguistic function. However, this seems a poor assumption to me. An agent's beliefs are clearly not directed toward sentences in a formal language. A better prospect is that an agent's beliefs are directed toward propositions.

Second, the interpretation is a logical interpretation because all probabilities are relative to the evidence that one has. This satisfies the intuitions of some legal theorists that there be some "connection" between evidence and conclusion. But the logical interpretation can be maintained without resorting to defining probability metalinguistically. Kyburg, however, apparently denies the advantage of more than one interpretation of probability. But I think it makes more sense to regard Kyburg's definition of epistemological probability as a function which constrains a rational agent's adoption of a c function, and hence, since this c function is adopted, of his degree of belief by way of the agent's belief in the chance of events.

Third, the interpretation of probability allows for the background knowledge and experience of the judge or jury to enter into the calculation of probabilities. Of course one might say that this is where some element of subjectivity comes in because it is each person's beliefs that determine the probabilities in question. This may be so, but this type of subjectivity is fundamentally different from the subjective account which does not allow for disagreement or agreement about probabilities. On this view disagreement about probabilities can be settled to the extent that any empirical disagreements can be settled.
Kyburg also allows, with some degree of resistance, that probabilities are based upon chances. However, as I argued, chances are evidenced by relative frequencies in possible worlds. So, in order to accommodate the view that probabilities are based upon chances, all that needs to be done is to consider frequencies in possible worlds as well as actual worlds. One way to tell if frequencies in the actual world reflect the chance is if the frequency ratios are invariant to a high degree under counterfactual conditionalization.

3.7 Conclusion.

I have examined five different interpretations of the use of probability in law. The main criterion was that the theory not assign arbitrary probabilities. The logical theory, developed in most detail by Carnap (1950) promised to offer an interpretation where this is a non-arbitrary relation of degree of entailment between evidence and conclusion. This is especially suitable as an interpretation of probability for law where probability is conceived as an objective logical measure between evidence admitted and the proposition at issue. The approach initiated by Carnap suffered from the problem that there are an infinite number of measures of probability and apparently no non arbitrary choice of measures. For many philosophers, including Carnap, the arbitrariness of logical measures of probability led to a subjective view of probability where no unique choice of prior probability measure is justified.

Some of the most skeptical philosophers and statisticians, like De Finetti (1937) reject the existence of any form of objective probability including frequencies and logical probability. The subjective theory of probability is fundamentally based upon the coherence and dynamic coherence arguments. The difficulty with these arguments is that they fail to apply to situations in which one has no reason to believe there is a cunning bettor. Another difficulty is that the subjective interpretation does not sufficiently single out a non arbitrary probability. Finally, the subjective interpretation does not express a "connection" between evidence and conclusion that legal reasoning requires.

The limiting relative frequency theory of probability suffers from the defect, as does the logical and subjective theories, that probabilities are arbitrary. This arbitrariness is due to the fact that the limiting relative frequency depends upon the order in which events occur; that a given finite frequency is compatible with any limiting relative frequency and since there is no "preferred" order of events, the probability based upon any particular ordering is arbitrary.
The main problem for finite frequency theories is that assignments of probability do not apply to single individuals. It is not sufficient that estimation of probabilities be correct more often than not in the long run. They must be correct in the very shortest run. Because individuals have a profound right not to be convicted if innocent, assignments of probabilities must apply to individual events. This requires the chance interpretation of probability. On the chance interpretation probabilities apply to individuals and are reflected by supporting counterfactual assignments of relative frequencies. The support of counterfactual frequencies explains the resiliency of frequencies under counterfactual conditionalization.

But the chance interpretation by itself did not solve all the problems with empirical interpretation. It does not provide an interpretation whereby probabilities are a logical relation between evidence and proposition at issue. The requirements that probability be a logical connection reflecting the chance of an event and expressing a rational degree of belief led to the consideration of Kyburg's theory. On Kyburg's theory probabilities are objective logical probabilities. This means that they conform to an inductive probability and function as inductive logic.

I construe Kyburg's theory as picking out which Carnapian c functions are rational degrees of confidence of a rational deliberating agent. These probabilities are not subjective and so do not suffer from any of the difficulties of subjective theories. But probabilities are legislative for rational belief. That is, probabilities constrain the beliefs that individuals have. On this view there is no need for precise probabilities to be used. The lack of a need for numerically sharp probabilities allows for a reasonable degree of precision in legal reasoning. The fact that probabilities are relative to the deliberator's knowledge provides a useful place for a deliberator's experience.
The Dynamics of Legal Deliberation.

4.0 Legal Deliberation.

In this chapter I examine the logic of legal deliberation. As I have argued there are a number of traditional principles which are accepted in law, however there is little agreement over the interpretation of these principles. For instance, a recognized principle is that a person be proved, according to the appropriate standard of proof, to have committed a particular act in order to be found liable. This is accepted by all sides. But the notion of proof here is notoriously ill understood in legal theory and practice. Proof may be construed to occur when a proposition at issue attains a high enough probability that it be accepted, or considered true. On a Bayesian view, however, proof does not confer upon the proposition at issue that it be considered true. To say that a person is proved guilty is only to observe passively that the probability of his guilt is sufficiently high, but not to regard the accused as guilty. I argue that principles of deliberation are best interpreted in a non-Bayesian manner.

A number of other principles can also be interpreted in a Bayesian or non-Bayesian manner. A number of principles will be examined including: conjunction: if A is proved and B is proved, then (A and B) is proved; deductive closure: if A is proved and A implies B then B is proved; consistency: the set of all proved propositions is consistent; completeness: proved propositions are formed based upon complete relevant evidence; presumption of innocence: an accused is presumed innocent at the beginning of proceedings; evidence is admitted only if it is relevant and reliable; presuming A: the "presumed" proposition A is inferred on the basis of a basic proposition B and background knowledge K; standard of proof: A is proved only if there is a critical number E such that the probability of A is equal to or greater than E. I argue that these principles be best interpreted in a non-Bayesian manner.

On both Bayesian and non-Bayesian views deliberation is a dynamic process in which the epistemic and conative states of a judge, jury or other decision-makers are forced through principles of epistemic and conative change to a decision. In the first section I present a model of deliberation by defining the epistemic and conative states of an agent and the notion of expansion, contraction and revision of epistemic and conative states through principles of deliberation. Deliberation can be understood in terms of the agent's epistemic and conative state and the changes in these states. Thus I relate the notions of proof, presumption, reasonable doubt, admission of evidence, judicial notice and presumptive inference, shifting burdens,
standard and burden of proof and so on in terms of the dynamics of deliberation. Hopefully both sides of the dispute can accept this general formulation of legal deliberation. The question that then arises is whether the dynamics of deliberation should be Bayesian or non-Bayesian.

The principles noted above will be evaluated according to whether or not an ideally rational agent involved in legal deliberation would be committed to them. In general, I argue that legal reasoning should conform a non-Bayesian pattern. In chapter three I already argued against the subjective view of probability as an explication of probability as used in the standards of proof. Section two contains a restatement of this position.

I examine the principles of conjunction, deductive closure and consistency. If one construes the system of beliefs as those beliefs which are adopted as true by the deliberator, then principles of deductive logic should constrain these beliefs. Deductive logic implies principles of conjunction, disjunction, negation and deduction for propositions. But Bayesian principles do not satisfy conjunction.

I also examine presumptive inference, judicial notice and admission of evidence. I show that the Bayesian dynamics for admitting evidence can not, in general, meet the traditional legal principle of conjunction except where full beliefs have the value of one. I also show how the presumptions of innocence is a non-Bayesian procedure because it involves a non Bayesian revision function. The type of presumptive inferences that are made in law by using a plausible principle of direct inference conflict with Bayesian procedures for evaluating evidence.

The principle that the probability of a proposition be based upon complete evidence is an important principle. However the notion of completeness is quite ambiguous. It may mean all the relevant evidence, all the admissible evidence, evidence sufficient to establish a proposition or evidence which has a sufficient weight. I argue that the Bayesian position is not able to capture the requirement that evidence be complete.

4.1 A Model of Deliberation: the Rational Deliberative Agent.

This section presents a model of legal deliberation. The basic idea is that there is an institution or agent who is called upon to make an epistemic judgement on issues which come to his attention. This is either a judge, jury, tribunal or other decision-maker. This epistemic judgement is an expression of one's epistemic attitude toward the proposition at issue given the
evidence. In legal proceedings these epistemic attitudes include belief, disbelief, partial belief, reasonable doubt, proof, disproof, prima facie prove or disprove, presume, (rebuttably or irrebutably) judicially notice, accept, reject, admit, not admit, judge, and suspend judgement. It is my contention that for legal purposes these attitudes can be defined in terms of partial belief and changes of partial belief.

An agent may express an attitude of categorical belief, qualitative belief or quantitative belief. A categorical or full belief is expressed when the agent believes \textit{(simpliciter)} that A or disbelieves that A. In a qualitative expression of belief an agent believes a proposition with a certain degree of confidence which can be ordinally graded. Quantitative belief is belief which is capable of cardinal measurement. The degree of belief in a proposition "A is X" is represented: \(B(A) = X\). The set of partial beliefs will be represented by \(S\).

Full belief, that is categorical belief, is related to partial belief. Sometimes full belief is thought to be a partial belief with complete confidence, or value one. This is an extreme position. In order to believe a proposition it is not necessary that an agent express total confidence in it. This is clear in legal reasoning where the level of confidence is often proof beyond a reasonable doubt or proof on a preponderance of the evidence. But this does not imply that no errors will occur in reaching a verdict. Full beliefs are related to partial belief by a rule that specifies the degree of partial belief required for full belief. The set of full beliefs will be denoted by \(K\) which is a subset of \(S\). The relation between the two sets is that \(K\) is a set whose degrees of beliefs is above a critical level \(E\) or above.

In order to represent these fundamental attitudes we can say that for a proposition \(A\):

\[
A \in K \text{ iff } A \text{ is believed} \\
-A \in K \text{ iff } A \text{ is disbelieved} \\
-A \in K \text{ and } A \in K \text{ iff } A \text{ indetermined}
\]

The epistemic state is fully specified when the function \(B\) is given for every possible proposition \(A\), and when the rule for relating full and partial belief is given. A rule of rationality defines the degrees of belief \(B\) for the agent at each time for each possible set of full beliefs \(K\). This rule specifies the confirmation function(s) that governs the epistemic states of the agent at every given instant. On the interpretation presented in chapter three,
probabilities are defined as relative to the set of beliefs \( K \) of an agent. A plausible rule of rationality is that \( P(A) = B(A) \) for all \( A \). On this view the degrees of belief at an agent at a time are relative to the information contained in his belief set. The probability is a logical \( \tau \) function which is determined by the information contained in \( K \).

Usually partial beliefs are taken to have the structure of the probability calculus. Thus a partial belief in \( A \) has a value equal to or greater than zero; a partial belief in a necessary proposition \( A \) has the value one; and a partial belief in the disjunction of contradictory propositions \( A \) and \( B \) have the value of the sum of the value of each partial belief of each proposition \( A \) and \( B \). The justification for such a rationality constraint is usually taken to be a Dutch book argument. (See Kyburg, 1983 and Skyrms, 1990) This is an argument which attempts to show that in order for an agent to avoid a sure loss on a bet he must conform his partial beliefs to the probability calculus. In a previous chapter I argued that coherence and dynamic coherence arguments do not succeed. But here I assume that the agent might adopt a convex set of confirmation functions due to uncertainty. A set \( Q \) of \( \tau \) function is said to be convex in the sense that if for any two \( \tau \) functions \( p \) and \( t \), if \( p \) is a member of \( Q \) then \( p + (1-t)p \) is a member of \( Q \) for all \( t \) between zero and one. The convex set of probability functions makes the agent's degree of belief interval valued, and so does not satisfy the standard probability calculus. But each confirmation function, though, does satisfy the standard calculus.

An important aspect of legal reasoning is its dynamic aspect. Epistemic attitudes change over time. Evidence which is admitted may then be contradicted and rejected. The rule of rationality also constrains the changes of epistemic states. Three important changes of epistemic states are expansions, contractions and revisions. An expansion occurs when a new piece of information is added to \( K \). This may be through admission of evidence or judicial notice, or presumption. A contraction occurs when a piece of information is deleted from \( K \). This may occur when a proposition is successfully rebutted or disproved. An important change is one of revision. Revision occurs when some beliefs are added and some beliefs are deleted. This may occur as a result of admitting evidence, presumptions, judicial notice, and shifting of the evidential burden of proof.

There are a number of different possible rules of dynamics of epistemic states. I will argue for the rule of expansion based upon the function \( + \). This function is the conjoining of \( A \) with \( K \) and forming the deductive closure. Forming the deductive closure means adding all the logical consequences of \( K \& A \). But a complexity enters in at this point because the set of beliefs need not have the structure of the propositional calculus. For example on the Bayesian view the set \( K \)
does not conform to the propositional calculus. It is difficult to know what propositions can be considered to be implied.

