Theory Based Explanation of Case Law Domains

Trevor Bench-Capon
Department of Computer Science
The University of Liverpool
Liverpool, UK
+151-794-3697
tbc@csc.liv.ac.uk

Giovanni Sartor
Faculty of Law
University of Bologna, Italy
+39-051-229405
sartor@cirfid.unibo.it

ABSTRACT
In this paper we put forward a formal description of theories which can be used to record understanding of, and explain decisions in, case law domains. We believe that reasoning with cases involves all of theory construction, use and evaluation, and that awareness of the theory which provides a context for case based arguments is essential to understanding such arguments. Moreover, our account of these theories includes a systematic link between factors and values, which we believe is necessary to explain why some arguments prove to be more persuasive than others. We begin by formalising the various elements that the theories contain, and then provide a set of theory constructors which allow theories to built up from the background of decided cases. We show how such theories can be used to explain decisions on particular cases. We discuss how theories can be compared and evaluated. We then show how the argument moves of HYPO and CATO can be understood in terms of our framework. We conclude with a brief discussion of an implementation of the framework, and a summary of the major features of our approach.

1. INTRODUCTION

Elsewhere, ([4], [5], [6], [15]) we have argued for the need to consider the values promoted or defended by case decisions when considering how these decisions should be applied to future cases. We will not argue this point in this paper, but simply take it that a satisfying account of case based reasoning will incorporate such values. We have also argued previously ([5], [6]) that arguments such as those used in case based reasoning cannot be considered apart from a context; the context being a theory of the law of the pertinent domain which explains the desired outcome in the case under consideration. Again we will not argue this point in this paper, but rather ask the reader to accept that this may, at least, be an interesting approach. What we will do here is set out, in a formal fashion, our conception of such theories, and how they can be constructed, used and evaluated. The approach used here is perhaps not the most general and abstract that could be developed,

but is intended to be tailored towards development into an implemented system, while covering the basic elements of the required theory. One possible extension is considered in section 6.

Throughout the paper we will illustrate our discussion with an example taken from [7] which consists of three cases involving the pursuit of wild animals. In all of those cases, the plaintiff (Π) was chasing wild animals, and the defendant (Δ) interrupted the chase, preventing Π from capturing those animals. The issue to be decided is whether Π has a legal remedy (a right to be compensated for the loss of the game) against Δ or not. In the first case, *Pierson v Post*, Π was hunting a fox in the traditional manner using horse and hound when Δ killed and carried off the fox. Π was held to have no right to the fox because he had gained no possession of it. In the second case, *Keeble v Hickeringill*, Π owned a pond and made his living by luring wild ducks there with decoys and shooting them. Out of malice Δ used guns to scare the ducks away from the pond. Here Π won. In a third case, *Young v Hitchens*, both parties were commercial fisherman. While Π was closing his nets, Δ sped into the gap, spread his own net and caught the fish. In this case Δ won.

2. THE ELEMENTS OF THEORIES
The essential building blocks of the theories are decided cases. Cases can be seen initially as a set of facts, and a decision made on the basis of those facts. But this has not typically been found to be the most useful way of representing cases for case based reasoning purposes. Facts are in themselves neutral and not necessarily relevant to the outcome. Explanation of outcomes has usually therefore been in terms of *factors,* introduced in the HYPO system and best documented in [3]. Factors are an abstraction from the facts, in that a given factor may be held to be present in the case on the basis of several different fact situations, and importantly are taken to strengthen the case for one or other of the parties to the dispute. In the above cases one such factor is whether Π has possession of his quarry. This abstracts from the hounds not yet having caught up with the fox, the ducks not yet having been shot and the fish still swimming in the sea rather than landed on the boat, to a single factor. That in none of the cases did Π have possession strengthens Δ's position in each case. We make use of factors, and assume that a prior analysis of the cases has been carried out, which determines a set of applicable factors, and for each case whether the factor is present or absent. Such an analysis of the example cases is given in [7].

Let us denote the set of all cases as *C*.

We adopt the analysis given in [7], and identify four factors:
- Pliv = Π was pursuing his livelihood (Keeble, Young) favours Π
- Pland = Π was on his own land (Keeble) favours Π

- Nposs = Π was not in possession of the animal (Pierson, Keeble and Young) favours Δ
-Dliv = Δ was pursuing his livelihood (Young) favours Δ

Let us call the set of all factors identified *F*.

We also need to link factors to values. We say that the reason a factor favours a party is because deciding for that party in a case where that factor is present promotes or defends some value, which it held that the legal system should promote or defend. In the example, again following [7], the factor NPoss helps to promote clarity in the law and so discourage needless litigation; factor Pland helps promote the enjoyment of property; and factors Pliv and Dliv help to safeguard socially desirable economic activity. We thus have three values:

LLit = Less Litigation
MSec = More security of possession
MProd = More productivity

Let us call the set of all values identified *V*.

We need to associate with each factor the outcome favoured and the value promoted. We therefore represent information about factors in the form of *factor descriptions.* For simplicity, in this paper we assume that each factor promotes only one value, although the framework here introduced can be straightforwardly extended to allow sets of values in factor-descriptions.

Definition 1: *A factor description is a three tuple $<f, p, v>$, where $f \in F$, $p \in \{\Pi, \Delta\}$ and $v \in V$.*

Let us call the set of all factor descriptions *FD*. Note that here, and in subsequent definitions, Π and Δ represent case outcomes, rather that the litigating parties themselves: Π indicates the recognition of a legal remedy to the plaintiff, and Δ the denial of such a remedy For ease of later notation, let as denote as $\sim p$ the complement of *p*, in particular when $p \in \{\Pi, \Delta\}$, $\sim\Delta = \Pi$, and $\sim\Pi = \Delta$.

Cases can now be defined:

Definition 2: *A case is a three tuple $<c, CFs, p>$ where $c \in C$, $CFs \subseteq F$, representing the factors present in c, and $p \in \{\Pi, \Delta\}$, representing the outcome of the case.*

Note that CFs may, and typically will, contain factors favouring *both* parties.

We can use these definitions to introduce some dependent notions.

Definition 3: *A primitive rule is a pair $<f,p>$ where $<f,p,v> \in FD$.*

Let us call the set of all primitive rules *PR*. Since every factor is a reason to decide for one side or the other, there is a one to one mapping between *FD* and *PR*.

Definition 3 is intended to make plain the fact that any given factor is a reason for deciding for a particular party. Note that, according to this definition, a rule is considered to be inherently defeasible. No suggestion that the presence of *f* conclusively determines a decision for *p* is intended. By calling this connection

between a reason and its output a "rule" we also do not intend to suggest that the rule prevents or excludes the consideration of other reasons. Though those stronger, and more specific notions of a rule are frequently used in legal theory, and indeed are relevant in many contexts, we do not need them to present our model.

Definition 4: *A rule is a pair $<A,p>$, where for every $a \in A$ there is a primitive rule $<a,p> \in PR$.*

Note that $A \subseteq F$. Let us call the set of all rules *R*. Note that *PR* is a subset of *R*, containing the members of *R* where *A* has only one element.

