A Mathematical Comment on the Fundamental Difference between Legal Theory Formation and Scientific Theory Formation

Ronald P. Loui

Department of Computer Science and Engineering and
Program on Legal Studies
Washington University in St. Louis
St. Louis, Missouri, USA 63130
loui@cs.wustl.edu

Abstract. This paper attempts to provide a mathematical formulation of the legal theory formation problem and a comparison to the scientific theory formation problem. This is intended to be both an original contribution to AI and Law, and a presentation of the problem to the philosophy of science community. In conclusion, some remarks are made about the shallowness of today's models in machine learning and knowledge discovery compared to the legitimate models in AI and law and philosophy of science.

This paper originates from a discussion with Henry Kyburg and John Pollock beside a tree at Tresider Union over a decade ago. I must also thank Thorne McCarty, Jeff Norman, Mark Mittleman, Ana Maguitman, Carlos Chesñevar, Fernando Tohmé, and Guillermo Simari's research group in Bahía Blanca, where we first discussed elision. This paper is dedicated to my intellectual paternity, from my teacher Henry Kyburg and his teacher, Ernest Nagel, who were concerned with the logic of science, to Nagel's advisor, Morris Cohen, who wrote on logic and law.

1 Introduction

Legal theory formation is no more restricted to law than scientific theory formation is restricted to science.

Scientific theory formation is a general kind of induction, which, as described and formalized by philosophers of science in the second half of the 20th Century, refers to a pattern of reasoning that may occur in everyday statistical reasoning, empirical observation, diagnostic reasoning, and ontological or taxonomical reasoning, that is not restricted to scientific laboratories or grandly unifying moments. Scientific theory formation refers to a deep pattern of appraising generalizations in the face of empirical observation for the purposes of adopting the right generalization or generalizations to summarize the data.

Induction (simpliciter) is ambiguous between mathematical induction, statistical reasoning about high probabilities, and theory-formation, the latter implying considerations beyond probability, such as simplicity. The considerations may include graded epistemological commitment to a generalization, the possibility of background or entrenched theory which biases induction, the role of probability in confirmation, the tradeoff between error and power, or between predictiveness and power, and the adherence to conventions such as causal explanation. It may go so far as to describe paradigm shifts, and be foundational or coherentist in its bias.

Today's machine learning (ML), computational learning theory (CLT), datamining, and knowlege discovery (KDD) fields of Artificial Intelligence (AI) and computer science, have developed largely in ignorance of the philosophical work on scientific theory formation. Nevertheless, many of the main issues of scientific theory formation have been rediscovered in ML and KDD.

Legal theory formation is also a general kind of induction. Like scientific theory formation, it is thought to be an occasional mode of legal reasoning. Just as normal science is supposed to be subsumption under generalizations, normal law is supposed to be subsumption under generalizations, which may require defeasible rather than deductive reasoning after general law has been applied to particular situation. In both cases, theory formation was supposed to occur less often, though ironically, both scientific theory formation and legal theory formation are now taken to be constantly operating in the rational scientific and legal minds, respectively.

Logical texts prior to the rise of deductive logic in the early 20th Century almost always included chapters on legal, ethical, or analogical reasoning. Isolated legal scholars, and whole schools of logical thought, have attempted to reduce legal reasoning to the logic of the day, most notably deontic modal logic. The formalization of legal reasoning is today in the most fertile stage in its history, with the activity in AI and Law and the restoration of the defeasible conditional (a pearl that managed to appear among the sea floor deposits of decaying non-monotonic logics).