But I argue that the beliefs in K are those beliefs that are taken to be proved and true. Hence if these beliefs are taken to be true they should conform to the standard propositional principles. It is difficult to see how the conformity could fail to apply if the objects of belief are taken to be propositions. On this view the standard theorems hold: K is consistent; if A and B are in K, then (A&B) is in K; if A is in K then the logical consequences of A are in K. The justification of these principles of belief are straightforward. If K is taken to be true beliefs, then the principles of deductive logic imply the principles of consistency, conjunction and deductive closure.

The Bayesian proposal for inductive reasoning is quite bold. In order to save the application of Bayes' theorem as a principle of belief change, Bayesians hold that we never adopt the view that our beliefs are true categorically. (Horwich, 1986) We do not regard theories about the world as true or even approximately true-but as having a certain subjective probability. But this doctrine is hard to swallow in legal contexts. On the Bayesian view a very high degree of belief does not commit us to regarding it as true that he committed a certain act. Thus insofar as we should only convict a person if we believe him to be liable, the Bayesian proposal for standards of proof will not work.

The agent has a conative state which consists of desirabilities toward propositions. Like epistemic states, conative states also have a particular structure which is determined by postulates of rationality. Generally a number of qualitative conditions on preferences are assumed from which it is proved that there exists a desirability function satisfying the axioms. On the model here, the deliberator's desires may be somewhat vague due to lack of information reflected in his interval valued probabilities. So I assume his desires are also interval valued. The agent adopts a set of D functions, where D is Jeffrey's desirability function.

There is some ambiguity in the notion of belief and this needs to be clarified. Beliefs can be thought of as dispositional or voluntary. On this model belief is taken to be voluntary, but the voluntariness is of an ideally rational agent. Thus one should not expect a normal person to voluntarily be able to accept all the deductive consequences of his beliefs. Suppose one accepts the axioms of set theory. Obviously he can not accept all the theorems in any voluntary way. The postulates of rationality determine what one would be committed to if one were ideally rational.
Beliefs are directed toward propositions. Sometimes, as in Kyburg's (1974) account, beliefs are taken to be directed toward sentences in a formal language. The reason for this seems to be that it allows a definition of probability to be syntactic. Since the nature of propositions is an issue as yet unresolved, there is a degree of clarity to be gained in having attitudes directed toward sentences in a well defined language. But I see no reason not to think that propositions are not those things that are expressed in that well defined language.

In summary, a rational deliberating agent is one who has a set of beliefs and desires and a rule for changing these beliefs and desires given new evidence. The deliberative state of an agent consists of elements \(< B, D, S, K, +>\) where \(B\) is the deliberator's degree of belief over a set of propositions \(S\); \(K\) is a subset of \(S\), the full beliefs; \(D\) is the agents desirability of propositions \(S\); \(+\) is a dynamical rule for changing the values for \(B(S)\), and \(D(S)\) in response to new information. \(+\) is the rule of conjoining \(A\) to \(K\) and forming all the deductive consequences.

4.2 Probability and Proof.

The need for some sort of standard of proof as reflecting a person's reasonable belief has been felt since the earliest times in evidence law. It was laid down in *Woolmington v. D.P.P* that it is the duty of the prosecution to prove the propositions at issue beyond a reasonable doubt. The epistemic attitude at issue is here one of reasonable doubt. The interpretation of "proof beyond a reasonable doubt" is not settled. According to Denning, the standard of proof in a criminal case is usually proof beyond a reasonable doubt. Denning himself defined the notion of reasonable doubt in terms of probability. If the proposition at issue is highly probable then it is proved beyond a reasonable doubt. (*Miller v. D.P.P.*) In the same case Denning also drew the distinction between the criminal standard of proof beyond a reasonable doubt and the civil standard of proof on a preponderance of the evidence. On this view, proof on a preponderance of the evidence occurs when the proposition at issue is more probable than not.

It was pointed out earlier that at a certain level of partial belief full belief is attained. This is implicit in legal reasoning and is reflected in the standard of proof outlined above. As the degree of belief rises past some critical point full belief in the propositions occurs and it might be considered proved. As Thayer put it: "[The evidence] must not barely afford a basis for conjecture but for real belief." (Thayer, 1898) A recent statement of this position is to be found in *Smith v Rapid transit* that findings of liability cannot rest solely on mathematical chances but
must be supported by evidence which gives rise to actual belief in the minds of the tribunal. The appropriate interpretation of the standard of proof, then is that at some sufficient level of probability, not only proof occurs but full belief as opposed to merely partial belief occurs. This gives rise to the definition of full belief or proof:

\[ A \in K \text{ iff } P(A) > E, \text{ for some number } E. \]

The model of deliberation given here is also able to incorporate the idea of shifting the burden of proof, and the idea of prima facie proof or a prima facie case. In the case of a prima facie case or prima facie proof, the judge examines only the evidence offered by the accusing party. Then on the basis of this examination determines whether if this were all the evidence there were would it be sufficient for proof. In this way there is a counterfactual judgement to the effect that if this is all the evidence there were, then it would be sufficient to prove the proposition at issue. Prima facie proof can be regarded as a sufficient probability for proof based upon less than complete information.

The shifting of the evidential burden has been described as follows:

... it is obvious that as the controversy in the litigation travels on, the parties from moment to moment may reach points at which the onus of proof shifts, and at which the tribunal will have to say that if the case stops there, it must be decided in a particular manner. The test being such as I have stated, it is not a burden that goes on forever resting on the shoulders of whom it is first case. As soon as he brings evidence which, until it is answered rebuts the evidence against which he is contending, then the balance descends on the other side, and the burden rolls over until again there is evidence which one more turns the scale. That being so the question of onus of proof is only a rule for deciding on whom the obligation of going further, if he wishes to win rests. (Abrath v. North Eastern Railway Company)

The evidential burden, then, falls upon the party who must introduce new evidence in order to win the case. The shifting of the burden of proof on the model presented here occurs when the proposition at issue changes from being rationally believed to being rationally disbelieved. The Bayesian conception of the shifting evidential burden is analogous their conception of proof. The Bayesian never regards a proposition as true, but only has having a sufficient probability. The burden shifts, according to the Bayesian view, when the probability of the proposition at issue has crossed the threshold of probability identified by Bayesians as sufficient for proof.
It is difficult to know whether a Bayesian can consider *prima facie* proof as proof based upon less than complete evidence. This is because the subjective probabilities which the Bayesian works with do not need to be assigned on the basis of complete evidence. For the Bayesian, if evidence is introduced in, say, the swearing of information initiating a trial and this evidence results in a high probability that the accused is guilty, then a Bayesian must consider the accused proved to be guilty.

4.3 Rationality, Deductive Closure and Consistency.

A minimal requirement for the content of epistemic states is that they be consistent. This requirement was stated by Hempel (1962, 150-1) as one of the three necessary conditions of rationality in the formation of beliefs; the other two being deductive closure and completeness given below. It seems obvious that such a requirement on $K$ is needed for legal purposes. If the propositions in $K$ are regarded as true, then they must be consistent. This is a principle of propositional logic. Pragmatic considerations also require $K$ to be consistent. For if $K$ was not consistent then by the deductive closure (which requires that every logical consequence of $K$ is in $K$) the negation of every proposition in $K$ is also in $K$ and there would be nothing useful about $K$ for directing us to truth or in making decisions.

In legal contexts the requirement of consistency can also be supported by principles of legal interpretation. In *Riggs v. Palmer* a passage from Blackstone was cited (Coval and Smith, 1986, 92): "If there arise out of them [the laws] collaterally any absurd consequences manifestly contradictory to common reason, they are with regard to those collateral consequences void." This principle also obviously applies to facts.

Consistency: $(A \& \neg A) \in K$

A principle which is unsettled in law is the extent to which a series of inferences can be made. (Tillers, 1983; Cohen, 1987) In law these types of inferences are sometimes referred to as concatenated inferences. The traditional legal view was that no inferences on inferences should be drawn. Wigmore attacked this view and agreed with the position in *New York Life Insurance v. McNeely* that inference upon inference can be made. (Twining, 1985, 182ff.) This principle was stated as:
When an inference of the probability of the ultimate fact must be drawn from facts whose existence is itself based upon only an inference or chain of inference, it will be found that the Courts have, with few exceptions, held in substance, though not usually in terms, that all prior links in the chain of inferences must be shown with the same certainty as is required in criminal cases, in order to support a final inference of the probability of the ultimate fact at issue.

In criminal cases this rule can be interpreted as a rule of deductive closure:

Deductive closure: \( A \in K \) and \( A \) implies \( B \) then \( B \in K \).

The rule appears to be too strong for civil cases because it requires every inference to be proved beyond a reasonable doubt rather than, say on a balance of the probabilities. Weaker versions of the principle have been proposed, which hold merely that each element of a chain of inferences be proved according to the standard which is appropriate. (Tillers, 1983, sec. 41)

These considerations might motivate the deductive closure principle for civil cases as well as criminal. The deductive closure principle is also supported by the rules of presumption in law. It is often stated that a presumption is an inference that is drawn from proved facts. (e.g. Zuckerman, 113) Thus for example one statute reads "where it is proved that the accused occupied the seat ordinarily occupied by the driver of the motor vehicle he shall be deemed to have care and control of the vehicle unless he establishes that he did not enter the vehicle for the purpose of setting it in motion." (1971 3 C.C.C. (2d)) It follows then, that if a proposition has been proved by way of presumption then there is some other proposition that is proved.

The deductive closure principle was advocated by Hempel (1962, 150) The reasoning behind this principle is quite clear, as Hempel notes. If those propositions which are in \( K \) are believed, or accepted as truths, then it follows deductively that if they are true then any logical consequence of these propositions are also true. So it is contradictory not to accept the deductive closure principle. This is the reasoning that non-Bayesians can use in order to justify the deductive closure rule.

But the deductive closure rule is quite contentious even among non-Bayesians. Inference upon inference has been rejected in a number of cases. In Standard Accident Co. v. Nicholas it was said that "a case cannot be established by piling one presumption or inference upon another presumption or inference." (Tillers, 1983, sec. 41) In spite of the considerations advanced by
Hempel, the deductive closure principle is not accepted by Kyburg (1974) because it leads to the lottery paradox. This problem will be discussed in the next chapter.

A prominent requirement is that of coherence of partial beliefs. (Ramsey, 1926, De Finetti; 1937, Teller, 1976; Skyrms, 1980) In the last chapter I rejected coherence as being a requirement on epistemic states. The reasons given were that the coherence argument is inapplicable to non-betting situations.

Instead I impose a principle of rationality that a person's degree of belief in A equal the epistemological probability of A.

Rationality: \( B(A) = P(A) \).

Bayesians do not require any principles such as rationality because they do not believe that our degrees of belief conform to any objective probabilities.

In summary, the principle of rationality, consistency and deductive closure are traditional principles of legal reasoning which are representable in our model. A Bayesian cannot support these principles however.

4.4 Presumptions, Judicial Notice and Admission of Evidence.

One of the most important epistemic attitudes in law is presumption. But legal theorists have been particularly unclear on what presumptions are. The main view, from Thayer on down, seems to be that presumptions are inferences of "presumed" facts from basic facts. (Thayer, 1898) But presumption is not an epistemic attitude in this sense. Another use of the term presumption is when it is used as in presumed fact. In this sense one makes a inference — a presumption -- from a basic proposition to a presumed proposition. Presumption then is the attitude toward the inferred fact. This is the sense of presumption used here. For the purposes of this model presumptions can be identified with beliefs.

\( A \models K \) if A is presumed.
Judicial notice is related to presuming. As Thayer put it: "... in very many cases, taking judicial notice of a fact is merely presuming it, that is, assuming it it until there shall be reason to think otherwise." (1898, 308-9) One could restrict the concept of judicial notice in some way other than that described by Thayer but there seems to be no reason to do so. If the purpose of judicially noticing propositions is to avoid the unnecessary task of admitting evidence which is not admitted by council but is technically needed to reach a conclusion then a wide concept is needed. Propositions that are judicially noticed tend to be common knowledge which enter into the belief not through legal sources but rather through ordinary knowledge. Thus, for instance it has been recognized that Victoria is in British Columbia, or that Kamloops is a smelly city due to the pulp mill.

Zuckerman points out that not every bit of evidence that is assumed in obtaining a verdict is a fact which is judicially noticed. (1989, 79) The vast majority of necessary assumptions remain unstated. But the sense of belief and assumption offered here are not necessarily descriptive of the process of deliberation, rather they reflect the attitudes of an ideally rational agent and what he is committed to. It makes sense to think of the assumption which the judge is committed to as being part of those judicially noticed propositions. One could draw a distinction between assumptions which the judge explicitly makes as noticed, and those which are implicitly assumed. But as far as the logic is concerned this makes no difference.

A question that arises for judicial notice, traditionally, is whether evidence is admissible to rebut judicially noticed propositions. It is often thought that only indisputable facts can be judicially noticed. Thus if the fact is indisputable, no evidence can be admitted to rebut what is judicially noticed. On the view presented here, a judicially noticed fact is no different from commitment to other propositions and can be rebutted by admission of evidence. No fact is indisputable, and each fact that the agent is committed to at a given time is open to being deleted from K.