We now introduce a way of getting from rules to values.

Definition 5: *The function val maps the elements of F to elements of V: $val(f) = v$ if and only if there is a factor description $<f,p,v>$.*

Definition 6: *The function ruleval maps the elements of R to elements of the power set of V, Pow(V): for all $v \in V$, $v \in ruleval(<A,p>)$ if and only if there is an $a \in A$ such that $val(a) = v$.*

Thus following a rule *r* will promote all the values in the set returned by *ruleval(r)*.

We now define the notion of conflict between rules.

Definition 7: *A rule, $<A_1,p_1>$ attacks a rule $<A_2,p_2>$ if and only if $p_1 = \sim p_2$.*

An attack may or may not succeed, depending on which rule is preferred. Preferences between rules are defined extensionally by the relation *pref*.

Definition 8: *The relation $pref(r_1,r_2)$ is a transitive binary relation on $R \times R$. It is intended to be read as "r_1 is preferred to r_2".* The set of all rule preferences is *Pref*. Note that preferences may exist between rules which do not attack one another.

We can now define defeat:

Definition 9: *A rule, r_1 defeats a rule r_2 if and only if r_1 attacks r_2 and not $pref(r_2,r_1)$.*

Values are also preferred to one another. Moreover combinations of values can be preferred to other combinations of values. Thus we have a relation *valpref* which is a transitive binary relation defined on $Pow(V) \times Pow(V)$. The set of all value preferences is *Valpref*. Whether a rule is preferred to another rule or not depends on the values it promotes or defends. Thus

Definition 10:
$pref(r_1, r_2)$ if and only if $valpref(ruleval(r_1), ruleval(r_2))$.

We are now in a position to define a theory:

Definition 11: *A theory is a five-tuple*
$<TC,TF,TR,Tpref,Tvalpref>$,
where $TC \subseteq C$, $TF \subseteq V$, $TR \subseteq R$, $Tpref \subseteq Pref$,
$Tvalpref \subseteq Valpref$.

The theory thus contains all the cases considered relevant by the proponent of the theory, all the factors chosen to describe those cases within the theory, all the rules available to be used in explaining the cases, and all the preferences between rules and values available to be used in resolving conflicts between rules. A theory is thus an explicit selection of the material available from the background.

3. CONSTRUCTING THEORIES

We assume that at the outset all of *TC, TF, TR, Tpref, and Tvalpref* are empty. The theory is then built up using a number of theory constructors. We will define these theory constructors in terms of their pre and post conditions. Essentially we need constructors to build up each element of the theory five-tuple. We begin by seeing how we can add cases.

Definition 12: *Include-case*
 Pre-condition: $<c,CF,p> \in C$,
    current theory is $<TC,TF,TR,Tpref,Tvalpref>$.
 Post-condition:
 current theory is $<TC \cup \{ <c,CF,p>\}, TF,TR,Tpref,Tvalpref>$.

Essentially we can select any case in *C*, and choose to include it. These are the cases that we aim to explain with our theory. Each party must include in the cases of his theory the current case, also called current situation, that is the case which is the object of the dispute. The current case has not yet been decided (or it is assumed so for the sake of the argument), and each party is claiming that it should be decided for their side. This is modelled here by assuming that two versions or the current case are contained in C, one with outcome Π (to be included in Π's theories) and one with outcome Δ (to be included in Δ's theories).

Cases bring with them factors, but we are not forced to consider in our theory all the factors associated with a case. We may believe some factors to be irrelevant. [9] has shown that it is not always obvious which factors should be considered when describing a case. We must therefore explicitly include each of the factors we wish to consider.

Definition 13: *Include-factor*
 Pre-condition: $<f,p,v> \in FD$,
    current theory is $<TC,TF,TR,Tpref,Tvalpref>$.
 Post condition: current theory is $<TC,TF \cup \{f\}$,
         $TR \cup \{<f,p>\},Tpref,Tvalpref>$.

Note that a factor, if included in the theory, is always a reason for deciding for one party or the other. Therefore the factor brings with it its associated primitive rule.

Cases typically contain several factors favouring a given party. Therefore we need a way of extending primitive rules so that they can be tailored to particular cases. These rules will contain more antecedents, and thus in general represent stronger reasons to decide for the favoured party than primitive rules.

Definition 14: *Factors-merging*
 Pre-condition: current theory is $<TC,TF,TR,Tpref,Tvalpref>$,
         $<A_1,p> \in TR$, $<A_2,p> \in TR$.
 Post-condition:
 current theory is $<TC,TF,TR \cup \{<\{A_1 \cup A_2\},p>,Tpref,Tvalpref>$.

A major role played by cases is to indicate preferences between rules. Assume that a theory *T* includes two conflicting rules, $<A_1, \Pi>$ and $<A_2, \Delta>$, with no preference between them, and a decided case $<c, CF, \Pi>$, to which both rules are applicable ($A_1, A_2 \subset CF$). As it stands the theory cannot explain the decision, since the conflicting rules attack each other and, in the absence of preferences, the attack is successful. But we can now ask: what does the case tell us about the relative merits of the two rules? We believe that the case, interpreted in the light of theory *T*, tells us precisely that the first rule was preferred to the second in that case. This is what one must presuppose, if one believes that theory *T* was the basis of the decision in *c*, i.e. that it prompted the decision-maker of case *c* to decide for Π. In other words, in the framework provided by *T*, one is authorised to assume or abduce that $pref(<A_1, \Pi>,<A_2, \Delta>)$, since this is required if *T* is to explain the decision in *c*. This assumption is not arbitrary, but rather grounded on the *evidence* provided by precedent *c* (similar to the way in which scientific theories are grounded in the evidence provided by empirical observations). Accepting this preference between two rules also commits us to a preference for the values promoted by the preferred rule over those promoted by the defeated rule. We therefore introduce a theory constructor to include such abductions based on the evidence of previous decisions in our theories.

Definition 15: *Preferences from case:*
Pre-condition:
    current theory is $<TC,TF,TR,Tpref,Tvalpref>$,
    $<c,CF,p> \in TC$,
    $<A_1, p> \in TR$ where $A_1 \subseteq CF \cap TF$,
    $ruleval(<A_1,p>) = V_1$,
    $<A_2,\neg p> \in TR$ where $A_2 \subseteq CF \cap TF$,
    $ruleval(<A_2, \neg p>) = V_2$.
Post-condition:
    current theory is $<TC,TF,TR,$
      $Tpref \cup \{pref(<A_1, p>,<A_2,\neg p>\},$
      $Tvalpref \cup \{valpref(V_1,V_2)\} >$.

We can also use value preferences to derive rule preferences. If we know that a value is preferred to another value, we may deduce from Definition 10 above that the rules promoting this value are preferred to rules promoting the other value.

Definition 16: *Rule preference from value preference*
Pre-condition: current theory is $<TC,TF,TR,Tpref,Tvalpref>$
      $r_1 \in TR, r_2 \in TR$,
      $ruleval(r_1) = V_1$
      $ruleval(r_2) = V_2$
      $valpref((V_1,V_2) \in Tvalpref$
Post-condition:
   current theory is $<TC,TF,TR,Tpref \cup \{pref(r_1,r_2)\},Tvalpref>$

Sometimes a case may lack some factors that were part of the antecedent of a rule used in a previous case. To make this rule applicable to the new case we must broaden it by dropping one or more of the antecedents. This is a common move in case based reasoning which we reflect in the following definition.