Legal theory formation is as likely to occur in day-to-day reasoning during the management of conflicting heuristics, the interpretation of discourse, the interpretation of open-textured predicates that appear in explicit, linguistically-specified rules, the interplay of principles and rules, and the formulation of coherent bureaucratic standards. Scientific theory formation is as likely to occur in decision-theoretic reasoning about an agent's preference, diagnostic and causal evidential reasoning, and mathematical modeling of engineered systems (like queues) that do not have a privileged nomological status (like objects on frictionless surfaces). Legal theory formation is almost certainly not what lawyers normally do, and may only be a rough parody of what judges, legislators, and solicitors do. Scientific theory formation is almost certainly *not* what scientists do. Nevertheless, these two patterns of reasoning cover the majority of inductive rational behavior.

In what follows, we aim to say as mathematically precisely as possible what these theory-formation problems are.

2 Scientific Theory Formation: Model I

A set of cases is a set, C, of circumstances paired with observations,

$$C = \{< c_i, o_i > \mid i = 1 \dots n\}.$$

Each circumstance, c_i, is a vector in $\Re^k$. Either all n observations, o_i are real-valued, in $\Re$, or all n are boolean, in $\{0, 1\}$. The possibility of boolean-valued circumstances is omitted here, for our convenience.

A theory is a function that is fit to the cases. A function, f, from a family, $\mathcal{F}$, is defined by parameters $p \in \Re^m$. Parameters are chosen to optimize the fit to the cases. f is a mapping $\Re^k \rightarrow \{0, 1\}$ or $\Re^k \rightarrow \Re$, depending on the values of o_i.

The error of f's fit with C is defined as a function of p, $E(f_p(C))$, or $E(p)$ for short:

$$E(p) = \Sigma_{i=1\dots n}(f_p(c_i) \neq o_i)$$

is just the number of disagreements between the value of the function and the empirically observed value. This makes most sense for boolean observations. For real-valued observations, a norm is often used, such as squared error:

$$E(p) = \Sigma_{i=1\dots n}\|f_p(c_i) - o_i\| \ .$$

The scientific theory formation problem is to find $p^* \in \Phi$ to minimize the error:

$$p^* = argmin_{p \in \Phi} E(p) \ .$$

For example, for $k = 1$, linear regression is a basic kind of curve-fitting, on $p =< m, b >$, and this is a kind of theory-formation. A classic example here may be the ideal fit of $PV = nRT$ to measurements on gases. A more practical example is the fit of a mean-risk decision rule to the observed decision behavior of an agent, or perhaps a linear multiattribute utility function. Another example is the training of feed-forward neural networks, where p is the weight matrix, C is the training set, k is the number of nodes on the input layer, $c_i|_j$ is the j-th input node's value on the i-th training datum, and o_i is the output value for a single output node for the i-th training datum.

The ML/KDD problem formulations of principal component analysis (PCA), support vector machines (SVM), and the CLT probabilistically approximately correct (PAC) derive from this model.

3 Scientific Theory Formation: Model II

A simplicity ordering on functional forms, $\prec$, permits one function, f to be compared to another, g on grounds of simplicity. For f_{p_1}, where p_1 has m dimensions or *degrees of freedom*, any g_{p_2} where p_2 has $m-1$ dimensions will be simpler, $g \prec f$. We might also require that $\mathcal{F}$ be related to $\mathcal{G}$, such as the relation between parabolas and lines, where any member of $\mathcal{G}$ can be expressed as a member of $\mathcal{F}$ by fixing one parameter.

A background theory, B is a set of functions which presumably have been fit to other sets of cases, for other circumstance/observation pairs, in other dimensions. The coherence of B and f_p may be defined as $B \cdot f_p$, which is not often expressed mathematically, but which may contain penalties, for example, whenever parameters fall outside anticipated ranges. More often, B contains axioms which permit k' inferences, so that each $c_i \in \Re^k$ becomes $c'_i \in \Re^{k+k'}$. For example, dimension 1 may be the measurement of an object's brightness, and B may contain a rule for deriving the difference between a reference object's brightness and the observed brightness, that difference constituting the implied circumstance in dimension k+1. If it is supposed in the background theory that the observed brightness must be greater than the reference object's brightness, a reasonable auxiliary hypothesis, then upon fitting a function to the derived $C' = \{< c'_i, o_i >\}$, one might add to the error if there is a positive difference in o_{k+1}, Yet another way of formulating coherence is to apply a penalty to f_p if it is inconsistent with B. In the notation of belief revision, $B \oplus f_p$ is the minimal contraction of B so that the addition of f_p is consistent. $B \ominus f_p$, the minimal set of statements to remove from B, can be measured.