The admission, acceptance or receipt of evidence can be understood as adding a belief to one's belief set. The usual rule is that a belief A is added to K iff A is relevant subject to exceptions.
Simple admission of evidence: \( A^+K \) iff \( A \) is relevant

Non-Bayesians define relevance as a change in conditional probability. But the interpretation of probability is quite different than the Bayesian interpretation. Kyburg (1974), for example, defined a type of conditional probability of \( A \) as the probability of \( A \) when \( A \) is conjoined to \( K \). The resulting probability is not necessarily equal to the usual measure of conditional probability.

The simple admission rule is quite different than Bayesian reasoning. The simple rule gets support from legal reasoning in the form of reasoning utilizing full beliefs. In *Metro Ry. Co. v. Jackson*, it was said: "It has always been considered a question of law to be determined by the judge whether evidence which, if believed, and the counter evidence if any, not believed, would establish the facts in controversy. According to this rule, the admission of evidence is more complicated than simple admission. First one conjoins \( A \) to one's set of beliefs and forms the deductive closure. Through deductive closure new information is generated which may change the probability values of propositions and hence modify beliefs. The relevance of a proposition \( A \) can then be defined as the probability of \( A \) relative to the new set of beliefs \( K^+A \).

A different rule of admission is through Bayes' theorem:

Bayes admission: 
\[
P(S) = P(S \mid E) = \frac{P(E \mid S) P(S)}{P(E)}
\]

This rule of admission has been advocated by many legal theorists including Lempert and Saltzburg (1977, 85). According to Lempert and Saltzburg "The Bayesian equation describes the way the law's ideal juror evaluates new evidence." This rule of admission, though, is obviously unacceptable because it assigns probability of one to admitted evidence, since \( B(E \mid E) = 1 \). On the basis of this rule, once eyewitness testimony has been admitted, it is not possible to contradict it through admitting new evidence. In this way it fails to illuminate how the burden of proof could ever shift or how the testimonial evidence could ever be rebutted by another witness if conditionalization is the rule of admission. This problem is overcome through probability kinematics of Jeffrey (1965).
Probability kinematics: 
\[ P(S) = P(S \mid E) = P(E) P(S \mid E) + (1 - P(E)) P(S \mid \neg E) \]

The simple rule of admission is closer to how legal reasoning has traditionally worked. As a rule of expansion, probability kinematics has the peculiar property that it satisfies the conjunction principle in general only if full beliefs are taken to be certain, that is \( B(A) = 1 \). This anomaly was pointed out by Cohen (1977). Suppose that two probabilistically independent elements of a civil case C and D are established with \( B(C) = .7 \) and \( B(D) = .7 \) on the basis of eyewitness testimony of 70% reliability. Then both C and D are believed separately but the conjunction \( B(C \text{ and } D) = .49 \) is not believed because it falls below the .5 standard for civil proof.

Dawid (1987) has disputed this point and claims that he can evaporate the paradox. Dawid shows that under a wide range of prior probabilities the posterior probability of the conjunction of a number of elements will be higher than the prior probability of each individual element. Thus the Bayesian rule for admission of evidence satisfies the conjunction principle in a wide variety of cases.

Thus, for example, given a posterior partial belief of .5 or higher and given 70% reliable testimony the required prior probabilities for the case as a whole are given in the following table (Dawid, 1987) (where \( n \) - number of elements):  

<table>
<thead>
<tr>
<th>( n )</th>
<th>1</th>
<th>2</th>
<th>5</th>
<th>10</th>
<th>20</th>
<th>50</th>
<th>( \infty )</th>
</tr>
</thead>
<tbody>
<tr>
<td>required prior</td>
<td>.3</td>
<td>.26</td>
<td>.226</td>
<td>.213</td>
<td>.206</td>
<td>.201</td>
<td>.198</td>
</tr>
</tbody>
</table>

Table 3
Where \( n = \infty \) the following table indicates how the required prior probability must vary given the reliability of witnesses.

<table>
<thead>
<tr>
<th>Reliability</th>
<th>.5</th>
<th>.6</th>
<th>.7</th>
<th>.8</th>
<th>.85</th>
<th>.9</th>
</tr>
</thead>
<tbody>
<tr>
<td>Required prior</td>
<td>.5</td>
<td>.353</td>
<td>.198</td>
<td>.062</td>
<td>.02</td>
<td>.002</td>
</tr>
</tbody>
</table>

Table 4

There is, then, a wide number of prior probabilities for which the conjunction principle will be met. This also shows that if the reliability of the evidence is high, a suitable presumption of innocence may be imposed. The required posterior partial belief indicates the wide degree to which the conjunction principle holds given Bayesian reasoning. A Bayesian could argue, then, that our fondness of the conjunction principle is due to the fact that this principle usually holds.

Thus Dawid does not evade the paradox, he simply does not accept the conjunction principle. Cohen is indeed correct that the conjunction principle can not be met in general. It can not even be met in many plausible cases. All one has to do is recall Cohen's example which Dawid does not dispute. In the case where element A and element B are independent and each have a probability of .7, the probability of A and B is .49. Hence given a standard of proof >.5, A and B are each proved and (A & B) is not. Hence the conjunction principle fails.

An interesting and controversial principle is the presumption of innocence. "The presumption must be in favour of innocence .... We are not to presume guilt because the prosecutor alleges guilt... and until guilt is established, we must hold the presumption to be in favour of innocence." (Re Mckinley Case) This is a well recognized principle but like the other principles it is quite ambiguous and thus open to interpretation.

It is often thought that there can be no presumption of innocence in legal argumentation in any ordinary sense of presumption. In an ordinary sense of presumption of innocence a person is assumed or believed to be innocent. On the model presented here the proposition that A did not commit act B is conjoined to K. How can one be presumed innocent and yet presumably there is sufficient evidence to prove the person guilty? So it is sometimes held that the presumption of innocence is merely a restatement of the dictum of who has the burden of proof beyond a
reasonable doubt. (Wigmore, s. 2511, Cross, 1979, 114-115, Morton and Hutchison, 1987, 7-8) Tribe says that the presumption of innocence is a normative principle whereby we treat the accused as if they were innocent, although we know they are not. (1970,1371)

The view that presuming innocence is a restatement of who has the burden of proof is consistent with the Bayesian position. Presuming innocence, on this view, does not introduce any new attitudes toward the defendant beyond the notion of proof. Presuming innocence on this account just means presuming that the subjective probability be below the threshold set in the standard of proof.

However if one is a non-Bayesian one can accept that people truly are presumed to be innocent during a trial. Nevertheless there have been thought to be difficulties with holding that people are presumed innocent. The reason for the rejection of the ordinary notion of presuming innocence is the view that it contradicts ordinary notions of proof. But it can be readily seen not to be the case. What is done is a judgement is made as to whether or not the accused can be proved to have committed the alleged act and then at the beginning to the proceeding the supposition that the person is not liable is added to K. Then the task of the Crown is to admit enough evidence so that it overwhelms this proposition in the sense that it is rejected and deleted from K.

Presumption of innocence: -P∈K at time 0 of the proceedings.

As evidence is gradually admitted, the probability of -P will change and the probability of P will change. At some point, if the evidence is sufficient, the presumption of innocence -P will be contracted from K and will be expanded by P. The presumption of innocence, in this sense, conflicts with the Bayesian view since on the Bayesian view each change of epistemic states is governed by a rule of conditionalization. So it is not possible simply to presume the innocence of the accused at the outset of proceedings, since this is not allowed by conditionalization.
4.5 Completeness of Evidence.

The view that probabilities are relative to the total relevant evidence is often regarded as a necessary principle of inductive reasoning. (Carnap, 1950; Hempel, 1962) According to Carnap, in the application of inductive logic to a given knowledge situation, the total evidence available must be taken as the basis for determining the degree of confirmation.

The requirement that all relevant evidence be admitted in legal proceedings is also often cited as a principle of legal reasoning. (Phillips v. Ford Motor Company) A.J. Ayer (1972) has questioned whether total relevant evidence principle is justifiable in inductive reasoning. But Ayer failed to distinguish between two distinct principles of completeness of evidence. One principle is that the assessment of probabilities are relative to all the evidence that one possess. On most theories of inductive logic this principle is simply a consequence of the definition of probability. On Carnap's view, for instance, c(H|E), E is meant to include all the relevant evidence that one possesses. On Kyburg's (1974) model, for instance, probability is always a measure that is a function of all the evidence in one's set of beliefs.

It seems fairly clear that deliberation in legal proceedings is based upon the total relevant evidence that is known to the deliberator. This principle finds its expression, for instance, in the requirement that a person make a prima facie case, by offering some evidence in favour of the proposition at issue. Thus, Cross says that the judge must "inquire whether there is evidence which if uncontradicted, would justify men of ordinary reason and fairness in affirming the proposition which the proponent is bound to maintain." (Cross, 1979, 77) In other words, evidence must be given in favour or against the proposition at issue. Thus the following principle is suggested:

Completeness of evidence: the probability of the hypothesis is relative to K.

A second completeness principle is the principle that all evidence, not yet available be admitted. Such a principle is implicit in the burden of production of evidence in legal proceedings. Thayer says the burden of production stands for "the duty of going forward in argument or producing evidence ...." (1898, 355) This is distinct from the first completeness principle, then, because it requires the admission of new evidence. Indeed it is often thought that much of the justification of the adversarial process is its function of producing evidence. In
Phillips v. Ford Motor Company it was said: "This procedure assumes that the litigants, assisted by their counsel will fully and diligently present all the material facts which have evidentiary value in support of their positions ... in order to arrive at the truth of the matter in controversy." This motivates a strong principle of completeness of evidence:

Strong completeness: $K + E$, that is, all relevant, reliable and available evidence be conjoined to $K$ and the deductive closure formed.

In spite of the statement in Phillips v. Ford Motor Company the strong principle of completeness is not satisfied in law. The reason is that it violates principles of admissibility. In many cases evidence which is relevant, in the sense that the probability of the proposition at issue given the evidence is not equal to the prior probability of the proposition, is inadmissible because of conflicts with other values. So, for instance, if relevant evidence is prejudicial to the accused it may be inadmissible in some proceedings.

There are a number of possible responses to the justification of the strong completeness rule. Suppose one is randomly drawing subsamples of things from a population. It is often thought that admitting evidence in legal proceedings is of this sort of reasoning. (eg. Dunham and Birmingham, 1989) Consider the hypothesis that the proportion of things in the sample with $A$ is equal to the proportion of things with $A$ in the population. If the sample (the evidence) is continually augmented by randomly adding new members of the population to the sample (the new evidence) then, due to the law of large numbers, the probability that the proportion of $A$s in the population falls within some small interval around the proportion of $A$s in the sample becomes higher and higher. Thus, the admission of more and more evidence will increase the accuracy of verdicts over the long run.

Another attempt at justifying the rule is offered by Good (1966) and with specific reference to philosophy of law by Davidson and Pargetter (1987) and Shoeman (1988). On this view the acquisition of new evidence is always desirable. Suppose that the judge has the choice between admitting evidence $E$ or not admitting $E$. It was shown by Good that if the admission of evidence is not undesirable in itself in terms of cost or otherwise, the admission of evidence is always desirable in terms of its value in discovering the truth of hypotheses.
While both these views provide a rationale for completeness of evidence they justify too much. First, the reduction of error in probability evaluations conflicts with principles of admissibility. It is not always desirable to admit new evidence because of reasons of prejudice, privileges, protection of the innocent and so on. Secondly, if not all relevant evidence can be admitted for proof due to legal constraints then a lesser amount of evidence must be considered sufficient. There needs to be some criterion of the degree of completeness for sufficiency of evidence. In this matter the previous criteria have not proved correct. Thirdly, it may be that all the available evidence has been admitted. The evidence is complete in the sense that the error and desirability of admitting further evidence can not be increased. It may also be the case that based upon this evidence there is a high probability of a proposition A. Yet the amount of evidence may not be considered sufficient to prove A.

The most serious problem that arises is the last case where the evidence is clearly insufficient for a verdict, yet there is a high probability of the proposition at issue. Probably the most familiar type of instance where the evidence is clearly insufficient yet there is a high probability is the case of "naked statistical evidence". In T.N.T Management Pty. v. Brooks, for instance, The respondent's husband drove a semi-trailer which collided with a pantechnicon on a curve in the highway. There were no witnesses to the accident and both drivers were killed. The respondent sought damages from the appellant company. The pantechnicon was found on the wrong side of the road, and Gibbs, Stephen, Mason and Aickin J J. reasoned that the driver was negligent to some degree on a balance of the probabilities. Murphy J. also found in favour of the respondent but reasoned slightly differently. He argued that the evidence from the wreckage was too weak to support any conclusions regarding negligence. He further argued that there were three possibilities, the driver of the semitrailer was at fault, the driver of the pantechnicon was at fault, or both were at fault. Of these three possibilities, two favoured the respondent and so she should succeed on a balance of the probabilities.