Definition 17: *Rule broadening*
Pre-condition:

current theory is <TC,TF,TR,Tpref,Tvalpref>,
$$<A_1,p> \in TR$$
Post-condition:
  current theory is <TC,TF,TR ∪ {<A_2,p>}, Tpref,Tvalpref>,
where $A_2 \subset A_1$.

Note that the rule obtained by *rule-broadening* could also be built up from primitive rules using *factors-merging*. In a sense therefore, this theory constructor is superfluous. We have included it, however, because it represents a move very common in accounts of case based reasoning.

Sometimes we will simply wish to assert a preference between rules, even though this cannot be justified on the basis of previous cases, or existing preferences between values. In doing so we commit to expressing a preference amongst the corresponding values.

Definition 18: *Arbitrary rule preference:*
Pre-condition: current theory is <TC,TF,TR,Tpref,Tvalpref>
$$r_1 \in TR, \ r_2 \in TR$$
Post-condition: current theory is
  <TC,TF,TR,Tpref ∪ {pref(r_1, r_2)},
  Tvalpref ∪ {valpref(ruleval(r_1),ruleval(r_2))} >

Similarly we may wish to assert a preference between values.

Definition 19: *Arbitrary value preference:*
Pre-condition: current theory is <TC,TF,TR,Tpref,Tvalpref>
$$f_1 \in TF, f_2 \in TF$$
$$val(f_1) = v_1, \ val(f_2) = v_2$$
Post-condition: current theory is
  <TC,TF,TR,Tpref,Tvalpref ∪ { valpref(v_1,v_2)}>

These arbitrary preferences are often required to justify a position when no position is determined by previous cases. What they do is make quite explicit the preferences that are being used to justify a position. In so doing they can pinpoint points of disagreement between the disputants, and which will be resolved when the case is decided.

The definitions 12 to 19 give us all we need to construct theories which can be advanced as explanations of particular case law domains.

4. USING THEORIES
The purpose of constructing a theory is to explain cases. We must therefore introduce the notion of explaining a case.

Definition 20: *A Theory <TC,TF,TR,Tpref,Tvalpref> explains a case c if and only if*
  <c,CF,p_1> ∈ TC
  <A_1,p> ∈ TR
  $A_1 \subseteq CF \cap TF$
For all <A_2,p_2> ∈ TR, such that $A_2 \subseteq CF \cap TF$, < A_2,p_2> does not defeat <A_1,p_1>

Informally, the definition says that a case is explained if we have a rule which allows us to conclude its outcome on the basis of the factors present in the case which are included in the theory, and this rule is not defeated by any other rule in the theory whose

antecedent is satisfied. The overall aim of a disputant is to construct a theory that explains the current case, with the outcome desired by the person advancing the theory.

Let us illustrate this by constructing some theories to explain the three wild animal cases. We will suppose that *Young* has not yet been decided, that is, *Young* is our current case. If we wish to argue for the plaintiff, we will include the case with desired outcome for the plaintiff *<Young,{Pliv,Nposs,Dliv},Π>* in our theory, and then construct a theory which explains it. Conversely if we wish to argue for the defendant we will include *<Young,{Pliv,Nposs,Dliv},Δ>* as the starting point of our theory.

A simple pro-defendant theory is:
T_1: <{<Young,{ Pliv,Nposs,Dliv},Δ>,
    <Pierson,{Nposs},Δ>},
    {Nposs},{<{Nposs},Δ>},{},{}>

T_1 can be constructed using *include-case* to add *Pierson* and *include-factor* to add Nposs.

This theory simply expresses the view that the plaintiff had no remedy (Δ) in *Pierson*, since he did not have possession of the animal (Nposs), which is indeed a reason for Δ (according to the rule *<{Nposs},Δ>*). Exactly the same reasoning also explains why the plaintiff should have no remedy in *Young* also. No preferences are necessary: In T_1, TR contains a single rule, and hence this rule is not attacked, and so cannot be defeated: it thus allows T_1 to explain both *Young* and *Pierson*.

The plaintiff can, however, produce a theory relying on *Keeble*, and subsuming T_1:
T_2: <{<Young,{ Pliv,Nposs,Dliv },Π>,
<Pierson,{Nposs},Δ>,<Keeble,{ Pliv,Nposs,Pland},Π>},
{ Pliv,Nposs },
{<{Nposs},Δ>,<{Pliv},Π>},{pref(<{Pliv},Π>,<{Nposs},Δ>)},
{valpref(Mprod,Llit)}>

This theory is obtained, starting from T_1, by including *Keeble*, including factor Pliv (*Π* was pursuing his livelihood), with rule *<Pliv,Π>*, and using *preferences-from-case* to get the required rule and value preferences from *Keeble*. As in T_1, T_2 implies that the plaintiff had no remedy in *Pierson* since he did not have possession of the animal. However, T_2 implies that the plaintiff had a remedy (*Π*) in *Keeble* since he was pursuing his livelihood (Pliv). Although the rule *<{Nposs},Δ>* applies to *Keeble*, this is defeated since Pliv supports Π more strongly than not having possession of the animal (Nposs) supports Δ (from the preference *pref(<{Pliv},Π>,<{Nposs},Δ>)*). According to the same reasoning, T_2 implies that *Young*, which shares with *Keeble* factors Pliv and Nposs, should also be decided for *Π*. Note that it is *pref(<{Pliv},Π>,<{Nposs},Δ>)*, derived from *Keeble*, which allows the rule *<{Pliv},Π>* to defeat the rule *<{Nposs},Δ>*. This means that the theory can explain why *Keeble* was decided for *Π* and why *Young* should be decided in the same way. Note also that the additional *Δ*-factor in *Young*, i.e. Dliv, has not been included in T_2, and is therefore is not available to contest the explanation. Similarly, the theory does not consider the additional *Π*-factor in *Keeble*, i.e Pland (*Π* was on his own land). For the proponent of T_2, neither of these factors is considered relevant. The defendant

can, however, make use of those factors and respond to T_2 in two different ways, depending on the factor he chooses to include. First he might add *Keeble* and factors Pliv and Pland to T_1 to get T_{3a}:

T_{3a}: <{<Young,{ Pliv,Nposs,Dliv },Δ>,<Pierson,{Nposs},Δ>,
    <Keeble,{Pliv,Nposs,Pland },Π>},{ Pliv,Nposs,Pland },
    {<{Nposs},Δ>,<{Pliv},Π>,<{Pland},Π>},{},{}>

