Reasoning With Legal Cases as Theory Construction: Some Experimental Results

Alison Chorley   Trevor Bench-Capon
Department of Computer Science, The University of Liverpool
Liverpool, UK
{*T.J.M.Bench-Capon, A.Chorley*}*@csc.liv.ac.uk*

Abstract. In this paper we report some experiments designed to clarify some issues and to test some of the assumptions in the model of reasoning with legal cases advanced by Bench-Capon and Sartor. We identify the questions to be explored, briefly describe a tool developed to support these experiments and report the results of a series of experiments based on Aleven's analysis of US Trade Secrets cases. We then consider what light the experiments have thrown on our questions, and propose some directions for future work.

1 Introduction

In [3] a model of reasoning with legal cases was proposed in which reasoning with cases is seen as a three stage process:

1. constructing theories to explain past cases in terms of the factors they exhibit;
2. evaluating the competing theories in terms of their coherence, explanatory power, simplicity, etc.
3. applying the chosen theory to a new case.

In [6] and [7] we described a tool designed to support the construction of theories, and some preliminary experiments made using it. In this paper we report on the progress of this work, which embraces some extensions to the tool and some further experiments designed to cast light on a series of specific questions. Section 2 will introduce the questions we address. Section 3 will describe the current version of the tool. Section 4 will describe the experimental set up, and Sections 5 and 6 give the results of the experiments, together with some discussion. Section 7 summarises our conclusions and gives pointers to future work.

2 Questions Addressed

While [3] lays out a general approach to modelling reasoning with legal cases, it leaves a number of points of detail open. This paper is intended to cast light on how these issues might be resolved.

Three issues concern theory construction. First there is the question of how past cases are used in constructing theories, and how the rules of the theory are extracted from these cases. We would hope that there is some systematic approach to theory construction which can be applied, perhaps even automatically. In [6] we describe three approaches, which we call the *safe* approach, the *simple* approach and the *value driven* approach. We have again used these three approaches to construct different theories. Our first question is therefore:

Alison Chorley and Trevor Bench-Capon, 'Reasoning With Legal Cases as Theory Construction: Some Experimental Results' in D. Bourcier (ed.), *Legal Knowledge and Information Systems. Jurix 2003: The Sixteenth Annual Conference*. Amsterdam: IOS Press, 2003, pp. 173-182.

Q1: How should we select cases and extract rules for inclusion in the theory?

Within the general approach it is possible to include or exclude factors that appear in the cases and the domain analysis generally. This gives rise to the second question:

Q2: Should we be inclusive or exclusive with regard to factors?

In [3] preferences between factors determine preferences between values, which can then in turn determine preferences between other factors relating to those values. If this is so, we should expect to be able to include additional factors without major revisions to the theory. Thus we may pose a third question:

Q3: Is there evidence to suggest that values can be used to determine the relative importance of factors?

A key role of the theory is to explain preferences between rules in terms of a comparison between the sets of values promoted by the factors contained in the rules. The fourth question therefore is:

Q4: How should sets of values be compared?

A related question concerns what should be done when a case contains two factors relating to the same value, but favouring different sides. In such cases we must prefer a factor rather than a value. Our fifth question is

Q5: Is it possible to use a general principle to pre-order factors within a value?

In [3] it is assumed that factors promote values to an equal degree: if a rule contains a factor then following that rule is held to promote the value associated with the factor. It is not impossible, however, that different factors will promote values to different degrees, and that this possibility needs to be considered. This leads to the sixth question:

Q6: Is there evidence to suggest that factors promote values to different degrees?

Again, [3] does not allow for accumulation of factors: if following a rule promotes a value, then it is not relevant whether this is the result of the presence of one factor or several. Perhaps, however, a value is promoted to a greater extent if several factors are present. The seventh question is thus:

Q7: Is there evidence to suggest that values and/or factors have a cumulative effect?

We see these questions as central to the effective modelling of reasoning with cases in law in the manner of [3]. Obviously the experiments on a single domain with a limited number of cases described here cannot produce definitive answers to them: our hope, however, is that the experiments may help to move the debate forward, and indicate which lines may be worth pursuing.