Murphy acknowledged the implications of such an argument. He asks us to suppose that there is a lorry carrying a driver and three passengers and is that it is driven negligently. In the resulting accident all evidence indicating who the driver was is destroyed. Then he, says, if all three were employees of the lorry owner then the spouses of each employee could each recover damages by arguing that on balance of the probabilities each of the deceased were only passengers.

Cohen (1988) has advocated that evidence is complete in the sense that the probability has sufficient weight. Davidson and Pargetter have agreed with Cohen and have said that proof
beyond a reasonable doubt is attained only when there is a sufficient probability and sufficient weight. The following notion of weight is quite helpful (for closely related definitions see Davidson, and Pargetter, 1987; Skyrms, 1980):

\[ P(A) \text{ is Resilient : } \text{the resiliency of } P(A) \text{ being a be .5 or greater where} \]
\[ \text{the resiliency of } P(A) \text{ being a is } 1 - \max \{ P(A) - P(A|Q) \text{ for all } Q. \}

But the Bayesian calculus cannot handle the requirement that probabilities be resilient. Hence a Bayesian cannot account for the principle that evidence be complete. Skyrms (1980), a committed Bayesian, advocates the use of a type of resiliency but it is difficult to see what rationale a Bayesian could give for resiliency. The dutch book argument shows, according to Bayesians, that rational deliberators must comply with the axioms of probability in updating probabilities. But resiliency requires an additional constraint. Thus where a Bayesian would adopt a particular degree of belief toward \( A \), \( B(A) \), and \( B(A) \) is not resilient, a resiliency advocate would not hold \( B(A) \). Thus there is a conflict between those who require, in addition to the standard probability calculus the requirement of resiliency.

In conclusion evidence needs to be complete in the sense that the probability is resilient. But the resiliency requirement does not appear to be justifiable on Bayesian principles since on Bayesian principles decisions are based upon degrees of belief regardless of their resiliency.

4.6 Presumptive Inference.

Both Presumptions of fact and presumptions of law are supposed to be inferences from a basic proposition to a presumed proposition. Presumptive inferences, then, are a form of expansion of epistemic states through inference. For example if a person has been missing for seven years an inference may be drawn that the person is dead. Evidently this can be put in terms of a simple deductive argument. "... It really comes down only to this, that we are to look at presumptions simply as deductions which men make from the facts which are laid before them as evidence." (Gardner v. Gardner) If \( A \) is missing for seven years, then \( A \) is dead. \( A \) has been missing for seven years, therefore \( A \) is dead. The major premise is a generalization from the judge's background knowledge \( K \), the minor premise a basic fact, and the conclusion the "presumed" fact.
An important type of presumptive inference is direct statistical inference. As an example of a presumptive inference based upon direct inference consider the following case. In Sargent v. Massachusetts Accident Co. a woman was involved in an accident with a bus at night. The darkness precluded her from identifying the colour of the bus. Suppose that no other evidence is available regarding the driver or other properties of the bus. Suppose also that it is known that 75% of the blue bus are owned by Blue Bus Co. A sues the Blue Bus Co. for damages. A direct inference can be made that the probability that the cab involved in the hit and run was owned by Blue Bus Co. is 75%.

According to Kyburg (1974) a rule of direct inference is of the following form. Major premise: the proportion of As that are Bs is known to lie in the interval \([p,q]\). Minor premise: x is known to be an A and this information represents the best statistical information that is known about x's membership in B. Conclusion: the probability that x is a B is in the interval \([p,q]\).

More precisely suppose the set K contains the propositions:

1. \(C \iff x \in B\)
2. \(x \in A\)
3. \(fr(A,B) \in [p,q]\)

therefore:

4. \(\text{the probability that } x \in A \text{ is in the interval } [p,q]\)

where x is an individual event, B is a property, and A is the reference class for x being a member of B, and (3) states that the proportion of As that are Bs is in the interval \([p,q]\).

The above example is represented as follows:

5. \(C \iff x \in \text{set of blue buses}\)
6. \(x \in \text{set of buses owned by Blue Bus Co.}\)
7. \(\text{frequency of blue buses owned by Blue Bus Co. is .75.}\)

therefore:

8. \(\text{the probability that } x \text{ is owned by Blue Bus Co. is .75.}\)

On this model, each premise of the syllogism is admitted as evidence through a rule of expansion. Then inferences are drawn on the basis of this evidence and the rule of direct
inference. The crux of the rule of direct inference is that the rule picks out what class is the reference class that represents the best statistical knowledge one has. In the above case the reference class is the set of blue buses. This is the best evidence that is available. Often however there is quite a lot of information and there arises a choice between potential references classes.

On Kyburg's model of direct inference two basic criteria are used for selecting reference classes. Narrower classes are preferable to broader classes. So if there is evidence in the form of frequency data regarding a reference class A which is included in a reference class B, then B is preferred to A. Narrower intervals are preferable to broader intervals. Recalling that Kyburg's probabilities are interval valued, if there is evidence in the form of frequency data for classes A and B and the interval associated with A is narrower than B then A is preferred to B. These two desiderata conflict, however, and further rules are formulated to deal with such conflicts. For example on the first rule alone, the reference class is always a singleton class with probability [0,1], and so the narrower interval rule takes precedence. But the general thrust of direct inference is clear enough.

It turns out that, given this plausible rule for picking out reference classes, that direct inferences sometimes conflict with the Bayesian procedures for admitting evidence. This occurs when narrow classes offer narrower probability intervals. Bayesian procedures ignore the narrower classes. Thus Bayesian presumptive inferences violate the principle to use the best evidence available by way of using the narrowest class of data. Consider the following simple example suggested by Harper (1983), although he does not draw the same conclusion.

The principle of conditionalization says that probabilities should be revised in accordance with Bayes' theorem (on pain of incoherence):

\[
P(S)_{t+1} = P(S|R)_{t} = \frac{P(S)P(R|S)}{P(S)P(R|S) + P(-S)P(R|-S)}
\]

For example:

At 12:00 there is an urn which is known to contain 50% red counters and 20% counters that are both red and square. Let S be the statement that an arbitrary counter x is square and R be the statement that an arbitrary counter x is red. It is also known that 100 draws from the urn have been made and it was found that 50% of the red counters obtained are square. Thus the following statements are true at 12:00:
(10) \( P(S \& R) = .2 \)
(11) \( P(R) = .5 \)
(12) \( P(S \mid R) = .4 \)

At 12:01 a draw is made from the 100 counters that were previously drawn from the urn. What is the probability \( X \) that an arbitrary counter \( x \) drawn from the 100 counters is square, given that it is red?

The reasoning by the principle of conditionalization gives:

(13) \( P(S)_{12:01} = P(S \mid R)_{12:00} = .4 \)

If the direct inference principle is used the reasoning is as follows:

(14) \( X \) iff \( b \in S \) (class of square counters)
(15) the reference class for "\( x \in S \)" is \( R \) (set of red counters)
(16) the frequency of \( R \) that are \( S \) is .5.

therefore:

(17) probability of \( S \) is .5.

This simple example shows that the direct inference rule conflicts with Bayesian procedures. The direct inference rule picks out the set of red counters as the appropriate reference class because it is a subset of the set of all counters. But this consideration conflicts with Bayesian principles. Thus presumptive inferences are best interpreted in a non-Bayesian manner.

4.7 Conjunction.

A rule of proof that is often given is that there are two elements of a crime or tort, the \textit{mens rea} and the \textit{actus reus} and that each must be proved to constitute proof of the tort of crime. (Smith and Hogan, 1986, 29) Phipson held that "The presumption of innocence casts on the prosecution the burden of proving every ingredient of the offence...." (1982 s.4. my italics) In
Canada the law, as stated in *R. v. Graham*, requires that "... each essential ingredient of the offence must be proved beyond a reasonable doubt." (Morton and Hutchison, 1987) There is some ambiguity as to the interpretation of this principle. On a strict reading of these sources the rule states that proof of each element of an action is required for the entire action to be proved. A more likely interpretation of the rule is that proof of each element is necessary and sufficient for the proof of the entire action.

The principle that the proof of each element of an act was sufficient to prove the entire act was adopted, in effect, by Cohen (1977). This principle can be stated as a conjunction principle: if proposition A is believed and proposition B is believed the proposition (A & B) is believed. As I previously argued, the conjunction principle is a principle of deductive logic which applies to those beliefs we regard as true; that is those propositions in K.

conjunction principle: A∈K and B∈K then (A&B)∈K

On causal views of action, actions are sequences of events. These sequences can be divided into a number – perhaps an infinite number – of events. Any party to a dispute would be quite free to question the existence of any part of this sequence. The conjunction principle implies another principle, the conjunctive closure principle in the case of more than two propositions which have to be proved. According to the conjunctive closure principle if A is believed and B is believed ... and N is believed then (A & B & ... N is believed) for all N.

It is instructive to look at how elements of an action can be split up in contentious cases and require the need for a conjunctive closure principle. The mental element itself which initiates the sequence may be quite complex itself and each of the elements can be called into question. So for instance, if A is accused of recklessly causing the death of B, the recklessness includes the intention to drive the car, and the carelessness in driving the car, as well as an intention to drive carelessly knowing the risks associated, but not the intention to cause the death. The intention is itself a complex mental state which may require looking at the motivation and knowledge of the accused.

After proving all the elements of mental state there still may be a large number of events which need to be proved to have occurred. There can be very complex causal chains where not
only is the occurrence of each event in that chain questionable, so is the existence of a causal link between the elements of the chain.

As was mentioned above, expansion through Bayes' theorem or probability kinematics does not satisfy the conjunction principle. Some eminent legal critics on the basis of this problem, such as Eggleston (1983), are prepared to abandon this traditional principle of proof with little regard for legal tradition or its implications. But there is much more powerful argument against the conjunction principle which was given by Kyburg (1961) and will be examined in the next chapter. This argument, in legal circumstances is an attempt to show that, the standard of proof is inconsistent with the conjunction principle. But as I will argue in the next chapter, the requirement that the evidence have sufficiency resiliency solves the problem of inconsistency and saves the conjunction principle.

4.8 Conclusion.

The law of evidence consists of a number of principles which are in need of a coherent interpretation. The standard of proof, completeness of evidence, conjunction, consistency, and admission principles, deductive closure, shifting burdens of proof, presumptive inferences and presumption of innocence are all principles which may have either Bayesian interpretations or non-Bayesian interpretations. I have argued that the non-Bayesian interpretation is better than the Bayesian interpretation of these principles.
The Gate-crasher Paradox

5.0 Standards of Proof and The Gate-crasher Paradox.

It is generally thought that standards of proof in law can be given a "pascalian" probabilistic analysis. A hypothesis at issue in legal proceedings is taken to be proved when the probability of the hypothesis relative to the admitted evidence is sufficiently high. (See eg. Miller v. Minister of Pensions, and Kaye, 1988) It should be noted on the epistemic interpretation of probability advocated in chapter three, a proposition is considered proved when the lower probability of the proposition is sufficiently high. On this view probability is a convex set of c functions which forms an interval. So the epistemic interpretation of probability does not satisfy the pascalian probability calculus. However each c function does satisfy the pascalian probability calculus. In particular the lower probability singles out a unique c function. Hence the standard of proof advocated here can be considered a pascalian probabilistic analysis since the standard of proof is defined in terms of the lower probability.

L.J. Cohen has questioned the assumption that probability, as expressed in the legal standards of proof, satisfies the pascalian probability axioms. Cohen points out a number of anomalies that result from adopting a probabilistic standard of proof, but the anomaly that has received the greatest attention is the paradox of the gate-crasher. The argument presented by Cohen, in brief, is that if proof on a balance of the probabilities means proof on a pascalian probability of greater than .5, then injustice results. Thus proof in a civil suit cannot consist in proof on a balance of the probabilities. For example if 501 of 1000 persons in a stadium are gate-crashers then on the civil standard of proof, and no other evidence to the contrary, all must be considered gate-crashers because the probability is greater than .5 that each is a gate-crasher.

On Cohen's solution we need a notion of probability which is not only of sufficient probability but gives a greater weight, in the sense of Keynes, to the evidence. The notion of weight is, as yet, only intuitively understandable. As Keynes (1921, 1952, 17) put it: "... an accession of new evidence increases the weight of an argument. New evidence will sometimes decrease the probability of an argument, but it will always increase its weight." As Peirce (1932) argued: "... to express the proper state of our belief, not one number but two are requisite, the first depending upon the the inferred probability, the second on the amount of knowledge on which the probability is obtained." So on Cohen's solution the weight of evidence needs to be specified more clearly.
The paradox of the gate-crasher has evoked mixed responses. This is mainly due to the strong desire to maintain pascalian probabilistic standards of proof. Glanville Williams (1980) and Richard Eggleston (1983) consequently oppose the assumptions of the gate-crasher scenario. They feel that such situations cannot occur. David Kaye (1980), on the other hand, disagrees with Cohen's conclusion that a measure of weight must be used in addition to probability. But Kaye unwittingly offers a solution to the gate-crasher paradox similar in essentials to Cohen. Schoeman (1987), another critic of Cohen, suggests that a measure the epistemic utility of adding new evidence be regarded as an additional variable in a standard of proof. This is offered in disagreement to Cohen, but is essentially an acceptance of the point that a pascalian probabilistic standard of proof is not sufficient for law. Davidson and Pargetter (1987), and Dunham and Birmingham (1989) agree that in addition to high probability there must be sufficient weight to the hypothesis. However the analyses of weight contained in these two papers do not solve, or even attempt to solve, the problem of what degree of weight is required.