At this point, neither *Young*, nor *Keeble* is explained, since in the absence of preferences, rules attacking each other defeat each other (this is the case for <{Nposs},Δ>,<{Pliv},Π>, and <{Pland},Π>). Clearly, the defendant does not want to explain *Keeble* as the plaintiff did, i.e. by using the rule <{Pliv}, Π> with the preference *pref(<{Pliv},Π>,<{Nposs},Δ>).* This would lead, as we have just seen, to *Young* being decided for the plaintiff, on the basis of the same reasoning. He can avoid all of that, by using *factors-merging* to add the rule *<{Pliv,Pland}, Π>*, and *preferences-from-case* to add the preference derived from *Keeble*, taking into account these factors, *pref(<{Pliv,Pland }, Π>,<{Nposs}, Δ>)*. In this way the theory explains why *Keeble* was decided for Π without implying the same decision for *Young*: the plaintiff had a remedy (Π) in *Keeble* since he was both pursuing his livelihood (Pliv) and on his own land (Pland), and the *combination* of these two factors supports Π more strongly that not having possession of the animal (Nposs) supports Δ (according to the preference *pref(<{Pliv, Pland},Π>,<{Nposs},Δ>))* . Note that the preference derived from *Keeble* is now different: Keeble is explained by giving priority to the rule *<{Pliv, Pland},Π>* rather then to the rule *<{Pliv},Π>* . Therefore, the reasoning of *Keeble* cannot now be applied to *Young*, where there is only Pliv (and not Pland) to support decision Π.

T_{3b}: <{<Young,{ Pliv,Nposs,Dliv },Δ>,<Pierson,{Nposs},Δ>,
    <Keeble,{ Pliv,Nposs,Pland },Π>},
{ Pliv,Nposs,Pland },
    {<{Nposs},Δ>,<{Pliv},Π>,<{Pland},Π>,
      <{Pliv,Pland},Π>},
     {pref(<{ Pliv,Pland },Π>,<{Nposs},Δ>)},
      {valpref({Mprod,Msec},Llit)}>

Unfortunately T_{3b} does not explain why *Young* should be decided for Δ. For this purpose, one would need the rule preference *pref(<{Nposs},Δ>,<{Pliv},Π>)*, which would have to be either added arbitrarily or derived from the arbitrarily added value preference *valpref(Llit,Mprod)*. (Remember that one's preference is arbitrary when it does not explain any precedent, but only supports the decision one wishes to have in current case.)

T_{3c}: <{<Young,{ Pliv,Nposs,Dliv }, Δ>,<Pierson,{Nposs},Δ>,
    <Keeble,{ Pliv,Nposs,Pland },Π>},
  { Pliv,Nposs,Pland },
   {<{Nposs},Δ>,<{Pliv},Π>,<{Pland},Π>,
     <{ Pliv,Pland },Π>},
    {pref(<{ Pliv,Pland },Π>,<{Nposs},Δ>),
      pref(<{Nposs},Δ>,<{Pliv},Π>)},
    {valpref({Mprod,Msec},Llit), valpref(Llit,Mprod)}>

T_{3c} suffices for the defendant, but the resort to arbitrary preferences is not desirable. A different tack for the defendant would be to ignore Pland and add Dliv instead to T_2.

T_{4a}: <{<Young,{ Pliv,Nposs,Dliv },Δ>,<Pierson,{Nposs},Δ>,

<Keeble,{ Pliv,Nposs,Pland },Π>},
{ Pliv,Nposs,Dliv },
{<{Nposs},Δ>,<{Pliv},Π>,<{Dliv},Δ>},
{pref(<{Pliv},Π>,<{Nposs},Δ>)},
{valpref(Mprod,Llit)}>

Now, by merging the primitive rules for Nposs and Dliv, introducing the value preference *valpref({Mprod,Llit},Mprod)*, and using this to derive the rule preference *pref(<{Nposs,Dliv},Δ>,<Pliv,Π>)*, an explanation of *Young* can be obtained.

T_{4b}: <{<Young,{ Pliv,Nposs,Dliv },Δ>,<Pierson,{Nposs},Δ>,
    <Keeble,{ Pliv,Nposs,Pland },Π>},
  { Pliv,Nposs,Dliv },
  {<{Nposs},Δ>,<{Pliv},Π>,<{Dliv},Δ>,
    <{Nposs,Dliv},Δ>},
   {pref(<{Pliv},Π>,<{Nposs},Δ>),
     pref(<{Nposs,Dliv },Δ>,<{Pliv},Π>)},
    {valpref(Mprod,Llit), valpref({Mprod,Llit},Mprod)}>

Therefore, according to theory T_{4b} *Young* should be decided for Δ since in *Young* the rule *<{Nposs,Dliv},Δ>* is not defeated. This seems, according to [7], to be the theory used by the judges in *Young*. This explanation does rely on the introduction of a preference that is arbitrary, in the sense of not being supported by precedents. However it might be held that *valpref({Mprod,Llit},Mprod)* is not entirely arbitrary on a different ground, namely since *{Mprod,Llit}*, is a superset of *{Mprod}*. The idea is that if all values are good, then a more inclusive set of values must be better that a less inclusive one (cf. [13] and [15]). This idea could be adopted into our framework by adding a theory constructor which allows one to introduce preferences for any set of values over its own proper subsets. We believe that this assumption is reasonable in many contexts, but possibly not in all, because of interferences between values: if two values are incompatible, then promoting only one of them can be better then promoting the two of them at the same time. So, we do not wish to commit to this being a general and necessary feature of our approach. None the less we would expect a preference of this sort to be accepted in most cases.

5. EVALUATING THEORIES

In the above discussion we produced four theories, each of which would explain the decision in *Young*. How do we choose between them? Intuitively theories are assessed according to their coherence. We will not, however, even attempt to develop a precise notion of coherence in this paper. For coherence in law, there is a discussion in [2], and for a general discussion of coherence and theory change, see [16]. For a recent attempt to develop some formal criteria with which to assess theories see [8]. In this paper we will do no more than indicate some considerations which might lead to one theory being preferred over another.

Firstly, we demand as much *explanatory power* as possible from our theories. In this context explanatory power can be approximately measured by the number of cases explained. More exactly, since different cases may have different weights (one case being more recent, or having been decided by a higher court, etc.) we should consider also the relative importance of the sets of cases that the competing theories can explain. We cannot consider

here the details of the metrics for such a comparison, which is also dependant on the features of the legal system under consideration. At the very least, however, we can certainly say that theory T_1 has more explanatory power than theory T_2, if T_1 explains all precedents explained by T_2 and some more, i.e. if the precedents explained by T_1 are a proper superset of the cases explained by T_2.

Secondly we can require theories to be *consistent*, in the sense that they should be free from internal contradiction. Note that we allow theories to include conflicting rules applicable to the same case, and we assume that these conflicts are solved through preferences. The contradictions we wish to avoid are those concerning rule and value preferences, i.e. the *pref* and *valpref* relations. Thus we can require that theories do not contain both *pref(r_1, r_2)* and *pref(r_2, r_1)* in *Tpref*, and do not contain both *valpref(v,v')* and *valpref(v',v)* in *Tvalpref*. Such incoherence is explicit. There is also implicit incoherence when there is a value preference which would allow the introduction of a rule preference which would produce an incoherence in *Tpref*, or where the transitivity of the preference relations can be used to derive an explicit contradiction.