3 The Theory Construction Tool

The theory construction tool is a program, written in JAVA, which provides a graphical user interface to support theory construction. The program requires two input files, one recording the cases available for theory construction and their associated factors, and the other associating these factors with their values. Output comes in two forms: one is a file containing the elements that make up a theory as in [3], and the other an executable Prolog program consistent with the theory. The tool supports all the theory constructors specified in [3], and the necessary file management and editing functions. Some more detail and a screen shot is provided in [6]. Since [6] we have given additional options for the way in which the code can be generated. This will be further discussed in section 4. We have also extended the tool to handle dimensions in the manner described in [3].

4 Experiment Design

The domain chosen for our experiments was US Trade Secrets Law as modelled in HYPO [2], CATO [1] and IBP [4]. The domain was chosen because it is well known and much discussed, and because a complete domain analysis together with a reasonable number of cases is available.

First values were associated with the factors identified in [1]. The attribution of values was largely informed by a consideration of the relationship between factors and abstract factors in [1]. The rationale for this is discussed in [6], where a complete list of the factors and their associated values is given. Next we extracted from [2], [1] and [4] a number of cases which we divided in two groups. The first group of 14 cases were to be available for constructing a program: for building the theory and refining the resulting program. The remaining 9 cases were then available to provide an independent test of the worth of the program produced. We would therefore be able to see both whether it is possible to construct a theory to explain a set of cases, and whether the resulting theory has predictive worth. The cases used and their associated factors are shown in Table 1. The factor numbers are as in [1]: the values are as in [6]: CA - that explicit confidentiality agreements should be made and enforced; RE – that a person with a secret should take reasonable efforts to protect it; LM – that a person should be able to develop a product using legitimate means; QM – that a person should not use morally (or legally) dubious means to obtain a secret; and MW – that litigation should only take place if the secret was of some material worth.

To address Q1 we constructed theories according to three different principles. For *Theory 1* we used the cases to generate rules and preferences in the manner of [9]. Each case generates three rules: conjunction of pro-p factors favours p; conjunction of pro-d factors favours d; and a preference between these two rules dependent on the actual outcome. *Theory 2* was supposed to reflect the sort of thinking embodied in rule induction style approaches. Here we chose a minimal set of factors such that at least one pro-plaintiff factor occurs in every case decided for the plaintiff, and at least one pro-defendant factor occurs in every case decided for the defendant. Where both pro-plaintiff and pro-defendant factors were present in a case, a rule preference was included to give the desired outcome. *Theory 3* was based on first determining an order for the values and then selecting cases to ground a theory expressing this ordering. A more extended discussion these three methods can be found in [6].

The programs generated from the theory were then executed on the remaining cases in Group 1. Not all cases were solved correctly: to cater as far as possible for such cases the program was refined. Essentially problems came from one of two sources. Although the theory orders values, and so determines the relative importance of factors relating to different val-

Table 1: Cases Used, Their Factors, Values and Outcomes

	Pro-P factors	Pro-D factors	Pro-P Values	Pro-D Values	Outcome
Group 1					
Arco		10 16 20		RE LM	D
Boeing	4 6 12 14 21	1 10	CA RE QM	RE	P
Bryce	4 6 18 21	1	CA RE MW	RE	P
CollegeWat	15 26	1	LM QM	RE	P
Den-Tal-Ez	4 6 21 26	1	CA RE QM	RE	P
Ecologix	21	1 19 23	CA	CA RE	D
Emery	18 21	10	CA MW	RE	P
Ferranti	2	17 19 20 27	QM	RE LM	D
Robinson	18 26	1 10 19	QM MW	RE	D
Sandlin		1 10 16 19 27		RE LM	D
Sheets	18	19 27	MW	RE	D
Space Aero	8 15 18	1 19	MW LM	RE	P
Televation	6 12 15 18 21	10 16	CA RE LM MW	RE LM	P
Yokana	7	10 16 27	QM	RE LM	D
Group 2					
FMC	4 6 7 12	10 11	CA RE QM	RE LM	P
KG	6 14 15 18 21	16 25	RE QM LM MW CA	LM	P
Mason	6 15 21	1 16	CA RE LM	RE LM	P
MineralDep	18	1 16 25	MW	RE LM	P
NationalRej	7 15 18	10 16 19 27	QM LM MW	RE LM	D
Reinforced	4 6 8 15 21	1	CA RE MW LM	RE	P
Scientology	4 6 12	10 11 20	CA RE	RE LM	D
Technicon	6 12 14 21	10 16 25	RE QM CA	RE LM	P
Trandes	4 6 12	1 10	CA RE	RE	P