The failure of the previous critics to answer the paradox is due, I think, to notice the core argument in the gate-crasher paradox. I will argue that one version of the gate-crasher paradox is equivalent to the lottery paradox discovered by Kyburg (1961) and related to a remark of Hempel's (1962). The lottery paradox is that in some situations a pascalian probabilistic rule for accepting hypotheses yields contradictions. Now if standards of proof are regarded as in this theory, as a rule for when propositions are to be rationally believed, accepted or proved, then the gate-crasher paradox is an instance of the lottery paradox.

To be more specific the paradox results as follows. It is to be recalled that K conforms to a number of principles including: K is consistent: (A & -A) \in K, conjunction: if A \in K and B \in K then (A&B) \in K, and the standard of proof: if P(A) > E then A \in K for some number E. But the gate-crasher paradox is a counterexample to the ability of probabilistic standards of proof to satisfy the conjunction and consistency properties. And the practical outcome of this property is its injustice that Cohen notes.

A number of solutions to the lottery paradox have been proposed which demonstrably have the consistency property. Kyburg's solution is simply to reject the conjunction principle. The solution offered by Kyburg allows for the consistency property but at the expense of finding all the gate-crashers liable, but not believing that they are all guilty. This might seem like a ridiculous solution but given the range of options it is not implausible. The conjunction principle, as was noted, is difficult to believe to be false on a view where beliefs are taken to be about
single individuals. If the belief that A is guilty is about A, and the belief that B is guilty is about B, then it seems to follow that it should be believed that A and B are guilty.

On the view I advocate, probability is given an interpretation as an objective chance which requires a certain degree of resilience, or invariance under conditionalization. Resilience can be thought to capture some element of the weight of a probability. It is the weight of a probability that enables probability to be about unique events or individuals. Finally, given this interpretation of resiliency it can be shown with a chance of .5 or greater the standard of proof is strongly consistent.

5.1 What Are the Standards of Proof?

The classic statement of the criminal standard is due to Denning J. in Miller v. Minister of Pensions:

That degree is well settled. It need not reach certainty, but it must carry a high degree of probability. Proof beyond a reasonable doubt does not mean proof beyond a shadow of a doubt. The law would fail to protect the community if it admitted fanciful possibilities to deflect the course of justice. If the evidence is so strong against a man as to leave only the remote possibility in his favour, which can be dismissed with the sentence "of course it is possible but not in the least probable" the case is proved beyond reasonable doubt but nothing short of that will suffice.

In the same case Denning J. also drew the distinction between the criminal standard of proof beyond a reasonable doubt and the civil standard of proof on a preponderance of the evidence:

That degree is well settled. It must carry a reasonable degree of probability, but not so high as is required in a criminal case. If the evidence is such that the tribunal can say: "we think it more probable than not", the burden is discharged, but if the probabilities are equal it is not.

Legal proof, it should be emphasized, is not simply just a matter of inductive logic. There is a substantial value consideration built into proof. This value is built into proof through the standards of proof that are adopted. If one fails to notice that the degrees of proof reflect an ethical standard of protecting the community as Denning J. pointed out, or of protecting the innocent from conviction, one can be led to the view that these degrees of proof are illusory. As Hilbert J. said in R. v. Murtagh and Kennedy: "I personally have never seen the difference between the onus of proof in a civil and criminal case. If a thing is proved it is proved." But it is not proof simpliciter that the law is interested in, but a certain standard of proof for ethical
reasons. It may be the case that one can justifiably believe that some criminal act has taken place on a balance of the probabilities if accuracy is the only criterion of proof, but the law requires a higher standard in order to protect an innocent person from being convicted.

In reaching a verdict two types of mistakes can be made. The first type of mistake -- in criminal terminology -- is acquitting the guilty, and the second is convicting the innocent. On a simple analysis of standards of proof, the standard of proof on a balance of the probabilities is deviated from whenever the disutilities of the two types of mistakes are not equal. As a consequence of the ethical considerations brought into proof procedures by way of standards of proof the statement by Denning is open to some revision. The standards of proof are capable of varying from circumstance to circumstance depending upon ethical considerations. In the U.S. a third standard of clear and convincing evidence mediating between the criminal and civil standard is sometimes invoked. (McCormick, 1984, 989) In certain cases, involving fraud for instance, the standard of proof is adjusted to a higher level. (Bater v. Bater)

5.2 The Gate-crasher Paradox.

Cohen has two gate-crasher arguments but I will deal only with one of them. According to the first paradox, injustice results from having a standard of proof at .5 which implies that a person can lose a case when his case is arbitrarily close to the strength of his adversary; according to the second paradox, injustice results because every gate-crasher can be successfully sued for non payment. (Cohen, 1977, 75):

Consider for example, a case which it is common ground that 499 people paid for admission to a rodeo, and that 1000 are counted on the seats of whom A is one. Suppose no tickets were issued and there can be no testimony as to whether A paid for admission or climbed over the fence. So by any plausible criterion of mathematical probability there is a .501 probability on the admitted facts, that he did not pay. The mathematicist theory would apparently imply that in such circumstances the rodeo organizers are entitled to judgement against A for the admission-money, since the balance of probability (and also the difference between the prior and posterior probabilities) are in their favour.... Indeed, if the organizers were really entitled to judgement against A they would presumably be equally entitled to judgement against each person in the same situation as A. So they might conceivably be entitled to recover 1000 admission-moneys, when it was admitted that 499 had actually been paid. The absurd injustice of this shows that there is something wrong somewhere.

The second argument is that if the standard of proof is interpreted in terms of pascalian probability, then each rodeo fan would lose his case if brought before the court because there
would be a sufficient probability to prove that each was a gate-crasher. And so each would be forced to pay restitution to the rodeo organizers. This seems manifestly unjust.

Such examples may seem far fetched, but the unlikeness of the situation is irrelevant to the logical properties of proof and this is what is being investigated. In the next section I explore the relation between the lottery paradox and the gate-crasher paradox. The lottery paradox reveals that a purely probabilistic standard of proof can not meet the conditions of conjunction and consistency. So, insofar as the standards of proof are rules of inductive acceptance, the far fetchedness, if it is, of the example is irrelevant to the logical properties of the standard of proof.

5.3 The Lottery Paradox

As Cohen presented the gate-crasher paradox the problem is that in certain situations injustice results from the application of the standard of proof. But recognizing that the gate-crasher paradox is a lottery paradox reveals that it is not simply an injustice that results but that a contradiction results. The lottery paradox arises for any theory of inductive inference which uses a simple rule of inductive acceptance of hypotheses in terms of mathematical probability. (Kyburg, 1961) As an example, suppose that one accepts a hypothesis on the balance of probabilities. Consider a fair lottery with a million tickets numbered one to a million. Consider the hypothesis that ticket number 2 will lose. The probability of this hypothesis is very high and so if one accepts the hypothesis that ticket number 2 will lose. But the same argument could be run through with for every ticket. Thus one would end up accepting a million hypotheses, each one stating that ticket number n (n= 1 to a million) will lose.

By hypothesis that this is a fair lottery, however, one ticket will win. One may argue, as Kyburg does, that this is not yet a contradiction. The set of beliefs is not assumed to have the structure of the propositional calculus and its rule of conjunction. To yield a contradiction a principle of conjunction is needed: if S is a set of believed propositions, then the conjunction of every finite member of S is also believed. This principle is so plausible that it is difficult to imagine it being false. If one regards propositions as being true or false, then propositional logic should hold for these propositions. Perhaps one could define belief in a way that the conjunction principle does not hold as Bayesians apparently would have us do. The propositional content of beliefs would conform to a deviant logic. However, if the conjunction
principle is used, together with the other assumptions, the contradiction in one's set of beliefs that emerges is that no ticket will win and that one ticket will win.

I have argued that an attractive model of legal reasoning results from viewing legal deliberation as dynamic process of epistemic states. On this model the court or other decision maker's knowledge is represented by a set of beliefs K. The content of the set of beliefs is determined by rules of expansion, contraction and revision, which state under which conditions a propositions is believed. The rule of expansion that was argued for was the conjoining of A to K and forming the deductive closure. Assuming that the belief set conforms to the propositional calculus this implies that the conjunction principle, and consistency is met. Presumptions and judicially noticed facts would be included in this set of beliefs. These form the "experience" that supplements the logic of legal procedure. Evidence would enter into the set of beliefs through expansion. Once the court has some beliefs in K, various inferences can be drawn using standard logical principles and principles of direct inference.

Now the logic of the gate-crasher paradox can be examined. According to the relative frequency notion of probability the probability of x being a gate-crasher is the relative frequency of gate-crashers among total attendees. The court admits the evidence of money paid and total attendance to form part of K. On the basis of this information a direct presumptive inference can be made. The court can reason that the probability of being a gate-crashers in the stadium is .501.

The standard of proof in this case is proof on a balance of the mathematical probabilities. It is initially supposed, and forms part of the court's belief set K, that 501 out of 1000 fans are gate-crashers. On the standard of proof it follows that x is a gate-crasher. But one could continue to pick out individuals, and reason in the same manner. Each time a person is tried by the rule of expansion the conjunctive closure of propositions results. By the conjunction principle it follows that the deliberator believes that all attendees are gate-crashers. This conflicts with the initial supposition that only 501 fans were gate-crashers. On this reading the injustice is very concrete. The court awards each suit for nonpayment to the rodeo owner even though this contradicts the courts belief that only 501 fans were guilty of nonpayment. This is the point that critics of Cohen miss. Kaye for example says (1979, 104 my italics): "A court can consistently state that it accepts probabilistic proof, that the case has been established against an arbitrary spectator..."
The supposition that standards of proof are rules of inductive acceptance has support from two sources. One source of support is that legal proof is a form of inductive reasoning that has, in addition to assignments of probability to propositions, rules of acceptance. These rules of acceptance are rules specifying that at a certain level of probability one can accept or believe the truth of the proposition. Wigmore, a well known evidence theorist in law, was convinced that legal reasoning was inductive reasoning. (See also Twining, 1985)

A second source of support is the fact that in order for the successful reinforcement of legal norms verdicts must be understood to apply to individuals. This requires that propositions be believed fully and not simply with some degree of belief. It also requires that beliefs have the conjunction property. As has been seen, to regard a proposition as true implies the conjunctive closure principle. Furthermore, only with the conjunctive closure property can verdicts be thought to apply to unique individuals. Otherwise one does not know whether an arbitrary individual is convicted because he accidently fell into some class and was convicted due to an actuarial calculation. Only when it is regarded as true that a convicted individual committed the alleged offence can effective reinforcement of behaviour result. (See Nesson, 1985)

The view that verdicts should be based upon full belief or acceptance rather than on partial belief can be found not only in everyday life but in law as well. In Smith v Rapid Transit the court held that liability could not rest upon mathematical chances alone but rather a verdict could only be supported by "actual belief... in the minds of the tribunal." (Shaviro, 1989, 530) This might be interpreted as meaning that a person can be found guilty or liable only if it is believed fully and not just partially even to a very high extent. There must be in other words some level of partial belief at which one says this is sufficient to accept the truth of the proposition in question and act on that basis.

The lottery paradox shows that it allows inconsistent beliefs to be held in K. One way of dealing with this problem is to live with the inconsistency. Perhaps K being consistent is too much to ask for. Kyburg (1974) suggests that the conjunction principle be abandoned. In certain circumstances perhaps this is warranted. It does not seems warranted in legal proceedings, however, because the same result applies regardless of whether the conjunction principle is used. Each person brought before the court would be found guilty of non-payment and would have to pay restitution to the rodeo owner. The court would escape a contradiction but the unjust consequences would still follow. In effect the court would act as if it believed all accused were guilty, and yet only acknowledge that 501 were guilty.
In addition to the lottery paradox there is a second reason one might wish to give up the conjunction principle. The conjunction rule is dismissed by Bayesians as holding under only very special circumstances and not in general. (Dawid, 1987) Suppose that there are two facts A and B at issue in a civil suit. The standard of proof says that proof in a civil suit is a probability of greater than .5. Suppose that the facts A and B at issue are probabilistically independent and have a probability of .7 each. Then the probability of the case $C = P(A \text{ and } B) = .49$. This poses a problem. If the standard of proof is that a proposition at issue is established if its probability is greater than .5 then A is proved and B is proved. But by the multiplication axiom of the probability calculus the probability of A and B is less than .5. Hence A and B are not proved. This shows that a purely pascalian standard of proof does not satisfy the conjunction principle. One could take the other option and abandon or modify the probability calculus as Cohen does.