A third classically desirable feature of scientific theories is *simplicity*. This could be measured in terms of the number of factors in *TF*. If we can explain a set of cases without introducing a given factor, this is a simpler theory than one which does include that factor. Suppose we extend T_{4b} above to include factor Pland.

T_5: <{<Young,{ Pliv,Nposs,Dliv },Δ>,<Pierson,{Nposs},Δ>,
   <Keeble,{ Pliv,Nposs,Pland},Π>},
  { Pliv,Nposs,Pland,Dliv},
   {<{Nposs},Δ>,<{Pliv},Π>,<{Dliv},Δ>,<{Pland},Π>
      <{ Nposs,Dliv },Δ>,<{ Pliv,Pland},Π>},
    {pref(<{Pliv,Pland },Π>,<{Nposs},Δ>),
      pref(<{Nposs,Dliv },Δ>,<{Pliv},Π>)},
   {valpref({Mprod,Msec},Llit),
     valpref({Mprod,LLit},Mprod)}>

Suppose we now have a new case in which the facts of *Keeble* are present, except that the plaintiff is hunting on common land. T_{4b} would explain a decision for the plaintiff, whereas T_5 would not explain either outcome. To explain an outcome for the plaintiff, T_5 would need the value preference *valpref(Mprod,Llit)* ($T5a$), and to explain an outcome for the defendant, the value preference *valpref(Llit,MProd)* (T_{5b}), to get the required preference between the rules *<{Pliv},Π>* and *<{Nposs},Δ>*. In either case such an introduction would be arbitrary. We would therefore expect the plaintiff to rely on T_{4b}, whereas the defendant would advance the more complicated theory T_{5b}. If the case were to be found for the defendant, we could justify the complication of T_5 by its additional explanatory power, but if it were found for the plaintiff we should have no reason to complicate T_{4b}, since we get no gain in explanatory power. If decided for the plaintiff, there would be no reason to think that Pland was a relevant factor at all. Indeed [7] argues that Pland is a red herring with respect to the cases under consideration.

An argument could, however, be mounted for preferring theories with *more* factors. Whenever a theory does not consider a factor that was present in one of its cases, that factor can be introduced, so jeopardising any rule (and value) preferences included in the theory based on that case, and so threatening its ability to explain

its cases. The use of factor Pland in T_3 above to challenge T_2 is an example of this. Thus a theory is *safer* in proportion to the completeness of the factors it considers when using a case to derive a rule preference. Whether we should look for simplicity or safety depends on the status of the factors. If they have been used in the past decisions, completeness is desirable, but if, even though they do provide a reason, they have played no part in previous decisions, simplicity is to be preferred. Such a choice requires reference back to the full text of decisions, and cannot be settled in a general way.

Finally a theory is better in so far as less recourse to arbitrary preferences have been made. In moving from T_5 to T_{5a} and T_{5b} above it was necessary to add an arbitrary value preference. Such moves can only be justified externally to the theory, by an appeal to intuition or the like. In only one case does this seem to be entirely convincing, namely the arbitrary preference in T_{4b}, *valpref({Mprod,Llit},Mprod)*, does seem plausible because the preferred value is a superset of the other value. As we have said above, we might even wish to have an additional theory constructor legitimising the introduction of such value preferences.

6. ARGUMENT MOVES AND THEORIES

It is now interesting to relate the moves made in a HYPO style argument to the above account of theories. A reconstruction of two of these moves, in terms of its own formalism, has been given in [13]. Where appropriate we will make comparisons with this work.A key element of our perspective on case based reasoning, is that reasoning with cases involves a number of related, but distinct, activities: namely first *constructing* a theory, then *using* the theory to explain cases, and finally *evaluating* competing theories, so as to adjudicate between competing explanations. The above discussion was structured around these three elements. Given this perspective, it is possible that argument moves in traditional case based systems, which do not make this distinction, conflate these elements.

6.1 Citing a Case
Citing a case just involves extending a theory with one additional precedent case. Typically, however, when this is done for a purpose, citing a case also involves expanding the theory with rules and preferences so that it can explain the cited case, and others included in the theory. An example above is T_1, which cites *Pierson* in support of the defendant in *Young* by introducing the case <Pierson,{Nposs},Δ>, and a rule sufficient to explain it, that is *<{Nposs},Δ>*. This citation is a particularly simple one, since the theory does not contain any rule which would require the case to have a different outcome. If the theory already includes such a rule, than the citation of a case also involves the introduction of a preference which explains why the case deserved the decision it had as a matter of fact, through the constructor *preferences-from-case*. As an example of this more complex type of citation, consider where the defendant constructs theory T_2 by citing *Keeble*. At this stage he introduces *<Keeble,{Pliv,Nposs,Pland},Π>*, the rule *{Pliv},Π>*, and also the preference *pref(<{Pliv},Π>,<{Nposs},Δ>*, which enables the theory to explain *Keeble*. Pragmatically the best case to cite is the one which includes as many factors in common with the current case as possible. This allows the most specific, and thus safest,

rule to be constructed, and thus pre-empts several possible challenges. This citing a case is essentially a move of theory construction, although considerations as to which is the *best* case to cite looks forward to the evaluation of the theory. Moreover, as implemented in HYPO, the criterion for choosing the best case favours safety over simplicity in theory evaluation.

6.2 Counter Examples and Distinctions

HYPO permits two different responses to a cited case: providing a counter example and distinguishing the case. Providing a "trumping" counter example is the stronger move because it will include another case in an opponent's theory so as to licence rule preferences such that the resulting theory will explain both the counter example case and the cited case, besides giving the current case the result wished by the citing party. It thus wins on explanatory power. The use of *Keeble* in T_2 is an example of this move. Introducing counter examples is part of theory construction, but their strength derives from theory evaluation, in that an "as on point" counter example does no more that display a failure to explain certain cases on the part of the theory, whereas the trumping counter example gives rise to a new theory superior in explanatory power. In [13] the idea is that counter examples are evaluated not in terms of on-pointness, but in terms of a comparison between the values promoted. A trumping counter example will always succeed because it promotes at least as many values as the case to which it is a counter example (remember that for [13] a set of values is always preferred to its proper subsets). On the other hand, a non-trumping counter example both lacks a value present in the precedent and has a new value not present in the precedent, so whether it succeeds depends on how these values are compared. A counter example is dismissed if the required value preference cannot be added to the theory. Indeed the theory may already contain value preferences which show that the counter example is ineffective.