ues, when we have factors relating to the *same* value but favouring *different* outcomes, either can be preferred while remaining consistent with the theory. Since the program is generated following an arbitrary selection of the order of such factors, if the choice is wrong we may need to modify the order of these factors within values. Secondly – for theory 1 especially - it is typically the case that there are preferences between values not included in the theory, because they do not emerge from the broadest construal of the rules. Where a case turns on such a preference, it may be necessary to refine the theory by explicitly including this preference. The particular refinements required are described in the next section.

Once the programs had been refined so as to deal, as far as possible, with the cases in the first group, they were tried on the second group of cases, to test the extent to which the theory explaining the “existing” cases, generalised to the “new” cases, with the view that the theory could be considered better according as to how well it generalised to the unseen cases.

To address Q2 we generated two versions of each of the programs. In one version (*selective*) only factors explicitly included during the construction of the theory were used: this is one natural interpretation of [3]. In the other version (*complete*) we included all the factors identified in [1], while still enforcing the value preferences from the theory. The idea here is that since the factors were used in the analysis of [1] they should be presumed to be relevant, and hence included in the theory. Q3 can be answered positively in so far as this extension does not require reordering the values.

Q4 has been the subject of some discussion in [e.g. 8]. Our initial version simply ranked factors and their combinations in order as determined by values. Effectively this method weights a set of values according to its most highly weighted value, unless a particular set of values has been treated explicitly. Where a case contains the same value as the most highly rated for both the pro-plaintiff and pro-defendant sets, the outcome is determined by preferring factors. Much of the refinement process may be seen as tuning this ordering of factors with common values. A second possibility, following [8] is to discount values found in both the pro-plaintiff and pro-defendant sets, giving the set a weight based on its most weighty value not in the other set. This “cancellation” process is produced by incorporating the absence of factors into the Prolog rules generated. A third possibility is to consider all the values in a set by assigning all its members a weight relative to the importance of the value to which they relate. There are two possibilities here, according to whether we count only the values represented, or increase the weight according to how many factors representing the values are present (that is, we use “bags” of values rather than sets). To explore this possibility we must assign weights to values. To examine Q5 we used two principles to pre-order values. In order to explore Q6, we had to go further, assigning differential weights to factors with the same values. Q7 is addressed by whether the use of sets or bags gives better results.

5 Experimental Results: Boolean Values

The initial experiments were set up as described in [6]. For the Theory 1, 4 cases were used, 2 plaintiff cases, 2 defendant cases (We found that using more cases to construct the initial theory typically led to less refinement, but had no effect on ultimate performance). For Theory 2 there were only 2 cases with a conflict to resolve so those are the only cases used to build the theory. For Theory 3 we ranked the values as CA>LM, LM>RE, RE>MW and RE>QM and selected 4 cases which could be used to express these preferences.

Once the theory has been constructed the Prolog code associated with it can be generated in several different ways.

The *original* method of code generation just sorts the rules present in the theory according to the value order and rewrites the rules as Prolog clauses.

The *cancellation* method creates rules of the form *f1 and not f2 and not f3... and not fn* where f2..fn are factors which relate to the same value as f1 but favour the opposite side.