5.4 Attempts to Save A Purely Probabilistic Standard of Proof.

Cohen's own solution is to have a standard of proof based upon a sufficiently high Baconian probability. A Baconian probability is simply a grading of how far along a series of n increasingly complex tests a hypothesis can survive without being falsified. If there are n possible tests for a hypothesis and the hypothesis passes i tests then the inductive probability is $i/n$. One can formulate a standard of proof which says that at a suitable level of inductive probability a proposition is believed. Cohen says that on his interpretation of probability there can be no case against a randomly picked individual, such as a gate-crasher, because there is no inductively supported generalization from which one can infer that he is guilty. How this applies to the gate-crasher paradox is hard to follow, but apparently the level of inductive probability for the proposition that ticket number two will lose is very low because there are no tests that one could perform save drawing the ticket. There is no universal generalization under which tests could be performed. So the level of inductive support is 0. (1977, 318ff)

Cohen's own system is best understood as a method of measuring the weight of evidence. But the main problem with Cohen's system is that it is unable to give a non-arbitrary characterization of how tests are ranked in terms of complexity. Intuitively experimenters do assign degrees of complexity to experiments. But Cohen gives no non-arbitrary measure of degree of complexity. And since a measure of probability is simply a grading of a proposition at issue along increasingly complex experiments, the problem arises as to which ordering of tests is the
correct one. Hence the problem is which level of probability is the correct level. This problem needs to be overcome, and perhaps it can be, in order for Cohen's system to be implemented.

Glanville Williams (1980, 304) has responded to Cohen by saying that the problem with the standard of mathematical probability is not with its mathematical structure, but it is that in the gate-crasher incident the standard does not sufficiently distinguish the defendant from other possible suspects. Accordingly he feels that there must be a rule of proof that evidence must focus on the defendant only. (1980, 305) But Cohen would agree with this point. There is no evidence specifically against the accused. The point of division between Williams and Cohen is that Williams thinks that there is or should be a law of evidence saying that evidence must be directed toward a unique individual in addition to the probabilistic standard of proof. Whereas on Cohen's view the fact that evidence applies to the individual is due to an adequate interpretation of probability.

Eggleston (1983) agrees with Cohen that Williams' (1980) response that "evidence" be taken to mean evidence relating to only one defendant or hypothesis is incorrect as a matter of law. If Williams' solution were taken up, then one could not be arrested or sue another in cases where there is no evidence against a particular person. But a plaintiff is allowed to sue a number of defendants if he is unsure about which one is liable and let them fight it out. An example appears in Cooke v. Lewis for instance, where the plaintiff was shot while hunting by one of the two defendants, both of whom fired at different birds at the same time. The evidence is that one of these two hunters shot the third hunter, but there is no specific evidence against either of them individually. Yet the burden of proof is shifted and the defendants have the burden of proving that each did not fire the shot.

Eggleston goes on to say, in effect, that a case like the gate-crasher case could not happen. (1983, 41) But there are a number of cases which are similar to the gate-crasher paradox. The case of T.N.T Management Pty. v. Brooks, was already cited. Murphy J., in this case, accepted the principle that where there is purely "naked" statistical evidence high probabilities based upon this evidence is sufficient for proof. Murphy, by his own reasoning would convict all the gatecrashers. In his imaginary example he asks us to suppose that there is a lorry carrying a driver and three passengers and is driven negligently. In the resulting accident all evidence indicating who the driver was is destroyed. Then he, says, if all three were employees of the lorry owner then the spouses of each employee could each recover damages by arguing that on balance of the probabilities each of the deceased were only passengers. But then Murphy J. is
subject to the contradiction that only one person drove negligently and all three drove negligently.

In *Sindel v. Abott Laboratories* plaintiff brought a class action suit against eleven drug companies who manufactured and distributed DES. Plaintiff alleged that Abott knew or should have known that DES was ineffective in intended use of preventing miscarriage but caused carcinoma in the daughters of mothers who ingested DES. However, there was no evidence available on which company who actually manufactured the DES plaintiff received since the products form the eleven were sold in an indistinguishable form. The class action was dismissed. Later, on appeal the California Supreme Court reversed and allowed plaintiff to recover from each defendant in proportion to their market share. The judgement indicated that the market share served as an acceptable "... measure of the likelihood that ... the defendants supplied the product which allegedly injured plaintiff." Yet this measure of "likelihood" is based upon purely statistical evidence.

David Kaye has written a number of papers defending a subjective Bayesian approach to legal proof. He disagrees that weight of evidence must be considered when proving legal hypotheses and defends a purely pascalian standard of proof. But, Kaye says: "As soon as we reject the invitation to equate the proportion of gate-crashers in the audience to the probability on which the plaintiff's case should turn, the paradox evaporates." (1980, 108) His argument is as follows: (1980, 107)

Suppose a juror accepts the statistic about the number of paying customers at face value. For him, the subjective probability, P(X), that defendant did not pay is .501. But, if he stops to reflect upon the fact that this is all there is to the case, he should revise his probability in light of this new item of "evidence". Under the preponderance of the evidence standard, he should find for defendant if the revised subjective probability is P(X|E) is one half or less.

How is one to interpret this answer? It admits that once one sees that "this is all there is to the case" then the proof is affected. Kaye proposes that the probability itself should be lowered to below .501. Kaye seems to be guilty of not distinguishing between probability and the weight of evidence. What Kaye takes to be "evidence" is simply the fact that the evidence is incomplete. The measure of weight Kaye unwittingly offers seems to me to be exactly on the right track and is in the tradition of weight as a measure of invariance under conditionalization. So in effect, Kaye has agreed that a measure of weight of evidence is needed.

Ferdinand Schoeman (1979) argues a number of points against Cohen. One point he makes is that the court should assure that in gate-crasher type cases there is a high threshold of proof
and all relevant evidence is admitted. (1987, 85) But a higher standard of proof does not help solve the paradox because the gate-crasher paradox can simply be restated for the higher standard. Suppose, for instance, that there is undisputed evidence that one thousand people were on the seats at a rodeo, but that only 100 paid for admission. On the assumption that proof of a criminal act is established by a high mathematical probability this evidence alone should suffice against any person picked at random off the seats and tried for fraud if he has nothing to say in his own defence. Yet that would surely be unjust. So proof in a criminal trial cannot consist of establishing a high mathematical probability.

Schoeman's second response is that having all the relevant evidence before the court would allow one to have evidence which would sufficiently pick out one person from other possible suspects. (1987, 85) But the total evidence is not always available in the common law legal system. Memories fade, evidence is destroyed, and often the total evidence is not available. In the common law tradition there are numerous exclusionary rules where relevant evidence is excluded due to privilege, prejudice to the accused, public policy and other considerations. It is simply not possible to have all relevant evidence available all the time. But in addition this answer would not work even if all the available relevant evidence were admitted. In the gate-crasher paradox all the relevant evidence available is admitted and the paradox results. This is because the paradox has nothing to do with having the total relevant available evidence admitted. Rather the paradox is due to the fact that the principles of conjunction, consistency and the standard of proof can not be satisfied as a logical truth. So this line of reasoning fails.

A softened principle of total evidence is suggested by Schoeman (1987, 86). He advocates that in addition to a high mathematical probability it should be required that probabilities not be volatile. One could define volatility in terms of the expected value of new information. If the expected value of new information is high then the standard of proof has not been met. Criminal cases would then require a higher expected value of information for instance because the inherent undesirability of convicting innocents. Of course, this is just to agree with the point that a high probability is not sufficient for proof. But this suggestion obviously fails to solve the paradox. For each level of volatility one can specify, there exists a probability level that does not satisfy conjunction and consistency. Hence for each level of probability one can specify, a new gate-crasher paradox can be offered.
5.5 The Weight of Evidence.

The attempts surveyed so far fail to solve the gate-crasher paradox. They contain a number of suggestions, though, which tend to confirm the view that probabilities must have a certain amount of weight to apply to individuals. But the definitions of weight we have looked at fail to solve the problem. However, I believe a notion of resiliency, similar to the one in Skyrms (1980) and Davidson and Pargetter (1987) can solve the problem. This solution essentially involves not counting the frequencies in the gate-crasher scenario as probabilities. Rather the frequencies with a sufficient weight are counted as chances. Then probability is defined in terms of our knowledge of these chances.

The strongest suggestion is that an additional variable of weight be included in standards of proof. This does not automatically mean that the probabilistic rule does not conform to the Kolmogoroff calculus. Propensities or chances can be defined, as Skyrms does, as probabilities which have sufficient resiliency. Thus if probability in the standards of proof is understood as based upon our knowledge of propensities, a consistent standard of proof can be defined. But let's first compare this view with alternative views of weight.

Keynes discussed the notion of weight:

As the relevant evidence at our disposal increases, the magnitude of the probability of the argument may either decrease or increase, according to the new knowledge strengthen the unfavourable or favourable evidence; but something seems to have increased in either case— we may have a more substantial basis upon which to rest our conclusion. I express this by saying the an accession of new evidence increases the weight of an argument. New evidence will sometimes decrease the probability of an argument, but it will always increase its weight.

Sometimes the weight of evidence is identified with the error level of probabilities by eg. Carnap (1950). One can reason "inversely" via Bernoulli's theorem from the relative frequency in a randomly picked subsample of a population of interest that the relative frequency of the population is within a given interval, and that this result would hold, say, 95 times out of 100, when a subsample was randomly picked from the population. The error level in this case is given as 5%. Bernoulli's theorem guarantees that as the sample gets larger and larger the error level gets smaller and smaller.

However, the error level by itself, does not appear to be an appropriate measure of the weight of evidence although it does appear to be a component of weight. All that the error level of, say, 5% expresses is that if a sample were randomly picked from a population 100
times that 95 of those times, the relative frequency of that sample will be within some interval. Relative frequency is taken to be a property of actual sequences but error levels are not properties of sequences, rather error levels are properties of sets of possible sequences. It is interesting that error levels are counterfactualizable in this manner.

The problem with the error level as a measure of weight is quite subtle, I think. The problem is that if a probability has an error level of 95/100, it is possible that each of the 100 sequences of trials are very similar to the actual sequence of trials. In this way the weight that a set of sequences has is quite arbitrary depending upon which sequences have been picked out. Suppose, for example, that a poll is done of voter choices for political parties. It is reported that 50% of voters support the liberals and that this value is accurate, within a margin of error, 95 time out of 100. That is, if the poll were conducted 100 times the reported value would hold 95 times. This may be a comforting level of error, but what if the 95 trials contained the same respondents every time a random selection was made. Of course, one would expect agreement. What is needed for a measure of weight is a suitable measure of how the probability would change under many varied circumstances. There is no guarantee that a randomization procedure would guarantee the appropriate variety of sequences to be picked.

Davidson and Pargetter have agreed with Cohen (1987, 183) and have proposed a definition of weight. Davidson and Pargetter argue that to prove beyond a reasonable doubt means that the probability of the hypothesis on the evidence is high and that the probability of guilt must have a sufficient weight. According to this notion of weight, one considers the change of probability that a proposition undergoes that a new piece of evidence makes. If the difference is small then the resilience of the probability is high with respect to the evidence. The weight of evidence is taken to be high if the resiliency of the probabilities is high relative to all possible bodies of evidence that are probable. (1987, 187) The definition they give is that the weight of \( P(H) = 1 - P(E \mid P(H \mid E) - P(H))^1 \) for each possible body of evidence \( E \). Davidson and Pargetter thus include in the weight of \( P(H) \) the probability of the evidence on which it is based.

This approach is correct, I think. But there are some in adequacies to be corrected. The most basic inadequacy is that Davidson and Pargetter do not show how this notion of resilience solves the gate-crasher problem to which they allude. Secondly the requirement that the weight of evidence is relative to all possible evidence is too strong. Because then, the resiliency of a probability always must be 1. Since the hypothesis \( H \) must be included in all possible evidence \( 1 - P(E \mid P(H \mid E) - P(H))^1 = 1. \)
Let me specify a form of resiliency which meets these problems:

**resiliency of probability:** The resiliency of the probability of A being a is equal to $1 - \text{max} \mid P(A) - P(A \mid B) \mid$ for all B

If resiliency is defined as the degree of invariance under conditionalization, that is the resiliency of a probability $P(A) = 1 - \text{max} \mid P(A) - P(A \mid B) \mid$ for all B. Now one can define a chance as a probability which has a sufficient resiliency. The question is what body of possible evidence should resiliency be relative to? If it is relative to all possible evidence, then the resiliency is always 1 because the possible evidence always includes the hypothesis. If probabilities are dispositions of things or persons then it makes sense that one can vary elements of the surroundings except for the essential aspects of the object or person and the probability will remain the same. So the natural requirement is that B includes all propositions that do not describe properties that are essential to the object or agent.