There are two ways of distinguishing a case in HYPO. Either one points to a factor favourable to one's opponent present in the precedent and absent in the current case, or one points to a factor favourable to oneself present in the current case and absent in the precedent. Here we discuss only the first of these; similar considerations apply to the other case. One way of distinguishing a case involves introducing a new factor *f*, which is *in favour of the opponent*, and which is not already present in the opponent's theory. This factor is not contained in the current case, but is present in the precedent licensing the *preferences from case* move which produced the preference *pref(<A₁, p>, <A₂, ~p>)*, which allowed the opponent's theory to explain the current case. Once the new pro-opponent factor *f* is introduced, the old rule *<A₁, p>*, which explained why the precedent was decided for the opponent (and why the current case should be decided in the same way), is extended into *<A₁ ∪{f}, p>*, and a new preference *pref(<A₁ ∪{f}, p>, <A₂, ~p>)* is provided to explain the precedent. The latter preference does not apply to the current case (which does not contain factor *f*). Moreover, once the new, more specific preference is available, the old preference becomes unnecessary to explain the precedent, and so fails to provide a convincing ground for the decision of the current case. The introduction of factor Pland in T_3 above exemplifies the *distinguishing* move: by introducing this additional factor, the defendant transformed the rule *<{Pliv},Π>* into the rule *<{Pliv,Pland},Π>*, which he then used to explain the case *<Keeble, {Pliv,Nposs,Pland},Π>*,

according to the preference *pref(<{ Pliv,Pland },Π>,<{Nposs},Δ>)*. The new rule (and the corresponding preference) are not applicable to the current case, *Young*, which has factors Pliv, Nposs, Dliv, and so does not contain Pland, required if the new rule is to be applied. On the other hand, in this new theory, the old rule *<{Pliv},Π>* and the corresponding preference *pref(<{ Pliv },Π>,<{Nposs},Δ>)* can be dismissed as being redundant because it has no explanatory function. Therefore according to the new theory, a *Π* decision in *Keeble* is consistent with a *Δ* decision in Young, which is what the defendant wanted to establish. The move is less powerful that a trumping counterexample because it does not form the basis for a different decision of the current situation, but merely blocks the opponent's theory. In conclusion, this theory construction move involves a factor rather than a case. The effect of the move is to render the original theory weaker because it makes its rule preference arbitrary rather than grounded in a precedent. An as-on-point counter example can also be seen as the combination of a distinguishing move together with a case which grounds an alternative theory, based on different factors. This theory can, of course, then be subject to a distinguishing move itself. We thus end up with two theories which both require arbitrary preferences in order to explain the current case. To be effective the distinguishing factor must relate to a value which can be shown to be preferred, so that arbitrary preferences are not required. This is what happened above in T_{4b} when *Dliv* is used to distinguish *Young* from *Keeble*. This is an example of the second kind of distinguishing move (i.e. one introduces a new factor favourable to oneself), but its greater effect comes from the value associated with the distinguishing factor, not from it being an example of this other way of distinguishing.

6.3 Other Moves in CATO

There are four other argument moves introduced in CATO [1]: *emphasise strengths*, *show weaknesses not fatal*, *emphasise a distinction* and *down play a distinction*.

The first of these simply corresponds to introducing more cases which are explained by the theory, with factors shared with current case, thus increasing the theory's explanatory power. Again these moves can be seen as constructing a theory which will be evaluated as better. Showing weaknesses not fatal is perhaps more interesting, in that it seems to suggest a different understanding of the rules derived from cases from that described above. For the absence of a factor to be fatal, it would have to be a necessary condition, and as we have described the situation above, case law can never give us such conditions, but only defeasible rules. The move would also involve including cases found for the desired side, but this time containing factors favourable to the other side which lead to defeated rules. In our terms therefore it can be seen as an attempt to increase the safety of the explanations in the theory, by anticipating and pre-empting the introduction of additional factors. It is also possible that such cases may licence the introduction of preferences which contradict preferences arbitrarily introduced by an opponent.

Emphasising and downplaying distinctions involve recourse to a hierarchy of factors, introduced by CATO, which gives a more refined description of factors than Definition 1 above provides. A fully satisfactory account of these moves would also require a

more elaborated notion of a case being explained than the one provided above, so as to allow for arguments to be chained to an arbitrary length, which could have conflicts at different points in the chain. This is impossible in the framework we have so far presented, since we do not accommodate chaining of rules. A logic which would provide the necessary support is given in [11]. Here, however, we will take a step towards accommodating these moves by providing a simplified model where only two step arguments are allowed and it is assumed that there is no conflict between intermediate factors. We believe that this indicates how our simple model could be developed into the richer model required to explain chains of arbitrary length, with conflict between intermediate factors. So as to convey the required hierarchy information, let us first redefine factor-descriptions as follows:

Definition 1b: *A factor description is a four tuple $<f,p,v,m>$, where $m \subseteq F \cup \{\Pi, D\}$ represents the intermediate factors through which f contributes to p. In cases where f leads directly to p, $m = p$.*

When a factor with a description like this is included in a theory, three rules are added to the theory, rather than the single primitive rule of definition 13: one reflecting the impact of the factor on the ultimate conclusion, i.e. $<f, p>$ (the rule added in definition 13), one reflecting the impact of the factor on the intermediate factor, i.e. $<f, m>$, and one reflecting the impact of the intermediate factor on the ultimate conclusion, i.e. $<m, p>$.

We need to modify the definition of a case being explained by a theory so that it covers the situation where an intermediate factor is used. Let us first introduce the notion of a factor f being an immediate consequence of a set of factors FS, according to a theory T, with rules RS.

Definition 20b: *The set of the immediate consequences of a set of factors FS, with respect to a set of rules RS, is the smallest set $IC_{FS,RS}$ such that*

$FS \subseteq IC_{FS,RS}$ *and*

If there is a rule $<A,q> \in RS$ such that $A \subseteq FS$, then $q \in IC_{FS,RS}$

Definition 20c: *A theory $<TC,TF,TR,Tpref,Tvalpref>$ explains a case c if and only if*

$<c,CF,p> \in TC$

$<A_1, p\} > \in TR$

$A_1 \subseteq IC_{CF,TR}$.

and there is no $<A_2, \sim p> \in TR$ such that $A_2 \subseteq IC_{CF,TR}$ which defeats $<A_1, p\}>$

Downplaying the distinction of a precedent case c_{prec} in regard to the current case c_{curr}, consists in providing an explanation for both c_{prec} and c_{curr} through a rule including an intermediate factor, which is an immediate consequence both of the distinguishing factor, and of a different factor present in the current case. Suppose the plaintiff's theory contains:
- the current case $<c_{curr}, CF_{curr}, \Pi>$,
- a precedent case $<c_{prec}, CF_{prec}, \Pi>$,
- rules $<A_1, \Pi>$ and $<A_2, \Delta>$, with $A_1 \cup A_2 \subset CF_{curr}, CF_{prec}$,
- a preference $rulepref(<A_1,\Pi>,<A_2, \Delta>)$, which was introduced to explain c_{prec}.

According to this theory, the explanation of c_{prec} (c_{prec} had decision Π since it includes the A_1 set of factors, which supports Π more strongly then the set A_2 supports Δ) also applies to c_{curr} The defendant can reply by distinguishing c_{prec} from c_{curr} by introducing a new factor f, such that $f \in CF_{prec}$ and $f \notin CF_{curr}$, a rule $<A_1 \cup \{f\}, \Pi>$, and a preference $rulepref(<A_1 \cup \{f\}, \Pi>, <A_2, \Delta>)$. At this point case c_{prec} can be explained by appealing to the latter rule and preference, which are not applicable to c_{curr} (since to c_{curr} does not include f). In other words, the plaintiff's explanation of c_{prec} (c_{prec} had decision Π since it includes the set of factors $A_1 \cup \{f\}$, which supports Π more than A_2 supports Δ) does not apply to c_{curr} (which does not contain $A_1 \cup \{f\}$, but only A_1).