For Q5 we used two approaches to pre-ordering values, one based on the principle that exceptions should carry more weight, the other using information found in the CATO factor hierarchy. The *exception* method creates more rule preferences by ordering the factors within each value. This is done by counting the factors that relate to each outcome. The factors promoting the less common outcome are taken to be exceptions and these are preferred to the other factors. The *exception and cancellation* method uses the exception method to sort the rules and the cancellation method to create the Prolog code. The *CATO* method also sorts the factors within each value. The factor hierarchy from [1] contains thin and thick arcs and these are used to represent whether the factor is strong or weak. The strong factors within each value are preferred to the weak factors. The *CATO cancellation* method sorts the rules using the CATO arc method and creates the Prolog code using the cancellation method. For all of the above methods the first rule to match each case fires and gives the case an outcome.

There were no refinements to be made to the selective versions of Theory 2 and Theory 3. For the selective version of Theory 1 all the methods for code generation needed the additional preference LM > RE and the *original* and *CATO* methods also required CA > RE. For Theory 1 with all factors, the *original*, *cancellation* and *CATO* methods for code generation could all be refined to improve the results for the main cases. The refinements for the *original* method are to set LM>RE, CA>LM (to handle Space Aero) and F23>F21 (to handle Ecologix); the refinement for the *cancellation* method is to set LM>RE and the refinement for the *CATO* method is to enforce F23>F21.

For Theory 2 with all factors the *original*, *CATO* and *CATO can*cellation methods for code generation can be refined to improve the results for the Group 1 cases. The refinement for the *original* method is to make F6>F10; the refinements for the *CATO* method is set F6>F1 and RE>QM; and the refinement for the *CATO cancellation* method to ensure F6>F10.

For the complete Theory 3 the *original* and *CATO* methods for code generation can be refined to improve the results for the main cases. The refinement for both of the these methods is to make F23>F21.

The number of errors for each method of code generation before and after refinement is shown in table 2. The results are split into the errors for the Group 1 cases and the errors for the Group 2 cases. Where there are two values in a cell the first value is the number of errors and the second is the number of undecided cases. The case will be undecided when there is not enough information in the theory to decide the outcome of a case. For Theories 1 and 3 the undecided cases are FMC, Scientology and Trandes, for Theory 2 the undecided cases are FMC, Mineral Deposits and Trandes.

Table 2: Misclassified Cases and Abstentions

		Theory 1				Theory 2				Theory 3			
		Selective		Complete		Selective		Complete		Selective		Complete	
	Cases	Before	After	Before	After	Before	After	Before	After	Before	After	Before	After
Original	Grp 1	3	1	3	0	0	0	4	1	1	1	1	0
	Grp 2	6/0	3/0	6	2	1/3	1/3	5	4	2/3	2/3	3	3
Cancel	Grp 1	1	0	1	0	0	0	1	1	0	0	0	0
	Grp 2	1/3	1/3	2	3	0/3	0/3	3	3	1/3	1/3	2	2
Except	Grp 1	3	1	0	0	0	0	1	1	1	1	0	0
	Grp 2	6/0	3/0	2	2	2/3	2/3	4	4	2/3	2/3	3	3
Except/ cancel	Grp 1	1	0	0	0	0	0	1	1	0	0	0	0
	Grp 2	1/3	1/3	3	3	0/3	0/3	3	3	1/3	1/3	2	2
CATO	Grp 1	3	1	1	0	0	0	3	1	1	1	1	0
	Grp 2	6/0	3/0	3	3	1/3	1/3	4	4	2/3	2/3	3	3
CATO/ cancel	Grp 1	1	0	0	0	0	0	2	1	0	0	0	0
	Grp 2	1/3	1/3	2	2	0/3	0/3	2	3	1/3	1/3	2	2

From a consideration of the results for the Group 1 cases, we can see that it is possible to develop a theory which will cover all the cases using the approach of either Theory 1 or Theory 3. Theory 2, however, cannot be refined to deal with Emery as it begins by ascribing undue importance to the value RE. We see this as an example of the over fitting that can occur using this sort of method. Thus as regards Q1, we feel that the approach of Theory 2 is inferior because of this danger. As regards Q2 we conclude that it is preferable to use the complete set of factors: Theory 1 cannot handle Ecologix on the original selection of factors, and there are many errors on the restricted set of factors for the Group 2 cases. Theory 2 does no worse on the full set of factors, and Theory 3 improves. With regard to Q3, it does seem that the use of the value order derived from the limited set of factors can be generalized to the full set of factors. The refinements required additional value preferences, clarifying the partial order, rather than overturning any existing preferences. Thus we conclude that using either the method of Theory 1 or Theory 3, and a full set of factors, it is possible to construct a theory in the manner of [3] which will explain a given body of cases.