The probability of a propositions is said to be resilient if its invariance under conditionalization of other accepted propositions is greater than .5. If this is so, then the set of beliefs held by the adjudicator are consistent. Suppose there are two beliefs A and B that are inconsistent, then $P(A \mid B) = 0$. Further, suppose the probability of A is .5 or greater in order for A to be proved. The resiliency of $P(A) = 1 - \mid P(A) - P(A \mid B) \mid = .5$ or less. Thus, because the resiliency of $P(A)$ is not greater than .5, A is not a member of the set of beliefs. The standard of proof now states that there be a sufficient resiliency of probability as well.

This rule evades the gate-crasher paradox. In this scenario the probability of each person being a gate-crasher is .501. On the basis of a standard of proof of greater than .5 all the attendees are believed to be gate-crashers. But if it were not the case that they were all gate-crashers then the change in probability would be greater than .5. Hence the resiliency of the probability of being a gate-crasher is below .5.
5.6 Resiliency, Chance and Probabilistic Standards of Proof

Cohen complained that standards of proof could not be probabilistic. In a sense Cohen is correct. The probabilities need to be resilient. But if probabilities were required to be resilient then the objection would fail. There seems to be a natural relation between chances and resilient frequencies. Consider a newly minted coin. This coin has never been tossed and will be tossed and then destroyed. What is the probability of tossing heads on that toss? On the finite frequency account the outcome will be heads or tails and thus the probability, apparently, either 1 or 0. But the chance of tossing a head is still thought to be .5. Why is this?

Hacking (1965, 10) notes that:

" ... 'frequency in the long run' is all very well, but it is a property of the the coin and tossing device, not only that, in the long run, heads fall more often than tails, but also that this would happen even if in fact the device were dismantled and the coin melted. This is a dispositional property of the coin: what the long run frequency is or would be or would have been."

Unlike judgements of relative frequency, the chance that the coin will land heads supports counterfactual judgements. If the coin were tossed 100 times, then the relative frequency of heads sould probably be .5. In possible worlds just like this world except the newly minted coin is tossed 100 times, 50 outcomes would be heads is probably true. Whereas on the relative frequency judgement nothing can be said regarding the counterfactual situation where the coin is tossed one hundred times. Chance, then exhibits a certain resiliency under counterfactual conditionalization. The coin, whether it is ever tossed or not still has the property that if it were tossed, a certain proportion would likely turn up heads. So if one identifies the probability with a relative frequency, the test for whether that frequency is the same as the chance is whether the frequency is sufficiently resilient.

Just as coins have dispositions people have dispositions. Chances can be most naturally interpreted as the disposition that a person has in a given situation to commit a particular act. The disposition that someone has is most naturally evidenced by relative frequencies of similar events, and by one's mental characteristics such as character, habits, opportunity, and beliefs. So in the gate-crasher scenario, the probability that arbitrary A is a gate-crasher is better estimated from A's habits, character, ability, opportunity, than any non-resilient frequency data about the relative number of gate-crashers to attendees.
The interesting point about evidence of habits and character, which is evidence of a disposition, is that the admissibility of such information is unsettled. As Cross understands the law (1979, 25): "If the evidence is relevant only as showing the accused to be by his nature or disposition a person likely to commit the crime alleged, the evidence is inadmissible." But in *R v. Scopelliti* the judge ruled that "... the admission of similar fact evidence against an accused is exceptional, being allowed only if it has substantial probative value on some issue, otherwise than as proof of propensity (unless the propensity is so highly distinctive or unique as to constitute as signature.)"

5.7. Conclusion.

If legal proof is interpreted as inductive logic then standards of proof can be interpreted as rules of inductive belief or acceptance. Standards of proof then become subject to the lottery paradox and the gate-crasher paradox becomes an instance the lottery paradox. The objections of Williams and Eggleston ignore the central problem, that the pascalian standard of proof conflicts with consistency and conjunction. Cohen's view that legal hypotheses must have a sufficient weight is supported by Davidson and Pargetter and (unwittingly) by Kaye and Schoeman. Cohen's characterization of the weight of evidence depends upon a non-arbitrary degree of complexity of tests. But Cohen has yet to spell out how to measure the degree of complexity of tests.

If the weight of evidence is defined in terms of the resiliency of probabilities, that is, the degree of invariance under conditionalization, then the lottery paradox can be solved while the conjunction principle is maintained. In the gate-crasher example, none of the attendees can be successfully sued. Finally, if chances are interpreted as those probabilities with sufficient resiliency, then the probabilistic standard of proof can be maintained so long as a greater than .5 probability is required for proof.
6.0 Introduction.

A standard of proof states the criteria for rational belief, or proof of propositions at issue in legal proceedings. In rough terms, there are two standards of proof which are generally applied in law. In criminal proceedings the standard is that propositions at issue must be proved beyond a reasonable doubt and in civil proceedings propositions must be proved on a preponderance of the evidence. (Woolmington v. D.P.P.; Miller v. Minister of Pensions) But the probability required for proof defined in these cases is thought to follow from a more general ethical rule such as maximizing social welfare or protecting individual rights. The purpose of this chapter is to argue against the maximizing social welfare rationale for standards of proof.

In order to be more quantitatively definite than "beyond a reasonable doubt" or "on a preponderance of the evidence" the standard of proof can be measured in terms of epistemic probabilities. Many writers have pointed out that the standards of proof cannot be expressed in terms of the exacting standards of the standard probability calculus. (Tribe, 1970, 1358) The need by a judge or juror to express a precise degree of belief of .9 say does not reflect the fact of uncertainty that exists in such a judgement. But these criticisms are easily met by specifying the standard of proof in terms of epistemic probability. On the epistemic interpretation probability is an interval valued function so that there will exist an upper probability \( p^* \) and lower probability \( p \). The interval \([p, p^*]\) reflects the degree of uncertainty that a deliberator has regarding the proposition at issue. But the standard can also be stated in an exact manner by requiring that the standard of proof is stated in terms of the minimum lower probability that must be attained in order for a proposition at issue to be proved.

Once the standard of proof is understood as specifying the degree of lower probability that is needed for establishing a proposition the question then arises as to what probability is necessary for proof to occur. In order to assess the merits of alternative probabilistic standards of proof, the nature of the legal decision needs to be clarified. Verdicts are made in conditions of uncertainty. Two types of errors are possible. One error—in criminal terminology—is convicting an innocent person and another is acquitting a guilty person.

On one version of rule utilitarianism, the object of any rule is to maximize the welfare of society, in the sense that the welfare is the average utility of society. So in this view \( D(A) = \)
1/n \sum D_i (A) for each individual i=1 - n in society. (Hare, 1976, Harsanyi, 1977) On this view the standard of proof can be assessed in terms of the changes in the average utility of society. The chief difficulty with utilitarian based standards of proof, as I will show, is that the standards infringe upon the right of individuals not to be convicted if innocent.

This infringement of rights by standards of proof advocated by utilitarianism is most clear in the case of multiple defendants. In civil cases the rule of proof says that a person is proved to have committed an act when the lower probability that defendant A committed the alleged action is greater than that of the other defendants. In cases involving two defendants this rule implies that a person be shown to have committed the offence more probably than not. But in the case of more than two defendants a person can be found guilty even though it is extremely improbable that the defendant committed any illegal act.

According to a rights based political view like Rawls (1970) the satisfaction of individual rights in society enjoys a priority over any considerations of utility. The chief difficulty with the rights based view is how to show that a particular standard of proof is required. The assumption of a rights based view of the standards of proof is that an individual has a right not to be convicted if innocent. Now, there may be some question as to what level of accuracy a person is entitled to on this view, but certainly justice requires a person should be entitled to a level of accuracy which implies that it is more probable than not that he did the alleged act. But on a utilitarian view, as will be shown, one is not entitled to even this level of accuracy.

6.1 Utilitarian Standards of Proof.

On a utilitarian analysis of the standards of proof the deliberator's desires for a propositions A is D(A) = 1/n \sum D_i (A) for each individual i=1 - n in society. The social utility D(A) is simply the sum of the utilities of an act of the deliberator on each individual member of society and the deliberator ought to maximize D(A). On a model of rational behaviour the effects of verdicts are simple enough. Convicting an guilty person will have a reinforcing effect on legal behaviour because of the increased expectation of being convicted if guilty. Convicting an innocent person increases a person's expectation of being convicted if innocent and will have a reinforcing effect on illegal behaviour. Raising or lowering the standard of proof is the means by which these expectations are controlled.
However the supposition that society should maximize the social utility allows society to infringe on fundamental rights of individuals. This happens in a fairly straightforward way through the standards of proof in law. In this and the next section I show how utilitarian standards of proof are derived. Then in 6.3 I argue the utilitarian standards are unjust.

The idea to derive standards of proof in law through decision theoretic methods is due to Kaplan (1968). Kaplan’s model was accepted by Mr. Justice Harlan in In re Winship, where Harlan said: "The reasonable doubt standard plays a vital role in the American scheme of criminal procedure. It is the prime instrument for reducing the risk of convictions resting on factual error." However on Kaplan’s original model the relation between decision theory and ethics is not adequately treated. There appears to be simple conflation between the standard of proof society ought to adopt and the standard of proof that is in the interests of the judge or jury. Kaplan’s argument apparently results in the conclusion that it is in the decision maker’s interest to adopt a given standard of proof, and not that that the standard is ethically required in any sense.

There are two essential premises to the derivation of the preponderance of the evidence standard. First, the desirability of convicting an innocent or acquitting a guilty individual are considered equal. On the model of deliberation presented here this means that $D(\text{convicting an innocent individual}) = D(\text{acquitting a guilty individual})$, where $D(A) = \frac{1}{n} \sum D_i (A)$ for each individual $i=1 - n$ in society. Second, the deliberator is to choose the action which maximizes his desirability, which by Jeffrey’s expected desirability theorem, is equivalent to maximizing the conditional expected desirability of an action.

The following is a decision matrix for legal deliberation. (Using criminal terminology for ease of exposition.)

<table>
<thead>
<tr>
<th></th>
<th>Guilty</th>
<th>Innocent</th>
</tr>
</thead>
<tbody>
<tr>
<td>Convict</td>
<td>+?</td>
<td>-?</td>
</tr>
<tr>
<td>Acquit</td>
<td>-?</td>
<td>+?</td>
</tr>
</tbody>
</table>

Table 5
There are four types of cases, then, that one need be concerned about:

(1) C & I: X is convicted and X is innocent (for certain)
(2) C & G: X is convicted and X guilty (for certain)
(3) A & I: X is acquitted and X is innocent (for certain)
(4) A & G: X is acquitted and X is guilty (for certain)

Kaplan's argument (roughly) went as follows. Let the desirability of convicting an innocent person be denoted by D(C&I) and the desirability of convicting a guilty person be denoted D(C&G), the desirability of acquitting an innocent person D(A&I), and the desirability of acquitting a guilty person D(A&G). Then the deliberator has the choice of acquitting the accused or convicting the accused. Since he is a desirability maximizer he will choose to convict if and only if D(C) > D(A). (Let P be the lower probability, and D be the lower desirability by the same generalization.) Kaplan then claims that it follows that a jury will convict if and only if:

(5) \( P(C&G) \ D(C&I) > [1- P(C&G) \ D(A&G)] \).

From which it is claimed that:

(6) \( P(G) > 1/ [1+ (D(C&I)/D(A&G))] \)

The second assumption, it is to be recalled is that the desirability of convicting and innocent person is equal to the desirability of acquitting a guilty person. Then since D(C&I)=D(A&G), it follows that, P(G) is greater than 1/2.

But it is clear that Kaplan's assumption (5) is mistaken. The desirability of conviction is not just the product of the probability of innocence and the desirability of convicting an innocent (for certain). Recall that it is a consequence of Jeffrey's theorems that:
(7) \[ D(A) = P(S \mid A) D(S \& A) \] where \( A \) is a basic act and \( S \) is a state of affairs.

So if the basic act is convicting and the state of affairs is whether the person is guilty or innocent, the correct expression for the desirability of convicting is: \[ D(C) = D(C \& G) P(G \mid C) + D(C \& I) P(I \mid C) \]. And the analogous point holds for the desirability of acquittal, that is, \[ D(A) = D(A \& G) P(G \mid A) + D(A \& I) P(I \mid A) \]. In order to derive the more probable than not standard a further assumption is needed. Kaplan assumed that the desirability of convicting an innocent was equal to the desirability of acquitting a guilty person. If truth is all that one is interested in, then in addition to Kaplan's assumption, it should be assumed that the desirability of acquitting an innocent is the same as the desirability of convicting a guilty person, and that \( P(G) = P(G \mid C) \) and \( P(I \mid C) = P(I) \), then the result follows.

Since the deliberator will convict iff \( D(C) > D(A) \), it follows that he will convict iff

(8) \[ [D(C \& G) P(G \mid C) + D(C \& I) P(I \mid C)] > [D(A \& G) P(G \mid A) + D(A \& I) P(I \mid A)] \].