Let us however assume that f has description $<f, \Pi, v, m>$, and that c_{curr} contains a factor f', with factor *description* $<f',\Pi,v'$, $m>$: both f and f' promote decision Π through promoting the intermediate factor m. The plaintiff can then downplay the defendant's distinction (which was obtained by introducing f), by introducing f' and the rules $<\{f\}, m>$ and $<\{f'\}, m>$. This move allows the plaintiff to explain c_{prec} through the *rule $<A_1 \cup \{m\}, \Pi>$* using the new preference $pref(<A_1 \cup \{m\}, \Pi >, <A_2, \Delta>)$. The latter preference, while having an explanatory role in regard to c_{prec} (c_{prec} was decided for Π since factors $A_1 \cup \{m\}$ support Π more strongly then A_2 supports Δ), also supports decision Π in case c_{curr} (which also includes $A_1 \cup \{m\}$). So, the plaintiff achieves the result of disarming the defendant's attempt at distinguishing c_{prec} from c_{curr}: in his new theory the explanation of the precedent still supports the same decision in the current case. Let us now provide a formal definition

Definition 21: *Downplaying distinction of a precedent case c_{prec} with regard to the current case c_{curr},*
Pre-condition: current theory is $<TC,TF,TR,Tpref,Tvalpref>$

$\{<c_{prec},CF_{prec},p>, <c_{curr},CF_{curr},p>\} \subseteq TC$,

$\{<A_1 \cup \{f\}, p>, <A_2, \sim p >\} \in TR$,

$A_1 \cup A_2 \subseteq CF_{prec}, A_1 \cup A_2 \subseteq CF_{curr}$,

$f \in CF_{prec}, f \notin CF_{curr}$,

$pref(<A_1 \cup \{f\}, p>, < A_2, \sim p >) \in TR$,

$\{<f,p,v,m, <f',p,v',m>\} \subseteq FD$,

Post-condition: current theory is
$<TC,TF^D,TR^D, Tpref^D,Tvalpref^D >$
where,

$TF^D = TF \cup \{f', m\}$,

$TR^D = TR \cup \{<f,m>,<f',m>, <A_1 \cup \{m\}, p>\}$,

$Tpref^D = (Tpref - \{rulepref(<A_1 \cup \{f\}, p>, < A_2, \sim p>)\} \cup \{rulepref(<A_1 \cup \{m\}, p>, < A_2, \sim p>\}$,

$Tvalpref^D = (Tvalpref - \{valpref(ruleval (<A_1 \cup \{f\},p>), \{ruleval(<A_2, \sim p>)\}) \cup \{valpref(ruleval(<A_1 \cup \{m\}, p>), ruleval(<A_2, \sim p>)\}$

After downplaying, it is still possible to re-introduce a distinction, by claiming that the distinguishing factor also causes another intermediate consequence, favourable to oneself, which does not hold in the current situation. Remember that downplaying takes place after one party distinguished precedent c_{prec} from the current situation c_{curr}, by arguing that a factor f, which is only present in the precedent, was necessary for producing the outcome p of the precedent, according to a rule $< A \cup \{f\}, p>$. As we have just seen,

the opponent can downplay the distinction by showing that *p* was produced via an intermediate factor *m*, and that *m* follows both from the factor *f* in the precedent and the factor *f'* in the current situation. At this stage, however, the distinguishing party can try to reinstate the distinction, by showing that *f* also produces another intermediate factor m_2, which is not produced by *f'*.

For example, consider the case where a patient was cured by a doctor in a hospital, without there existing a contract, and the doctor behaved with minor carelessness, so causing a damage to the patient (the factors are: *Hospital*, *Carelessness*, *Damage*, *MinorCarelessness*). Assume that there is a precedent where a doctor was considered to be contractually liable for his carelessness, a contract being in place, even though the carelessness was a minor one (the factors in c_{prec} were: *Contract*, *Carelessness*, *Damage*, *MinorCarelessness*). Assume also that the patient explains both this case and the current situation (with output *Liability*) according to the theory that damage caused by medical carelessness produces the liability of the doctor (i.e. according to the rule *<{Carelessness, Damage}, Liability>*, which prevails over *<{MinorCarelessness}, NoLiability)>*). Finally, assume that the doctor distinguishes the precedent from the current situation by claiming that in the precedent there was a contract, i.e. she proposes a theory where a doctor's liability for minor carelessness requires a contract (i.e. she explains c_{prec} according to the preference *pref(<{Contract, Careless, Damage}, Liability>, <{MinorCarelessness}, NoLiability)>)*. The patient may downplay this distinction, by claiming that the existence of the contract implied that the doctor was warranting a careful performance, and that this was the real reason why a doctor was held liable for a minor carelessness (i.e. he adds to the theory the rules *<{Contract}, Warranty>*, *<{Careless, Damage, Warranty>, Liability>*, and substitutes the preference above, with a preference concerning the latter rule, i.e. *pref(<{Warrant, Careless, Damage}, Liability>, <{MinorCarelessness}, NoLiability)>)*. He also claims that the same warranty is also implicitly given by the practice of the medical profession in a hospital, regardless of the existence of a contract (he adds the further rule *<{Hospital}, Warranty>*), so that the doctor in the current situation would still be liable for the same reasons (i.e carelessly causing a damage under a warranty) as the doctor liable in the precedent. Now the doctor, to reinstate the distinction, may claim that the contract in the precedent also implied that there was a consideration, and that both consideration and warranty are required to ground the standard of care which is necessary for a doctor to be liable for minor carelessness {i.e. she adds the rules *<{Contract}, Consideration>*, *<{ Consideration, Warrant, Carelessness, Damage}, Liability>* and the priority *pref(<{Consideration, Warrant, Careless, Damage}, Liability>, <{MinorCarelessness}, NoLiability)>)*. The dispute may then go on, with the patient still trying to downplay this further distinction (e.g. for example by claiming that the doctor was going to be paid for her work at the hospital in any case, so that there was in a sense a consideration), and the doctor trying to introduce further distinctions, still based on the absence of a contract in the current situation.

If we wish to see downplaying a distinction as a theory construction move, we could use definition 21 as an additional theory constructor. This would mean that a downplayed distinction was never accepted, since when a distinction is downplayed, we always produce a theory with the same effect as before the distinction was drawn, by adding in the required rules establishing the links through intermediate factors. This theory would, however, be subject to attacks producing more refined theories along the lines of the above paragraph.

An alternative, although more restrictive, way of downplaying distinctions, given in [13], simply requires that some factor in the current case promote the same value as the distinguishing factor. If this is so, although the factors differ, the competing rules return the same values when given as arguments to ruleval, and thus can enjoy the same preference relations.

In contrast, emphasising a distinction does not give rise to new theories: it only draws attention to the non-availability of the downplaying move, and the consequent need for the opponent to resort to arbitrary preferences to repair his theory. In [13] this is expressed in terms of a difference between the values associated with the two sets of factors. The very difference in values alerts us to the significance of the distinction, which requires a consideration of the value preferences to resolve. The move is, of course most effective, if the distinction relates to a more highly prized value.