The worth of the theory as an explanation of unconsidered cases is shown by looking at the results for Group 2. Here the results are a little disappointing, in that we are never able to achieve better than 7 out of the 9 cases correct. Note, however, that Theory 2 performs worse than the other two, again indicating the problem of over fitting.

The cases that give problems differ for Theory 1 and Theory 3. Theory 1 tends to fail always on Mineral Deposits and Scientology without cancellation, and FMC and Technicon when cancellation is used. Theory 3 fails on Mineral Deposits and Scientology and also, without cancellation, National Rejectors. Mineral Deposits is something of a special case: its outcome cannot be explained on the basis of the factors used in CATO, but relies on some special circumstances [5][1]. Again for Q1 there does not seem much to choose between the approach of Theory 1 and Theory 3. With regard to Q2, we may note that if we consider abstentions and mistakes together, we always obtain better performance by using all the factors. With regard to Q3, we feel that the use of values to determine the significance of unconsidered factors is vindicated, since whilst we may need to extend the partial order, we never need to change it. Note also that the unrefined theory never generalised more successfully than the refined theory.

For Q4, the results from Theory 1 and Theory 3 for the versions with all factors seem to give conflicting results, in that Theory 1 generally works better without cancellation, and Theory 3 with cancellation. This can be explained by reference to the particular cases. Theory 1 encounters problems using cancellation when F12 and F10 co-occur (that is when the secret was disclosed to outsiders, but under restrictions). That such disclosures were subject to restrictions seems to be a strong point in favour of the claimant, but this strength is lost if the factors cancel each other out. Given that Theory 1 places stress on the value related to these factors, this results in the misclassification of such cases. Theory 3 benefits from cancellation since the two factors related to the value it typically stresses are not the strongest factors in National Rejectors, which is misclassified without cancellation: thus its ability to ignore them through cancellation is helpful. There is not really sufficient data here to come to any firm conclusions: however, we would suggest that cancellation is quite possibly undesirable, since the *combination* of F12 and F10 appears to improve the plaintiff's case, and the weakness of the pro-plaintiff factor F15 in National Rejectors might be better handled in some other way rather than hoping that there will be a cancelling factor also present. We feel that the approach of the next section provides a more promising line of attack.

For Q5, we should expect to find less refinement needed in the two approaches using pre-ordering on values. This is indeed the case: only once is refinement needed when we pre-order the factors within a value, whereas refinement is always required when we do not. The use of exceptions to pre-order looks promising, but the dataset is too small to say that it will always obviate the need for refinement. The pre-ordering does not, however, carry any gains for the ultimate performance: it merely simplifies theory construction.

6 Experimental Results: Numeric Values

In order to explore Q6 and Q7, we must again modify the code generation, and the execution of the Prolog program. Instead of relying on the order of the execution of the rules to enforce priorities, we attach numerical weights to each factor, and consider every factor present in order to determine the relative strength of the plaintiff's and defendant's cases. The size of the numbers produced may be an indicator of confidence in the predicted decision, but we lay no stress on this here.

The numeric method assigns weights to each value and hence to each factor within the value. The weight is positive for a plaintiff factor and negative for a defendant factor. The weight for each value is decided by the value preferences, a value which is preferred being given a larger weight than the value to which it is preferred. Three approaches are used:

- The first approach just counts *the values* present in each case. If the value is associated with a plaintiff factor then the weight is positive, otherwise the weight is negative. This will allow for the cumulative effect of values: several lesser values may together outweigh

[1]This could, of course, be explained by including an additional factor to represent these special circumstances, but it is probably dubious practice to include a factor applicable to a single case.

a stronger one. This addresses Q4 in that it allows all the members of a set to be considered, rather than just the strongest member. Note that this effectively incorporates the cancellation method above, since if factors favouring both sides are present the positive and negative values will cancel out.