Now assuming that \( D(A \& G) = D(C \& I) \) and \( D(C \& G) = D(A \& I) \) the equation reduces to:

(9) \[ D(C \& G) P(G \mid C) + D(A \& G) P(I \mid C) > D(C \& G) P(I \mid A) + D(A \& G) P(G \mid A) \]

It follows that if \( P(G \mid C) \) and \( P(G) \) are probabilistically independent, that \( P(G \mid C) = P(G) \), then \( P(G) > 1/2 \).

Rewrite (9) to reflect the probabilistic independence.

(10) \[ D(C \& G) P(G) + D(A \& G) P(I) > D(C \& G) P(I) + D(A \& G) P(G) \]
Since $P(I) = 1 - P(G)$

(11) $D(C & G) P(G) + D(A & G) [1 - P(G)] > D(C & G) [1 - P(G)] + D(A & G) P(G)$

which finally reduces to:

(12) $P(G) > 1/2$

So if the state of affairs, being guilty or not guilty is considered probabilistically independent from the basic action of convicting or acquitting the $P(G) > 1/2$ rule follows.

Obviously if the errors are not considered equal the probability deviates $1/2$. In order to derive a higher standard of proof it is assumed that convicting innocent persons is of much greater harm than acquitting guilty persons. In general the standard of proof on utilitarian assumptions is:

(13) $P(G) > \frac{D(A & I) - D(C & I)}{D(C & G) - D(C & I) - D(A & G) + D(A & I)}$

Given this general equation one can make the suitable assumption that the desirability of convicting an innocent person is far worse than the desirability of acquitting a guilty person in order to derive a higher standard of proof. Suppose that in a given situation the judge decides the following values apply:

(14) $D(C & G) = 10 D(A & I) = 10$

(15) $D(C & I) = 10 D(A & G) = -100$
It follows that in this situation the standard of proof is $P(G) > .83$. In these special cases the utilitarian rationale appears to give the correct standards of proof.

6.2. The Standard of Proof in the Case of Multiple Defendants.

An assumption implicit in the above argument is that the number of defendants is one. In the case where the accused number more than one, the above utilitarian argument leads to a standard of proof that says a proposition is proved if it has the highest probability among a number of alternative possibilities. (See Kaye (1982) for a similar argument.) But the consequence of this argument is that in wide variety of situations a person's right not to be found guilty if innocent is seriously violated.

Consider the case of *Sindell v. Abott Laboratories*. In this case the plaintiff brought a class action suit against 11 drug companies that manufactured and marketed DES. It was claimed that these companies knew or should have known that ingestion of DES would cause carcinomas in the daughters of mothers who took DES. In this case, then there are not two possible errors as represented in Kaplan's argument. The proof proceeds much like above but in a more general way. There are $2^n$ possible errors which are:

1. $(C_i \& I_i)$ finding against $i$ and $i$ is innocent for all $i = 1 - n$
2. $(A_i \& G_i)$ finding for $i$ and $i$ is guilty for all $i = 1 - n$

In this general case the desirability of an action $A$ is:

$D(A) = \sum P(S_i \mid A) D(S_i \& A)$ for all $i = 1 - n$.

If the assumption of independence, that $P(S_i \mid A) = P(S_i)$ for all $i = 1 - n$ and the assumption that convicting any defendant for certain is equally desirable, that $D(C_i \& I_i) = D(C_j \& I_j)$ for all $i = 1 - n$ are made, then a generalization of the civil standard can be made. The defendant will be found guilty iff the desirability of convicting a given defendant is greater than the desirability of convicting any other defendant.

$D(C_1) > D(C_i)$ for all $i = 2 - n$

This proposition holds only if:
(5) $D(C_1 \& G_1) P(G_1) + D(C_1 \& I_1) P(I_1) >$
$D(C_i \& G_i) P(G_i) + D(C_i \& I_i) P(I_i)$ for all $i=2-n$.

Since $D(C_i \& G_i) = D(C_j \& G_j)$ and $D(C_i \& I_i)$ and $D(C_j \& I_j)$ the following holds:

(6) $D(C_i \& G_i) P(G_1) + D(C_i \& I_i) P(I_1) >$
$D(C_i \& G_i) P(G_i) + D(C_i \& I_i) P(I_i)$ for all $i=2-n$.

Since $P(G_i) = 1-P(I_i)$ the following is obtained:

(7) $D(C_i \& G_i) P(G_1) + D(C_i \& I_i) [1-P(G_1)] >$
$D(C_i \& G_i) P(G_i) + D(C_i \& I_i) [1-P(G_i)]$ for all $i=2-n$.

Which reduces to:

(8) $P(G_1) > P(G_i)$ for all $i=2-n$.

In other words, the adjudicator convicts the defendant with the greatest probability of guilt. This property leads to the crux of the difficulty with utilitarian standards of proof. The solution to maximizing average utility allows for a defendant to be convicted when it is more probable than not that he did not commit the alleged act. In the case of one defendant $P(G_1) > P(G_2)$. Hence $P(G_1) > 1/2$. So the utilitarian rule gains some plausibility in the case of one defendant.

6.3. Rights Based Standards of Proof.

Each individual has a right not to be convicted if innocent. Thus, it is thought that each individual has a right to a certain level of accuracy in legal proceedings. One might ask if this right implies the right to the most accurate procedures that exist regardless of the cost of such procedures. (Dworkin, 1985) It might even be thought that high standards such as proof beyond a reasonable doubt is not required. The derivation of standards of proof is a definite problem for a rights based analysis of standards of proof. But the right not to be found guilty if innocent certainly seems to require, at least, that it be more probable than not that a particular defendant did the alleged act. To find a person guilty on any lesser probability means that it is
more probable than not that he did not do the alleged act. On the rights based view this finding would indicate a profound lack of respect for individuals.

In this chapter I raise the objection that the utilitarian rational for standards of proof has the effect of infringing upon a person's rights not to be convicted if innocent. The type of argument that pits justice against maximization of utility is quite common. I will continue in that tradition by giving another example in terms of the adoption of utilitarianism in regards to standards of proof. As was shown in the last section maximizing desirability leads to the result that a defendant can be convicted even though the probability that he did the alleged action is very low. Let's examine how this rule works in practice and give an obvious case of injustice due to this rule.

In the *Sindell v. Abott* case 11 drug manufacturers were sued in a class action suit for damages relating to the use of DES. Because the drug was marketed in an indistinguishable form by manufacturers, there was no evidence beyond market share evidence as to which manufacturer cause the damage to each plaintiff. The California Supreme Court held that market share was an effective measure of the likelihood that the defendants supplied the product which injured the plaintiff. But as a dissenter to the judgement pointed out "... a particular defendant may be held proportionately liable even though mathematically it is much more likely than not that it played no role whatever in causing plaintiffs' injuries." This is the crux of the criticism of utilitarian standards of proof.

So a rule which says to maximize the desirability of a verdict which involves a number of plaintiffs will have the effect of infringing on the rights of to a fair trial. The rule is to convict the defendant with the greatest probability of having committed the alleged act. In the case of 11 defendants whose market share is roughly equal the rule is to convict the defendant whose probability of guilt is slightly higher than $1/11$. In other words the utilitarian standard of proof in civil cases with multiple defendants allows for a finding against one even though it is highly improbable that he committed the alleged act.

6.4 Conclusion.

The standard of proof has been given a probabilistic interpretation as the minimum lower probability that a proposition at issue must attain in order to be regarded as proved. There are two main views for determining what the level of probability should be for a proposition to be
regarded as proved. One strategy is to follow utilitarianism and maximize the average utility. This means that the deliberator's desire function should be treated as a utilitarian social welfare function.

In a civil case, where the errors that can be made are treated as equally bad, a rational deliberator will convict, in the case of two or less defendants, on a balance of the probabilities. In the case of more than two defendants the defendant with the highest probability will be found guilty or liable. But this rule shows that the utilitarian conception implies that a person's rights are violated since, even where it is extremely improbable that a person did something enjoined by law, he can be found guilty or liable.

The second view is to consider a verdict to be an action to protect the rights of the citizens in society. Individuals have a right not to have their property taken, and not to be injured. Counterbalancing implementation of these rights is the right that individuals have a right not to be convicted if innocent. The difficulty with this rational for standards of proof is determining what these standards should be. But it seems clear that if a person is found guilty it must be more probable than not that he is guilty. Otherwise it is more probable than not that he did not commit the alleged act. Such a rule involves a profound disrespect for individuals. Yet, as I argued, the utilitarian view allows people to be found guilty when it is more probable than not that they did not commit the alleged act.
CONCLUSION

This thesis examined deliberation in legal proceedings. The law has a number of principles for application in various types of proceedings such as criminal, civil and administrative proceedings. I argued that, although everyone would agree to these principles, underneath the superficial agreement there is a profound disagreement about how these principles are to be interpreted.

A traditional principle is that in order to find one liable it must be proved that the accused committed the alleged act. I argued that when a proposition at issue is proved that it be regarded as true on the basis of a sufficient probability. Since the proposition is regarded as true, propositional logic should also be regarded to apply to it. It follows that the theorems of logic hold for proved propositions. On the alternative Bayesian interpretation proved propositions are not regarded as true. The propositions which are regarded as proved do not conform to principles of deductive logic.

All views hold that, in some sense, a proved proposition be sufficiently probable. I argued that the epistemic view is particularly suited to legal reasoning. On this view probability is conceived as a mind independent logical relation between evidence admitted and the conclusion reached on the basis of that evidence. The opposing subjective view of probability, where probability is conceived as a degree of belief is completely arbitrary. Chance is the disposition of an object or person to behave in a way in a given circumstance. Unlike relative frequency views chance is a property of a single individual or a single object in its environment and not a property of a set of individuals or objects. This means that an individual's right to be treated as an individual is upheld in legal proceedings.

In order to be convicted a person must be proven to have committed a given act. I argued in favour of causal theories of action. Causal theories view an action as a part of a sequence of causally related events. It is that part of the sequence with the properties that are represented to the agent in his causally efficacious mental state. The causal view is able to distinguish a number of elements in the ontology of action that legal theorists have had trouble with. The causal view distinguishes between events, actions, omission of actions, causing something to happen, being in a state, the actions of others, ascribing responsibility and having something happen to one.
All theorists can regard deliberation as dynamic principles for modifying a rational agent's mental states in response to new information. There are two plausible models of these dynamics: Bayesian and non-Bayesian. On Bayesian theories all changes of belief are by Bayes' theorem or generalizations thereof. On a non-Bayesian view beliefs are changed by accepting new beliefs, conjoining them with the old beliefs, and modifying the old beliefs on the basis of the new ones. As an interpretation of legal deliberation the Bayesian view has a number of disadvantages. The Bayesian theory does not allow for the conjunction principle, does not regard proved propositions as true, does not allow for all relevant evidence to be admitted, and does not allow for plausible presumptive inferences.

The principle that a person is proved to have committed an act if it is sufficiently probable that he committed such an act gave rise to a difficulty. The problem amounted to how one can have a theory of deliberation which meets three principles of legal reasoning: the deliberating agent's beliefs are consistent, the agent believes a proposition A if the probability of A is sufficiently high, and if the agent believes A and believes B then he believes A and B. I show how this problem is resolved by requiring probability to be resilient. The most natural interpretation of resilient probabilities are chances.

A person is proved to have committed an act if the probability of having committed that act reaches an appropriate standard of proof. I considered the utilitarian standard as an interpretation of "standard". In this sense the standards of proof follow from a more general standard of the deliberator to maximize the average utility level in society. There are two errors that a deliberator can make: he can convict an innocent person or acquit a guilty person. If these two errors are equally bad, then utilitarian implies in the case of one or two defendants the more probable than not standard. In the case of more than two defendants and the assumption that the two errors are equally bad, that the defendant with the highest probability of being guilty be convicted. However, I argued that a utilitarian rational for standards of proof violates a person's right not to be convicted if innocent. This is due to the fact that a person can be convicted by a utilitarian deliberator even though it is more probable than not that he did not commit the alleged offence.


Philosophy of Science. 54: 76-91.


Harvard Law Review. 103: 530-54.


**CASES**


Commonwealth v. Crow, 1931. 303 Pa. 91


Gardner v. Gardner, 1877., 2 App. Cas. 723 H.L.

In Re Winship, 1970. 397 U.S. 35


New York Life Insurance v. McNeely, 1938., Pac. 2d. 948

Partington v. Williams, 1975. 62 Cr. App Rep 220, DC


Riggs V. Palmer, 1889. 115 N.Y. 566; 22 N.E. 188 N.Y.C.A.

R v. Shrimpton, 1851. 5 Cox. C.C. 387, 3 Car & Kir. 373 C.C.R.


R. v. Lewis (1903), 60. L.R. 132, 7 C.C.C. 261 (C.A.)


Sargent v. Massachusetts Accident Co., 1940. 307 Mass. 246, 250, 29 N.E.22d 825, 827
Sindel v. Abott Laboratories, 1980. 26 Cal. 3d 588, 607 P.2d 924, 163 Cal Rptr. 132, cert. denied, 449 912
U.S. v. Guiteau, 1882., Mackey 498