A recent paper by Roth [14] suggests some further moves that can be made to augment the notions of downplaying and up-playing distinctions. These mostly turn on a richer account of intermediate moves in arguments: for example a factor may only promote an intermediate factor in the presence of some other factor. His example is that the potential to find another job may be an intermediate factor, but whether a particular job promotes this factor depends on the current state of the labour market, Such moves would require a more sophisticated logic to use theories. It would also be an interesting exercise to see what extensions to our basic formalism would be necessary to represent the information needed to make such moves.

7. IMPLEMENTATION

We have implemented the framework defined in Definitions 1-20 above in PROLOG. The process of implementation was relatively straightforward as the definitions have a fairly direct mapping into PROLOG. For example, the implementation of Definition 18 is:

```
arbitraryRulePref(Theory,R1,R2):-
    retract(theory(Theory,TCases,
      TFactors,TRules,TPrefs,TValPrefs)),
      member([R1,[F1,O]],TRules),
      member([R2,[F2,O2]],TRules),
      ruleval(F1,V1), ruleval(F2,V2),
asserta(theory(Theory,TCases,TFactors,
        TRules,[pref(R1,R2)|TPrefs],
          [vp(V1,V2)|TValPrefs])).
```

Once we implemented the definitions we embedded them in a menu system. The menu simply provides access to calls to the theory constructors, together with options to display the theory, list the cases explained by the theory, and to list the background cases and factor definitions. The program is perhaps not very exciting, but it does allow a user to construct theories ensured to be correct in accordance with the definitions, and to check that the expected consequences can be delivered.

A second version of the program constructs an audit trail during theory construction. On the basis of this we can produce the metrics necessary to evaluate the theories. In particular we record the cases and factors used to establish preferences, to assess explanatory power and simplicity/safety, and the number of arbitrary preferences which have been included. A further straightforward refinement would be to check for contradictions in the rule and value preferences.

The ultimate aim would be to construct a program which produces the best theory for a given side in a given new case against a given background of precedents and factors. Such a program would require both heuristics to pick the theory constructors to use, and to evaluate the theories constructed. This would, of course, require a precise description of the evaluation criteria. Experiments might be tried with different weight being given to the various criteria. All this is reserved for future work, although we feel that we have some firm foundations on which to build such a study.

8. SUMMARY

The main object of this paper has been to present a formal account of case law as a theory, and to provide constructors to build such a theory. We have illustrated the formalism using a well known example. We have shown how such theories could be used to argue for and against positions with respect to as yet undecided cases, and to account for a body of decided cases. Finally we have shown how the basic moves in the currently most accepted account of reasoning with case law can be expressed in our formalism. We have implemented a system which supports a user in the construction and evaluation of such theories. We believe that the work provides insight into reasoning with cases in the following ways:

§ We see reasoning with cases as involving all of the construction of a theory, the application of that theory, and the evaluation of competing theories. The above gives a precise account of all these elements, while maintaining a clear conceptual separation between them.

§ The formal account is sufficient to reconstruct the reasoning and argument moves of HYPO, making clear at which points theory construction, application and evaluation are involved.

§ The formal account can be extended to accommodate systems, such as CATO, which go beyond HYPO style reasoning. We have sketched a limited extension to indicate how a distinctive move of CATO could be incorporated.

§ A problem with many accounts is that they indicate what arguments can be made, but fail to account for why an argument might be persuasive. Our account distinguishes two levels of persuasion. First we can show some arguments to be persuasive in terms of values: a value preference derived from a rule preference exhibited in a decided case can be used to derive a new rule preference applicable to other cases. Second arguments can be found persuasive by a comparative evaluation of the competing theories from which they derive. We have suggested some criteria which could be used to compare theories.

Of course, much work is needed to capture the full richness of current case based systems. None the less we think we have provided some firm foundations on which such work can be built.

10. REFERENCES

[1] Aleven, V., (1997). Teaching Case Based Argumentation Through an Example and Models. PhD Thesis. The University of Pittsburgh.

[2] Alexy, R., & A. Peczenik. 1990. The Concept of Coherence and Its Significance for Discursive Rationality. *Ratio Juris* 3: 130-147.

[3] Ashley. K.D., (1990). Modelling Legal Argument. Bradford Books, MIT Press, Cambridge, Mass.

[4] Bench-Capon, T.J.M., (2000). The Missing Link Revisited: The Role of Teleology in Representing Legal Arguments. Submitted to *Artificial Intelligence and Law*.

[5] Bench-Capon, T.J.M., and Sartor, G., (2000). Theory Based Dialectics for Resolving Disagreement in Law. ECAI workshop on Computation Dialectics, Berlin, August, 2000.

[6] Bench-Capon, T.J.M., and Sartor, G., (2000). Using Values and Theories to Resolve Disagreement in Law. In J. Breuker, R. Leenes and R. Winkels (eds*), Legal Knowledge and Information Systems: Jurix 2000*. IOS Press, Amsterdam, pp 73-84.

[7] Berman, D.H., and Hafner, C.L., (1993). Representing Teleological Structure in Case Based Reasoning: The Missing Link. In Proceedings of the Fourth International Conference on AI and Law, ACM Press, New York, pp 50-59.

[8] Hage, J.C. (2000). Goal Based Theory Evaluation. In J. Breuker, R. Leenes and R. Winkels (eds*), Legal Knowledge and Information Systems: Jurix 2000*. IOS Press, Amsterdam, pp 59-72.

[9] Levi, E.H., 1948. An Introduction to Legal Reasoning. University of Chicago Press.

[10] Perelman, C., (1980). Justice, Law and Argument. Reidel: Dordrecht.

[11] Prakken, H. & G. Sartor. 1996. A Dialectical Model of Assessing Conflicting Arguments in Legal Reasoning. *Artificial Intelligence and Law* 4: 331-368.

[12] Prakken, H. & G. Sartor. 1998. Modelling Reasoning with Precedents in a Formal Dialogue Game. *Artificial Intelligence and Law* 6: 231-287.

[13] Prakken, H., (2000). An exercise in formalising teleological case based reasoning. In J. Breuker, R. Leenes and R. Winkels (eds*), Legal Knowledge and Information Systems: Jurix 2000*. IOS Press, Amsterdam, pp 49-57.

[14] Roth, B., (2000). New Reasoning Patterns in Analogical Case Based reasoning: An Informal Investigation. In J. Breuker, R. Leenes and R. Winkels (eds*), Legal Knowledge and Information Systems: Jurix 2000*. IOS Press, Amsterdam, pp 113-122.

[15] Sartor, G., (2000). Telelogical Arguments and Theory Based Dialectics. Submitted to *Artificial Intelligence and Law*

[16] Skalak, B.B., and Rissland, E.L., (1991). Arguments and Cases: An Inevitable Intertwining. *Artificial Intelligence and Law*, Vol 1 No 1, 3-44.

[17] Thagard, P. 1992. Conceptual Revolutions. Princeton(NJ): Princeton University Press.