- For the second approach *all the factors* present in the case are counted. Again the weights for the plaintiff factors are positive and negative for the defendant factors. In this approach several factors for the same value will have a cumulative effect, addressing Q7.
- The third approach gives a basic weight to the value which is then modified by the factor depending how much the factor promotes the value. For a strong factor the weight is multiplied by 10, for a weak factor the weight is divided by 10 and for a normal factor the weight is unchanged. This is to allow for different degrees to which a factor promotes a value, addressing Q6. Two methods of identifying strong and weak factors were used. In one case we used the "KO" (i.e knockout) and weak factors of [4] and in the other we gave strong weight to the factors we identified as "exceptions", as above.

The results for just counting the value weights and counting all the factors is shown in Table 3 and the results for using differential values in Table 4. Misclassifications in both tables are shown in bold.

If we compare the theories using all factors, we can see that for Theory 1, the cumulative value approach is superior to any of the approaches previously tried, failing only on the problematic Mineral Deposits. For Theory 3, there are two errors, as with the best previously seen. On this basis we may feel that Q4 is answered by the suggestion that all the values in the set should be considered. With regard to Q7, however, there seems no gain in allowing factors to have a cumulative effect: Theory 1 has an additional error, and Theory 3 also does not improve.

Turning to Q7, we can see that using the weights from [4] we have a problem with Space Aero. This is because of the presence of two opposing KO factors. Except for this problem, using these differential values seems to improve the situation, especially when compared to cumulative factors. This is to be expected perhaps: without differentiation, accumulating weak factors can give them too much importance. This is also indicated by the comparatively poorer performance of the exception method which does not identify weak factors, with consequent misclassification of National Rejectors. (It also fails to pick out F20, Information Known to Competitors, as a strong factor and so misclassifies Scientology, which relied on this factor).

We tentatively conclude therefore that values should be considered cumulative, and that if factors are to be accumulated they should be given different weights, which should identify both strong and weak factors.

7 Summary of Results

With regard to our seven questions we have come to the following tentative conclusions. These conclusions are tentative because our dataset is not large enough to allow tests for statistical significance. On the other hand, the small set of cases allow for detailed consideration of aberrant cases, which gives insight into the problem. We do, therefore, believe our results to be indicative.

Q1: We feel that the approach of Theory 2 is not suited as it overfits the data. The approaches of Theory 1 and Theory 3 both have some good and some bad aspects.

