A Quantitative Approach to Theory Coherence

Trevor J.M. Bench-Capon
Department of Computer Science
The University of Liverpool
Liverpool
UK
t.j.m.bench-capon@csc.liv.ac.uk

Giovanni Sartor
Department of Law
The Queen's University of Belfast
Belfast
Northern Ireland
sartor@cirfid.unibo.it

Abstract. In this paper we describe an approach which can attempt to give a quantitative measure of the coherence of legal theories. Starting from the notion of a theory as described in [3], we show how some ideas of Thagard can be adapted to give a measure of how coherent such a theory is. Some experiments, using the wild animals domain and theories concerning it discussed in [3] are described. We then identify some issues, both of detail and principle, that are raised by this work.

1 Introduction

In [3] we articulated a view of reasoning with cases as a process of theory construction, evaluation and application. On this view, reasoning with cases involves the construction of competing theories to explain the body of decided cases, evaluating these theories to select the best, and then applying that theory to new cases. That paper focussed primarily on the components of theories and how they could be constructed, but we did offer also a brief discussion of what made one theory better than another, detailing several aspects which together make up the *coherence* of a theory. That discussion, however, used only qualitative considerations. In this paper we will explore an approach to providing a quantitative measure of theory coherence.

Our approach is inspired by the work of [5]. In that book Thagard attempts to account for concept change in science in terms of more coherent theories replacing less coherent ones. His essential idea is to represent the evidence to be accounted for by a theory and the tenets of a theory as nodes connected by links representing support and conflict. A set of initial values (between 1 and -1) is assigned to these nodes, and these values are then propagated, support links increasing the values of nodes, and conflict links decreasing them. Moreover, links are subject to a rate of decay so that isolated nodes decrease in value. This propagation is continued through a number of cycles, until the values of the nodes stabilise. In Thagard's interpretation of this process, nodes which end with a high activation can be considered part of a coherent, and hence acceptable theory, while those with a low activation do not form part, and so should be rejected. In this paper we will follow a slightly different approach, which is

Trevor J.M. Bench-Capon and Giovanni Sartor, 'A Quantitative Approach to Theory Coherence' in Bart Verheij, Arno R. Lodder, Ronald P. Loui and Antoinette J. Muntjewerff (eds.), *Legal Knowledge and Information Systems. Jurix 2001: The Fourteenth Annual Conference*. Amsterdam: IOS Press, 2001, pp. 53-62.

more in line with our adversarial view of legal argument. Our purpose is not to extract the best theory out of the body of existing evidence and hypotheses, but to measure the comparative coherence of competing theories. Therefore, we do not see nodes in a theory having a low level of activation as rejected: they remain in the theory but lower its overall coherence.

We will adapt this idea to legal theories as described in [3]. In section 2 we will recapitulate some of the essential ideas of that paper, the example cases used and some of the example theories concerning them given there. Section 3 will give details of the nodes links that need to be included to provide a graph structure representing a theory, and the program used to propagate support and conflict through these nodes and links. Section 4 will describe the behaviour of the program with respect to the theories given in section 2. This work throws up a numbers of issues of detail and principle. These will be discussed in the concluding section.

2 The Example Domain

The example used in [3] was based on three example cases concerning the pursuit of wild animals used in [4]. In all three 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 δr not. In the first case, *Pierson v Post*, π was hunting a fox on open land in the traditional manner using horse and hound when δ killed and carried off the fox. In this case 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, shooting them, and selling them for food. Out of malice δ used guns to scare the ducks away from the pond. Here π won. In the 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 on.

A theory can be seen [3] as consisting of five elements; a set of cases, a set of factors (in the sense used in CATO [1]) used to describe those cases, a set of defeasible rules relating those factors to case outcomes, a set of preferences between the rules, and a set of preferences between values justifying the rule preferences.