In any of these cases, the measurement of fit and the simplicity/coherence measures must somehow be combined. For example, is the simplicity ordering respected unless there is no tolerable fit for any p within a family $\mathcal{F}$? Or is there a combination of measures (e.g., linear combination),

$$\alpha E(f_p(C)) - (1-\alpha) B \cdot f_p$$

which is minimized simultaneously in the search for p^*?

I know of only one author who attempts to answer this question formally. Kyburg's theory is that the augmentation B with f_p does not require a contraction to make the union consistent. Instead, $B' = B \cup \{f_p\}$ defines an error rate for each observational predicate. The cost of simplicity is a poor fit to empirical observation. The cost of a poor fit is an increased error rate for observation and prediction. The cost of these errors is the reduced epistemological commitment (level of acceptance) that can be attributed to any observation or any prediction. In our model, there is a single dimension which is predicted using any f_p that is fit to any C. The "top-level" error rate thus applies to any use of f_{p^*} to predict the observable given a new circumstance. The error also propagates to any of the dimensions $1 \ldots k$ in the range of f. These dimensions may be used in the observations of other cases to which other functions are fit. The combined rate at which a dimension participates in erroneous and error-free function fitting will define the error rate for the dimension. Functions with many parameters will fit more data without introducing much error. The cost of fitting many-parameter functions is that they require more observations to determine parameters p, and thus permit predictions in fewer observational situations.

For example, we might say that all 3-2 pitches are fastballs, but this implies a high rate of error for correctly observing curve balls, since $B \cup \{f_p\}$ contains both the raw observation that the last pitch was a curve ball and the prediction, post-adoption of the generalization, that it must actually have been a fastball, since it was a 3-2 pitch. That is, the web of belief can accomodate the generalization, but only at a reduced self-confidence in the ability to distinguish fast balls from curve balls. Perhaps at a level of acceptance associated with probability .9, observations of curve balls and fastballs on 3-2 pitches can no longer be considered veridical. Alternatively, we might say that all 3-2 pitches are fastballs unless the previous curveball was a strike (this is actually a reasonable baseball batter's hypothesis). This theory fits with less error, but cannot be used for prediction unless, on the previous pitch, the observations of curveball/not-curveball and strike/not-strike were made. This example obviously uses boolean dimensions for c_i instead of the real-valued dimensions of our model.

Let $\delta \in \Re^k$ be the error rates for each observational dimension, $1 \dots k$. Cn is a consequence-closure function on each $o_i \in Re^k$, $Cn(o_i) = o_i \pm \delta$ (To Kyburg, this inference happens because the probability of a true value in the j-th dimension, in the range $o_i|_j \pm \delta|_j$ given that it is observed to be $o_i|_j$ is greater than the acceptance level, $1 - \alpha$.) The problem is to find p^* to maximize the inferrable content of $B \cup \{f_{p^*}\} \cup Cn(\Omega)$, where Ω is the union of all observations. This inference is probabilistic, and is essentially the same as Cn. Thus, find:

$$p^* = argmax_{p \in \Phi} Cn(B \cup \{f_{p^*}\} \cup Cn(\Omega)) \ .$$

4 Legal Theory Formation: Model I

The principal difference in legal theory formation is that a function fit to a set of cases should disagree nowhere with the cases and their outcomes. It is possible for legislators and judges to ignore a precedent, regarding it as an erroneous or otherwise non-binding decision. But most often, in non-appellate situations, an explicit distinction must be made to avoid a putative disagreement with precedent. The opinion of the case must explain why it is *not* governed by a rule that appears to counsel an opposite result. Anyway, we can assume that a set of cases, C, is authoritative with respect to the problem of legal theory formation; deviation from a perfect fit is not possible.