Table 3: Cumulative Values and Cumulative Factors

	Value Results						Factor Results					
Theories	t1s	t1c	t2s	t2c	t3s	t3c	t1s	t1c	t2s	t2c	t3s	t3c
Arco	-0.3	-1	-1.5	-2.2	-0.7	-1	-0.3	-1.7	-1.5	-3.7	-0.7	-1.7
Boeing	0.4	0.8	0.3	**-0.3**	1.5	1.6	0.1	1.5	0.3	0.7	1.5	3.1
Bryce	0.5	0.8	0.3	0.4	1.6	1.6	0.5	1.5	0.3	0.7	1.6	3.1
CollegeWat	0.5	0.5	1.5	0.9	0.8	0.5	0.5	0.5	1.5	0.9	0.8	0.5
Den-Tal-Ez	0.5	0.8	0.3	0.4	1.6	1.6	0.5	1.5	0.3	0.7	1.6	3.1
Ecologix	**0.4**	-0.3	-0.4	-0.7	**1.2**	-0.3	**0.1**	-0.6	-0.4	-1.4	**1.2**	-0.6
Emery	0.5	0.5	0.3	**-0.3**	1.6	1.3	0.5	0.5	0.3	**-0.3**	1.6	1.3
Ferranti	-0.3	-0.9	-2.2	-2.1	-0.3	-0.9	-0.6	-1.9	-2.9	-4.3	-0.3	-1.9
Robinson	-0.1	-0.1	-0.7	-0.5	-0.1	-0.1	-0.7	-0.7	-0.7	-1.9	-0.1	-0.7
Sandlin	-0.3	-1	-0.7	-2.2	-1	-1	-1.2	-1.9	-1.4	-4.3	-1	-1.9
Sheets	-0.2	-0.2	-0.7	-0.6	-0.2	-0.2	-0.5	-0.5	-1.4	-1.3	-0.2	-0.5
Space aero	0.5	0.5	0.8	0.9	0.5	0.5	0.2	0.3	0.8	0.3	0.5	0.3
Televation	1.2	0.8	1.8	0.4	1.6	1.6	1.2	1.1	1.8	1.1	1.6	1.9
Yokana	-0.3	-0.9	-0.7	-2.1	-0.7	-0.9	-0.6	-1.2	-0.7	-2.8	-0.7	-1.2
FMC	**-0.3**	0.1	0	**-1.1**	0	0.9	**-0.3**	0.4	0	**-0.4**	0	1.2
KG	1.5	1.2	1.8	1.2	1.6	2	1.5	0.5	1.8	**-0.3**	1.6	1.3
Mason	1.1	0.7	1.8	0.3	1.5	1.5	1.1	0.7	1.8	0.3	1.5	1.5
MineralDep	**-0.2**	**-0.9**	0	**-2.1**	**-0.6**	**-0.9**	**-0.2**	**-1.6**	0	**-3.6**	**-0.6**	**-1.6**
NationalRej	**0.5**	-0.1	**0.8**	-0.5	-0.2	-0.1	-0.1	-0.7	**0.1**	-1.9	-0.2	-0.7
Reinforced	1.1	1.5	1.8	1.9	2.2	2.3	1.1	2.2	1.8	2.2	2.2	3.8
Scientology	-0.3	0	-1.5	-1.2	0	**0.8**	-0.3	-0.4	-1.5	-2	0	**0.4**
Technicon	0.4	0.1	0.3	**-1.1**	0.8	0.9	0.4	**-0.3**	0.3	**-1.9**	0.8	0.5
Trandes	**-0.3**	0.7	0	0.3	0	1.5	**-0.6**	0.7	0	0.3	0	1.5

Table 4: Differential Factor Weights

	Weights from Brüninghaus and Ashley						Weights from Exceptions					
	t1s	t1c	t2s	t2c	t3s	t3c	t1s	t1c	t2s	t2c	t3s	t3c
Arco	-0.03	-7.1	-15	-15.22	-0.07	-7.1	-0.3	-1.7	-1.5	-3.7	-0.7	-1.7
Boeing	0.64	2.04	0.3	1.96	1.5	3.64	0.1	6.9	0.3	13.3	1.5	8.5
Bryce	0.77	1.77	0.3	1.33	1.6	3.37	0.5	4.2	0.3	7	1.6	5.8
CollegeWat	1.67	1.67	1.5	2.43	1.7	1.67	6.8	6.8	15	14.4	7.1	6.8
Den-Tal-Ez	1.67	2.67	0.3	2.23	2.5	4.27	0.5	4.2	0.3	7	1.6	5.8
Ecologix	-2.33	-3.03	-6.7	-7.07	-1.5	-3.03	**0.1**	-6.9	-0.4	-4.1	**1.2**	-14.1
Emery	0.77	0.77	0.3	0.33	1.6	1.57	0.5	0.5	0.3	**-0.3**	1.6	1.3
Ferranti	-6	-13.6	-29	-30.4	-3	-13.6	-0.6	-1.9	-2.9	-4.3	-0.3	-1.9
Robinson	-1.96	-1.96	-7	-6.04	-1.9	-1.96	-0.7	-0.7	-0.7	-1.9	-0.1	-0.7
Sandlin	-6.06	-6.13	-14	-14.29	-3.07	-6.13	-1.2	-1.9	-1.4	-4.3	-1	-1.9
Sheets	-5.9	-5.9	-14	-13.9	-2.9	-5.9	-0.5	-0.5	-1.4	-1.3	-0.2	-0.5
Space aero	**-2.23**	**-1.23**	**-5.5**	**-4.47**	**-2.2**	**-1.23**	6.5	6.6	14.3	13.8	6.8	6.6
Televation	1.47	2	1.8	3.08	2.23	2.8	7.5	12.8	15.3	27.2	7.9	13.6
Yokana	-3.03	-3	-7	-7.12	-0.07	-3	-0.6	-1.2	-0.7	-2.8	-0.7	-1.2
FMC	**-0.03**	0.67	0	0.23	0	1.47	**-0.3**	5.8	0	12.2	0	6.6
KG	1.5	1.13	1.8	1.05	2.23	1.93	7.8	9.5	15.3	19.5	7.9	10.3
Mason	1.37	1.6	1.8	2.28	2.13	2.4	7.4	9.7	15.3	20.1	7.8	10.5
MineralDep	0.07	**-0.7**	0	**-1.62**	0.03	**-0.7**	**-0.2**	**-1.6**	0	**-3.6**	**-0.6**	**-1.6**
NationalRej	-5.23	-5.2	-12.5	-12.52	-2.27	-5.2	**6.2**	**5.6**	**13.6**	**11.6**	**6.1**	**5.6**
Reinforced	1.37	3.37	1.8	3.73	2.2	4.97	7.4	11.2	15.3	22	8.5	12.8
Scientology	-0.03	-6.43	-15	-14.87	0	-5.63	-0.3	**5**	-1.5	**10.6**	0	**5.8**
Technicon	0.67	0.6	0.3	0.08	1.43	1.4	0.4	5.1	0.3	10.7	0.8	5.9
Trandes	**-0.06**	1.24	0	1.56	0	2.04	**-0.6**	6.1	0	12.9	0	6.9