We used four factors in the construction of our theories:

- π*Liv* = π was pursuing his livelihood (Keeble, Young), favouring Π
- π*Land* = π was on his own land (Keeble), favouring Π
- π*Nposs* = π was not in possession of the animal (Pierson, Keeble and Young), favouring Δ
- δ*Liv* = δ was pursuing his livelihood (Young), favouring Δ

The factors were based of three values:

Llit = Less Litigation, promoted by π*Nposs*
Prop = Enjoyment of property rights, promoted by π*Land*
Mprod = More productivity, promoted by π*Liv* and δ*Liv.*

On this basis we constructed a number of possible theories to explain the domain. The assumption here is that *Young* is as yet undecided, so we are trying to explain different outcomes for *Young*. Four of these theories are summarised in Table 3. The outcome for Young is shown in brackets in the cases column; arbitrary preferences are marked with an asterisk.

Our assessment of the theories was that T4 was best, followed by T3, which makes arbitrary assumptions, followed by T2, which is unsafe because it includes too few factors, followed by T1, which includes too few cases. This qualitative assessment will be used as the basis for discussion of the quantitative assessments in the experiments below.

3 Implementation of the Approach

In order to apply Thagard's method, we need to represent theories as graphs. The nodes of these graphs will correspond to certain elements of theories, cases, rules, preferences between rules, and preferences between values. Factors and values themselves are not included in the graphs, since they are intended to be not the propositions of a theory, but rather the vocabulary out of which the theory is constructed.

Table 1: Summary of Theories in [3]

Theory	Cases	Factors	Rules	Rule Prefs	Value Prefs
T1	Pierson Young (d)	π*Nposs*	r1: π*Nposs*$\rightarrow \Delta$		
T2	Pierson Keeble Young (p)	π*Nposs* π*Liv*	r1: π*Nposs*$\rightarrow \Delta$ r2: π*Liv* $\rightarrow$ Π	r2 > r1	Mprod> LLit
T3	Pierson Keeble Young (d)	π*Nposs* π*Liv* π*Land*	r1: π*Nposs*$\rightarrow \Delta$ r2: π*Liv* $\rightarrow$ Π r3: π*Land* $\rightarrow$ Π r4: π*Liv* & π*Land* $\rightarrow$ Π	r4 > r1 r1 > r2*	Llit > Mprod* Mprod + Msec > LLit
T4	Pierson Keeble Young (d)	π*Nposs* π*Liv* δLiv	r1: π*Nposs*$\rightarrow \Delta$ r2: π*Liv* $\rightarrow$ Π r5: δLiv $\rightarrow \Delta$ r6: π*Nposs* & δLiv $\rightarrow \Delta$	r2 > r1 r6 > r2	Llit > Mprod Mprod + Llit > Mprod

The nodes will be connected according to whether they are used together in theory construction or application. Rules are constructed from, and applied in, cases. Rule preferences are constructed from cases and used to defeat the application of rules in particular cases. Value preferences are derived from, and allow the derivation of rule preferences. All these therefore represent connections.

We also need to recognise that connections may have different strengths. First, Thagard suggests that conflicting links should be given a higher strength than supporting links [5], if convergence is to be achieved. Second, if a factor contains several factors for the same side, it could support several rules. For example, if a case contains factors A and B and was decided for the plaintiff, then it could be explained by any of the three rules A$\rightarrow$P, B $\rightarrow$ P or A&B$\rightarrow$P. Clearly the case will give full support to the inclusive rule, which is entailed by the other two rules, but its support for the two primitive rules must be divided between them. In all these cases, support is considered bi-directional. Finally we note that the rule A$\rightarrow$P

logically implies A&B→P. Therefore there should be a very strong connection between them, but running only in that direction.