Again, a set of cases is a set, C, of circumstances paired with decisions, $C = \{< c_i, d_i > \ | \ i = 1 \dots n\}$. Each circumstance, c_i, is a vector in three-valued logic, $\{0, 1, ?\}^k$. The third-value is used when a dimension is not reported or is considered irrelevant. The possibility of real-valued circumstances and decisions is omitted here, for our convenience. d_i are boolean, in $\{0, 1\}$.

The error of f's fit with C is defined as a function of p, $E(f_p(C))$, or $E(p)$ for short:

$E(p) = \Sigma_{i=1...n}(f_p(c_i) \neq d_i)$

as above, the number of disagreements of f with the decisions.

The problem at first appears to be to minimize the simplicity of $\mathcal{F}$ subject to the constraint that there exists a p which zeroes the error function.

$min_{\prec}(\mathcal{F})$ s.t. $\exists p \in \Phi : E(f_p(C)) = 0$.

5 Legal Theory Formation: Model II

The model above would not even be recognizable to Joseph Raz, who is the first to this author's knowledge to formulate the problem as rule-induction in a propositional logical form (von Wright had a more mature framework in deontic logic, but perhaps did not actually formulate a theory-formation problem).

It is important in law to extract a set of rules from a set of cases, so one may think of a *set of rules*, R, *of the cases*, C. Furthermore, the rules may be defeasible, so that a rule "$a \wedge b \longmapsto c$" defeats a rule "$a \longmapsto \neg c$".

A case may be rewritten in terms of the dimensions that receive a 0 or 1 value; each such supporting dimension is known as a (propositional) factor; a 1 value is an assertion of the factor; a 0 value is its negation. A set of defeasible rules that fits the cases with no error is therefore the set which takes the antecedent of each defeasible rule to be the factors of the case. Let $\phi(c_i)$ be the factors of the case, c_i.

$\bigwedge_{\phi(c_i)} \longmapsto d_i$

If there are directly conflicting cases, two cases with the same factors and opposite decisions, then there is no error-free fit of rules to cases.

Unfortunately, the set of rules of interest is usually less numerous than the set just proposed, and one may also want to induce rules that are not in the set. Kevin Ashley annotates each factor as inherently pro-d_i or contra-d_i. ϕ^+ and ϕ^- are the pro- and contra- factors, respectively. The set of rules that can be extracted from a set of cases is now extended to include any rule:

$\bigwedge_{\phi^+(c_i)} \wedge \bigwedge_D \longmapsto d_i$

for any $D \subset \phi^-(c_i)$. I.e., any rule can be extracted that can be derived from another extracted rule by dropping any number of contra-d_i factors. This separability and monotonicity of factors presumes that two positive factors cannot in combinaction act as a negative factor.

Jeff Norman and I describe, in a series of papers, how the case should be represented as a set of arguments, not just a list of factors.

The circumstances of a case $c \in C$ is a set of arguments, $\{a_1, \ldots, a_n\}$ each argument a is a propositionally-labeled tree, rooted in a decision, $\theta(a)$, with each leaf's proposition, $\nu(a^j)$ recognized as evidence (uncontested assertion) in the case, and each internal node arguable using the defeasible rules of the argument, $\rho(a)$. The set, $\nu(c) = \bigcup_{a_i \in c} \nu(a_i)$, collects all the evidence used in all the arguments of the case, and should correspond to the factors of the case, $\phi(c)$.