Q2: Should we include all factors? We feel that this is preferable: the relevance of our factors has already been established by the analysis of [1], and further selection does ignore relevant factors. Note, however, that considering only the factors appearing in a selection of cases does succeed in establishing a value order.

Q3: No especial problems were encountered in using all factors according to the ordering of values established in a selected few cases: this is consistent with the assumption of [3] that the importance of a factor is determined by the value it promotes.

Q4: We feel that there is evidence to suggest that all values in a set should be considered when comparing sets.

Q5: We feel that pre-ordering factors within a value saves a considerable amount of refinement, simplifying theory construction. Exceptions can help as a general principle, but perform less well that the expert judgement of [4].

Q6: We feel that there is evidence, also suggested by [4], that factors promote values to differing degrees. This becomes important if it is wished to ascribe a cumulative effect to factors

Q7: While, as in Q4, we feel it is important to give values a cumulative effect, the case for giving factors a cumulative effect seems less persuasive.

For future work we intend to attempt to confirm the tentative conclusions given above by a trial on a larger data set, either in this domain or some other, which will enable some measure of significance to be attached to the findings. Additionally, we have already extended the theory construction tool to handle dimensions as described in [3]. We intend to perform some experiments to explore the benefits of operating at this level. In particular we feel that this may give some insight into identifying strong and weak factors.

References

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

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

[3] Trevor Bench-Capon and Giovanni Sartor, A model of legal reasoning with cases incorporating theories and values. *Artificial Intelligence*. In Press.

[4] Stefanie Brüninghaus and Kevin D. Ashley. Predicting Outcomes of Case-based Legal Arguments. Proceedings of the Ninth International Conference on AI and Law: ACM Press, New York, 2003, pp233-42.

[5] Stefanie Brüninghaus. Private Communication. September 2003.

[6] Alison Chorley and Trevor Bench-Capon, Developing Legal Knowledge Based Systems Through Theory Construction. Technical Report ULCS-03-013, Department of Computer Science, The University of Liverpool, 2003. http://www.csc.liv.ac.uk/research/techreports/tr2003/ulcs-03-013.pdf

[7] Alison Chorley and Trevor Bench-Capon, Developing Legal Knowledge Based Systems Through Theory Construction. Proceedings of the Ninth International Conference on AI and Law: ACM Press, New York, 2002, pp85-6.

[8] 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.

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