The program was implemented in two parts. One part holds a model of a particular theory, describing the nodes, the connections between them and the strength of these connections. Two lists of connections are maintained for each node, one containing the positive connections, and one containing the hostile connections. Note that in the case of a hostile connection between a rule and a case, the connection can itself be in conflict with a rule preference. These connections need therefore to be also modelled as nodes, supported by their rule, and attacked by the preference which defeats them. Again these links are uni-directional. The second part was a generic program to propagate values through these models. The algorithm is fairly straightforward. Its psuedo-code is as follows:

```
For each node
    find current node value
    set temp node value := node value
    find connections
    apply decay factor
    for each supporting node
      temp node value := temp node value +
            connected node value * strength
    for each hostile node
      temp node value := temp node value -
            connected node value * strength
    if temp node value > 1 then temp node value := 1
    if temp node value ¡ -1 then temp node value := -1
    if node is hostile link and temp node value ¡ 0 then
            temp node value = 0
    store temp node value
for each node
    node value := temp node value
```

Note that the values of hostile links can never drop below 0; a defeated rule cannot support a case.

The above algorithm was embedded in a simple interface, which after each cycle displays the node values, and executes another cycle or halts as the user requires. A history list of all node values is maintained. After the last cycle a number of statistics such as the average node value is produced.

4 Experimental Results

The program was executed on four models of theories, representing the four theories in Table 3.

We needed to choose a number of specific values to represent connection strengths. The values used are shown in Table 2.

In all cases, except where mentioned below, nodes representing decided cases were initialised to 1, and all other nodes to 0.

Table 2: Parameters Used in the Experiments

Parameter	Value
Decay factor	0.1
Normal Support Link	0.4
Entailment support Link	0.8
Conflict Link	0.6

4.1 Theory 1

This is the simplest theory comprising only two cases and a single rule. The connections and their initial weights are shown in Table 3:

Table 3: Connections in Theory 1

Nodes	Pierson	Young	r1
Pierson			0.4
Young			0.4
r1	0.4	0.4	

Table 3 shows the initial weights of the connections in Theory 1. Each row shows the connections from the node in the left hand column from the nodes in the column header. An empty cell indicates no connection. Tables 4-6 provide similar information for theories 2-4 respectively.

Executing this model results in all 3 nodes attaining the value 1 in seven cycles. This suggests that the theory is entirely coherent. We should expect this, since the problem with this theory is not in how well it hangs together, but that is does not account for some of the cases.

4.2 Theory 2

Theory 2 introduces the third case, and an additional rule. There is now a conflict between the rules which is resolved by a rule preference, and a corresponding value preference.

The connections, and their initial values is shown in Table 4

This model converges after 10 cycles. All the nodes achieve 1, except the hostile connections between r1 and Keeble and Young, resulting in an average activation of 0.8. Essentially the theory is can be seen as coherent, but requiring a rule to be defeated in two of its applications.

4.3 Theory 3

Theory 3 improves on Theory 2 by including an extra factor, and hence extra rules. It contains 3 cases, 4 rules, 2 rule preferences (one of which is arbitrary) and two value preferences (again, one arbitrary). The connections and their initial weights in theory 3 are shown in

Table 4: Connections in Theory 2

Nodes	Pierson	Keeble	Young	r1	r2	r2>r1	Mprod > Llit	r1→ -Keeble	r1→ -Young
Pierson				0.4					
Keeble				0.6	0.4	0.4			
Young				0.6	0.4				
r1	0.4	0.6	0.6						
r2		0.4	0.4						
r2>r1		0.4					0.4	0.6	0.6
Mprod > Llit						0.4			
r1→ -Keeble						0.6			
r1→ -Young						0.6			

Table 5. Note here we have some uni-directional columns: thus, for example, a connection runes from r2 to r4, but not *vice versa*.

The execution of this model converges in 11 cycles. 3 nodes, the two arbitrary preferences and Young, end as -1. Two nodes, the attack of r1 on Keeble and attack of the arbitrary preference on the attack of r2 on Young become 0. All others become 1. The overall activation is 0.5.

From this we can see that the resort to arbitrary preferences has made the theory much less coherent than its predecessors. Indeed the preferred nodes of the theory after propagation suggest a victory for the plaintiff in Young.