For example, the argument tree

$a =< d, << f_1, < f_2, f_3 >>, f_4 >>$

is a tree rooted with the decision d, with evidence $\nu(a) = \{f_2, f_3, f_4\}$ and the two rules $\{f_2 \wedge f_3 \longmapsto f_1, f_1 \wedge f_4 \longmapsto d\}$. $\theta(a) = d$. $\phi(c) = \nu(c) = \bigcup_{a \in c} \nu(a)$.

A rule grounded in a case, c is any

$$\bigwedge_D \longmapsto d$$

such that $D \subset \phi(c)$. But not all grounded rules are permitted generalizations. Some such rules have excessive specificity. In defeasible reasoning with implicit defeaters, the specificity of the antecedent of a rule determines what rules it can defeat; not only is a narrow generalization less broad than should be extracted, but it is also implicitly stronger than conflicting rules that ought to be its peer. Meanwhile, some grounded rules might have insufficient specificity, which is equivalent to too broad a generalization. Overly broad generalization is less of a problem for defeasible rules than for a deductive rules (because more particular exceptional circumstances can be governed by more specific contrary rules) but such a rule may still be unwarranted.

We define the set of rules which are proper elisions of the case, $< c, d >$, $R^*(c)$, is defined inductively as the smallest set that contains as subsets:

1. $\{\bigwedge \nu(c) \longmapsto d \}$.

2. $\{\bigwedge(A - B) \longmapsto d \mid$
   $\bigwedge A \longmapsto d \in R^*(c),$
   $\exists a \in c, B' \subset \nu(a):$
   $\theta(a) = \neg d,$
   $B = B' \cup \bigcup_{i \in \Delta_c(a)} \nu(i),$
   $B \cap \bigcup_{i \in (c-a)} \nu(i) = \{\}$
   $B' \neq \{\},$
   $\}$.

where $\Delta_c(a)$ is the set of arguments in c that are dialectically relevant to a, i.e., any counterarguments, their counterarguments, and so forth. I.e., $\Delta(a) = \mu\{a' \mid a' \rhd a, \text{ or } \exists a'' \in \Delta(a) : a' \rhd a''\}$, where μ denotes the smallest (inductively defined) set satisfying the equation, and $\rhd$ is the counterargument relation.

Case 1, the base case, the maximally specific rule of the case, says that one may collect all the evidence from all the arguments for the decision of the case and conjoin their evidence as the antecedent any argument arguing for the decision of the case.

Case 2, the inductive case, says that an extracted rule can be weakened by removing any factor from the antecedent that participates as evidence only in an argument contrary to the decision. This is essentially the consideration of Ashley, taken into the more detailed representation of cases that permits a more nuanced look at factors (a factor may participate in an argument pro and in an argument contra at the same time). However, the set B of factors that must be removed includes, in addition to the evidence of the contrary argument, all evidence participating in its counterarguments (and so on). This is because we can no longer guarantee that the factor of a reinstating argument is relevant, once the counterargument is no longer possible. (This expression actually makes some simplifying assumptions, but is appropriate for this level of exposition.)

For example, the case might decide d=*banning political signs on private houses abridges free speech* (e.g., *Ladue v. Gilleo*) on the grounds that it is $\neg v_1$=*not content neutral*, and $\neg f_2$=*does not provide alternate means of communication*, the evidence for which is v_3=*ban does not provide alternate times or manners*. A counterargument that is considered in the case, was that f_2=*does provide alternate means of communication* because v_4=*home owners may place political signs elsewhere* and f_5=*a sign has the same impact wherever it is placed*, with evidence v_6=*the content of a message is unchanged when it is moved*. A counterargument to the counterargument holds that $\neg f_5$=*residential signs are a uniquely important and distinct medium of expression* because v_7=*location of the sign provides information about the identity of the "speaker"*.

The rules used in arguing the case, $\bigcup_{a \in c} \rho(a) = \rho(c)$ are:

$$\neg v_1 \wedge \neg f_2 \longmapsto d$$
$$v_3 \longmapsto \neg f_2$$
$$v_4 \wedge f_5 \longmapsto f_2$$
$$v_6 \longmapsto f_5$$
$$v_7 \longmapsto \neg f_5 \ .$$

The proper grounded elisions are just:

$$\bigwedge\{\neg v_1, v_3, v_4, v_6, v_7\} \longmapsto d$$
$$\bigwedge\{\neg v_1, v_3, v_4\} \longmapsto d$$
$$\bigwedge\{\neg v_1, v_3, v_6\} \longmapsto d$$
$$\bigwedge\{\neg v_1, v_3\} \longmapsto d \ .$$

Two separability properties in the calculus of arguments permit the extension to proper grounded elisions: first, if an argument has two counterarguments, at least one must suffice as an undercutter if the argument be undercut; second, similarly, if a proposition can be argued in multiple ways, at least one argument must warrant its conclusion if the proposition be warranted. A different calculus of argument that weighs arguments by their number, or that combines multiple logically unrelated weak arguments, would not permit such extensions.

With a better notation for operations on trees, the set of proper elisions can include non-evidential extensions and elisions of defeasible rule antecedents, and defeasible rule consequents that are not the decision of the case, but are intermediary decisions that are mandatory for any who would accept the reasoning of the case.

The legal theory formation problem is now the extraction of a set of rules, $R \subset R^*(c)$ such that for each case $< c_i, d_i > \in C$, d_i is defeasibly warranted under R, i.e.,

$$\forall < c, d > \in C : \nu(c) \mapsto_R d \ ,$$

i.e., some R that is a sufficient subset of grounded proper elisions of the cases.

Roth-Verheij and Prakken-Sartor, among others, give even more nuanced analysis of reasoning with cases. Our purpose here is to stay as close to scientific theory formation as possible.

6 Observations Toward Legal Theory Formation: Model III

The rules that underlie the arguments used in the cases must belong to a background theory, B, where $\bigcup_{c \in C} \bigcup_{a \in c} \rho(a) \subset B$. These are the rules so far recognized; the judgements of the cases permit adding to these rules. As in scientific theory formation, the background theory may influence which generalizations are permissible, and considerations of simplicity and coherence may cause revision of B in light of generalization from C.

We have seen that the difference between scientific and legal theory formation begins with a different attitude toward error, even before the logical difference (defeasible versus deductive) rules must be encountered. Not surprisingly, the improvement of the scientific reasoning model (e.g., Kyburg) pays close attention to error and predictiveness, where probability is the main tool, while the improvement of the legal reasoning model pays close attention to improper over-generalization and improper under-generalization, where defeasible argument is the controlling idea. In law, the strict attitude toward error is made possible by the defeasible rules. In science, the quantification of error (and predictiveness) is made possible by probability.

In science, the generality of the rules is treated as a consideration of simplicity. Simplicity on its face is not even a primary consideration in legal theory formation, as the emphasis is on determining the entire set of permissible generalizations supported by a set of cases. Were the set of rules to be published explicitly, as a revised code, it would make sense to prune "unneeded rules." These can be defined trivially as rules that are subsumed by more general rules, where the general rule is not likely to be countered in a way that makes relevant the more specific rule (a theory of likely relevant countering can be provided). It is probably correct to extract just those grounded proper elisions that correspond to complete removals of arguments, rather than removals of parts of arguments. However, the law returns to reasoning upon cases precisely because unanticipated, unlikely interactions occur.

In the defeasible reasoning community, it has long been suggested that the more the defeasibility, the less simple or less cogent the overall set of rules. This is only partly true. A sequence of rules that can, in the appropriate circumstances, defeat each other is indeed suspect, e.g.:

$a \longmapsto x$, but
$a \wedge b \longmapsto \neg x$, but
$a \wedge \wedge c \longmapsto x$, but
$a \wedge \wedge c \wedge d \longmapsto \neg x$, but
$a \wedge \wedge c \wedge d \wedge e \longmapsto x$, etc.