Table 5: Connections in Theory 3

Nodes	Pie	Kee	You	r1	r2	r3	r4	r4 > r1	r1 > r2	L > Mp	Mp + Ms > L
Pie				0.4							
Kee				0.6	0.2	0.2	0.4	0.4			
You				0.4	0.6						
r1	0.4		0.4								
r2		0.4									
r3		0.4									
r4		0.4			0.8	0.8					
r4 > r1		0.4									0.4
r1 > r2										0.4	
L > Mp									0.4		
MP +Ms > LL		0.4						0.4			
r1→ -Kee								0.6			
r2→ -You									0.6		

4.4 Theory 4

The final, and most elaborate theory is Theory 4. This theory contains 3 cases, 4 rules, two preferences and two value preferences. In this case the value preference is not arbitrary, but

rather is taken as axiomatic, since the less preferred value is a proper subset of the preferred value. This time, therefore, this value preference will be initialised as 1. The connections and their initial weights are shown in Table 6.

If we run this model it converges after 8 cycles, with an average activation of only 0.47. In this case two nodes, those representing the defeated applications of r1 to Keeble and r2 to Young go to zero. The rule preferences are enough to defeat these. However, r5 , r6 and Young also go to -1, which is not what we anticipated. The reason for this is that the decay factor is such that r1 does not grow sufficiently quickly to counter the negative impact of the decay on Young, on r5 and r6. Because there is no independent evidence for r5, its impact on r6 and Young is rather negative. This could be interpreted as suggesting that a factor introduced solely to explain the current case is detrimental to the theory. (An alternative suggestion would be that values should be included in the model to support the introduction of new factors, but this would require a different approach to representing the theories as graphs from that taken in this paper).

The crucial element in producing this behaviour is the decay factor. This can be seen as representing the degree of intrinsic believability or otherwise of propositions. The higher the decay factor the closer must be the connection of a proposition with some evidence, otherwise the scepticism represented by the decay factor will overcome the distant evidence.

Table 6: Connections in Theory 4

Nodes	Pie	Kee	You	r1	r2	r5	r6	r2 > r1	r6 > r2	Mp > L	Mp+L > Mp
Pie				0.4							
Kee				0.6	0.4			0.4			
You				0.2	0.6	0.2	0.4				
r1	0.4		0.4								
r2		0.4									
r5			0.4								
r6			0.4	0.8		0.8					
r2 > r1		0.4								0.4	
r6 > r2											0.4
L > Mp		0.4						0.4			
MP +Ms > LL									0.4		
r1→ -Kee								0.6			
r2→ -You									0.6		

Suppose, however, that we are somewhat less sceptical, and more prepared to take beliefs on trust. This would mean a smaller decay factor. If we reduce the decay factor to 0.05, we find that all of Young, r5 and r6 go to 1 after 18 cycles, with an average activation of 0.87. This highlights the significance of the precise choice of parameters, but given this choice this theory comes out as superior to theories 2 and 3.

4.5 A Final Experiment

For our last experiment we constructed a model which contained essentially the union of all the preceding theories. As well as Keeble and Pierson it contained two versions of Young,

one with the outcome π and one with the outcome δ. All six rules were included, together with the possible preferences between them, and the various value preferences which would licence them. All rules were connected to the cases according to their applicability to the cases.

In this model cases were set to 1, and value preferences to 1 if the lesser value was a proper subset of the preferred value, to -1 if the preferred value was a proper subset of the lesser value, and 0 otherwise. Connections were given values as in Table 2.

Obviously the theory represented by this model is *not* coherent. The idea here was to step through the model to see which cases and rules were believed and which disbelieved. In some ways this is closer to Thagrard's original use of the technique, in which he always represented competing theories in a single model.

Two runs were performed, one with a decay factor of 0.1 and one with a decay factor of 0.5.

The results for the cases and rules are given in Table 7.