If such a cascade of exceptions to exceptions were unmotivated and could be replaced with a less hierarchical set of rules, where the same set of cases could be fit by making distinctions on other factors, it would be desirable to improve cogency by making that replacement. However, we find in the law many examples of motivated cascades of exceptions, precisely because there are principles that come into conflict. It is not always desirable to simplify when simplification deletes the legitimate expression of a right, privilege, or obligation.

A better next step in the improvement of models of legal theory formation would consider the relation between principles and generalizations. One strategy for simplifying rules is to introduce new dimensions for cases. A set of cases may seem to have no clear general pattern of decision until a new concept such as “contributory liability” or “knowable interchangeability” permits reorganization of the cases. Often, however, these dimensions are artificial, and are chosen for their practicability rather than the social values they promulgate, for example, “third strike.” I propose that a legal theory is more cogent than another if its rules are aligned with broader social principles. “No vehicles in the park” is a good rule to add to a legal theory if it is the particularization or instantiation of more general principles regarding fair sharing of public resources, not just because various special interests each have separate grounds for restricting vehicles from parks, all of which can be collected under the umbrella of the rule. A rule that is a special case of a general principle adds to the validity of the theory even if it creates organizational havoc. Properly tailoring rights so that they are neither too broad nor too narrow may just require a lot of rules arrayed in hierarchies of potential exceptions.

For scientific theory as well as legal theory, much needs to be done mathematically in the modeling of coherence (or cogency).

This paper has attempted a side-by-side comparison of two of the most important forms of induction. Scientific theory formation gives primacy to the rules, and uses cases to identify optimal rules. Rules would be simple, few, and aesthetically pleasing if not for the empirical evidence, which causes asymmetries, non-linearities, and hidden variables. God knows why the rules are the way they are (or for that matter, why deductive logic is preferred). Rules are real, and optimal fit is a heuristic for uncovering reality. Legal theory formation takes cases (actual or hypothetical) to be deontically primary, and induces rules only as a practical compilation of edict. It does so for the purposes of prescriptive clarity, structural discovery, or linguistic brevity. Rules are as numerous as they need to be, in order to depict the complexities that justifiably differentiate circumstances. The moral and political intuitions contained in the cases almost always require a flexible logic, such as defeasible reasoning, and a detailed representation of *why* a case was decided the way it was.

7 On Machine Learning and The Proper Modeling of Phenomena

Modern machine learning attempts to erase all distinctions between various kinds of induction from data, as if all learning from cases can be expressed as the selection of features followed by the determination of an appropriate set of separating hyperplanes in that feature space. In this recent research programme, the intellectual question is no more than what is the preferred mathematical transformation of feature space that permits convenient re-representation of data, e.g., what is your favorite kind of nonlinear support vector kernel?

Nothing could be more historically or intellectually insulting. There are different purposes for different kinds of induction and in a few cases, notably, the cases of scientific induction and legal induction, the modeling of the nuances has led to great ideas in the history of logic and reasoning. This reductivist travesty of induction is not actually the fault of the mathematical or computational emphasis of the machine learning field. Computational and mathematical modeling are valuable exercises, especially when they enable automation or improve the clarity of discussion. The intellectual vacancy we see in machine learning is the direct result of an unwillingness to study the phenomena first, before doing the mathematical modeling.

In this paper, the objective has been to develop and compare mathematical formulations worthy of law and science, without discarding the essence of either.

8 Bibliography

C. E. Alchourrón and E. Bulygin. Normative Systems, Springer-Verlag, 1971.

Trevor Bench-Capon and Giovanni Sartor. Theory Based Explanation of Case Law Domains, Proceedings of the 8th International Conference on Artificial intelligence and Law, St. Louis, Missouri, pp. 12 - 21, 2001.

T. J. M. Bench-Capon and G. Sartor. A Model of Legal Reasoning with Cases Incorporating Theories and Values, Artificial Intelligence, 150:1-2, pp. 97-14, 2003.

Morris Raphael Cohen and Ernest Nagel. An Introduction to Logic and Scientific Method, London, Routledge and Kegan Paul, 1936.

J. C. Hage. Reasoning with Rules : An Essay on Legal Reasoning and its Underlying Logic, Kluwer Academic Publishers, 1997.

H. L. A. Hart. The Concept of Law, Oxford : Clarendon Press, 1961.

Henry Ely Kyburg. Theory and Measurement, Cambridge University Press, 1984.

Henry Ely Kyburg. Science amd Reason, Oxford University Press, 1990.

Edward Hirsch Levi. An Introduction to Legal Reasoning, University of Chicago Press, 1949.

Isaac Levi. The Enterprise of Knowledge, MIT Press, 1980. R. P. Loui.Logic of Arguments, and Arguments of Cases Proc. XVIII World Congress of the International Association for Philosophy of Law and Social Philosophy (IVR), Buenos Aires, 1997. Also R. Loui and J. Norman. Eliding the arguments of cases, Washington University CS Technical Report, 97-12, 1997.

R. P. Loui. Process and Policy: Resource-Bounded NonDemonstrative Reasoning, Computational Intelligence, 14:1, pp. 1, February 1998.

R. P. Loui. Case-Based Reasoning and Analogy, MIT Encyclopedia of Cognitive Science, R. Wilson and F. Keil, eds., MIT Press, 1998.

R. P. Loui and Jeff Norman. Rationales and Argument Moves, Artificial Intelligence and Law, 3:3, pp. 159-189, September 1995.

L. Thorne McCarty. Some Arguments about Legal Arguments, Proceedings of

the 6th International Conference on Artificial intelligence and Law, Melbourne, Australia, pp. 215 - 224, 1997.
Ernest Nagel. The Structure of Science: Problems in The Logic of Scientific Explanation, Hackett Pub. Co., 1979.
Aleksander Peczenik. On Law and Reason, Kluwer Academic Publishers, 1989.
Henry Prakken and Giovanni Sartor. Reasoning with Precedents in a Dialogue Game, Proceedings of the 6th International Conference on Artificial intelligence and Law, Melbourne, Australia, pp. 1 - 9, 1997.
R. A. Posner. The Problems of Jurisprudence, Cambridge/London, 1990.
Joseph Raz. The Authority of Law, Essays on Law and Morality, Oxford University Press, 1983.
W. Quine and J. Ullian. The Web of Belief, Random House, 1978. Edwina L. Rissland, Kevin D. Ashley. A Note on Dimensions and Factors, Artificial Intelligence and Law 10:1-3, pp. 65-77, 2002.
Edwina L. Rissland, Kevin D. Ashley, R. P. Loui. AI and Law: a Fruitful Synergy, Artificial Intelligence 150:1-2, November 2003, pp. 1 - 15, 2003.
Bram Roth and Bart Verheij. Cases and Dialectical Arguments An Approach to Case-Based Reasoning, Lecture Notes in Computer Science, On the Move to Meaningful Internet Systems 2004: OTM 2004 Workshops: OTM Confederated International Workshops and Posters, GADA, JTRES, MIOS, WORM, WOSE, PhDS, and INTEROP 2004, Agia Napa, Cyprus, October 25-29, 2004. Proceedings, Robert Meersman, Zahir Tari, Angelo Corsaro, eds., p. 634, Springer-Verlag, 2004.
Cass R. Sunstein. Legal Reasoning and Political Conflict, Oxford University Press, 1998.
Cass R. Sunstein. Of Artificial Intelligence and Legal Reasoning, The Law School, University of Chicago, 2001.
G. H. von Wright. Norm and Action. Routledge, 1963.
Richard A. Wasserstrom. The Judicial Decision: Toward a Theory of Legal Justification. Stanford, CA: Stanford University Press, 1961.