Table 7: Results for Combined Model

Node	decay = 0.1: 15 iterations	decay = 0.05: 15 iterations
Pierson	1	1
Keeble	1	1
Young(π)	-1	-1
Young(δ)	-1	1
r1	0.33	1
r2	1	1
r3	-1	-1
r4	-1	1
r5	1	1
r6	-1	1

Consider first the sceptical version. Here we end up disbelieving both outcomes for Young. Moreover, the disbelief in Young extends itself to the rules which could establish the conclusion. Doubt is even present in case of the perfectly well founded r1, despite the presence of Pierson. Note also that it is r2, that the land is owned which is held to explain Pierson; r3 is doubted by its contamination from Young.

The less sceptical version gives a theory we might prefer. Now we believe the outcome δ or Young. Note, however, that again r2 is the key element in Keeble: r3, the plaintiff earning his livelihood, is not considered sufficient to explain Keeble for π. Is this reasonable? It probably does go against the analysis of [4], but does accord with some readings of the original decisions (see [2]).

Probably we should not try to draw very definite conclusions from this small experiment, but it does suggest intriguing possibilities which would need to explored in the context of larger models.

5 Issues and Future Work

From a limited set of experiments performed on a small domain, probably not too much can be concluded. None the less, two surprising (at least to the experimenters) points arose.

- The issue relating to the decay factor called into question the practice of using a factor present only in the current case;
- The final experiment suggested that it was the ownership of the land which was crucial in Keeble. Our original theories did not take this view, but it is arguable, and worthy of consideration.

This work raises a number of questions which require further exploration:

- The program requires us to choose a number of values for certain parameters; the decay factor, the weights initially assigned to support and conflict links, and the initial values for our case nodes. Changing these parameters can significantly affect the outcome. We need to do some systematic investigation of these effects.
- We have a number of different types of link. In our initial study we used a higher value for conflict links that normal support links, and a higher value still where support can be seen as logical inference. Again there is scope for distinguishing more kinds of link, and changing these relative weights.
- In a large body of cases there will be cases which do not fit with the theory, either because the decisions have been superseded, or because they were ”rogue” decisions. For our initial experiment, we used only a consistent set of cases. We need to further explore the effects of aberrant cases. We also need to consider setting different initial values to represent the different importance of cases, according to their recently, the seniority of the court and the like.
- Considering activation alone seems not to reflect adequately the explanatory power of the theory, or the simplicity of the theory. We need to consider how we can combine the activation measure with measures of these aspects; for example a positive effects from the number of cases nodes included in the theory, and an adverse effect from the total number of nodes.
- The models used in the experiments derive their connections from an idea of which nodes are used to construct other nodes in theory construction, and which nodes are applied to explain and derive other nodes during theory application. We have already begun work on a more formal and disciplined account of how theories can be transformed into graphs representing the theory and the explanation and inference relations within it. This account will be used to give a stronger rationale for the connections in the theory models.
- In general we need to conduct further exploration through experiments on larger models, on models constructed with different views on connections, and with different values for the various parameters.

We find this approach interesting and holding promise of more interesting results. The question, of course, arises as to whether the approach has any cognitive validity with respect to the ways in which lawyers evaluate theories. Once the technical questions above have been answered, some kind of empirical study will be required to see how far the judgements on theories given by this approach can be seen as reflecting a legal consensus.

References

[1] Aleven, V. (1997). *Teaching Case Based Argumentation Through an Example and Models.* PhD Thesis. The University of Pittsburgh.
[2] Bench-Capon and Rissland (2001) *Back to the Future: Dimensions Revisited*. Submitted to JURIX 2001.
[3] Bench-Capon, T.J.M., and Sartor, G., (2001). Theory Based Explanation of Case Law Domains. In *Proceedings of the Eighth International Conference on AI and Law,* 12-21. ACM Press: New York.
[4] 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,* 50-59. ACM Press, New York.
[5] Thagard, P. 1992. *Conceptual Revolutions.* Princeton (NJ): Princeton University Press.