Artificial Intelligence Review **12:** 393–427, 1998.

Understanding Similarity: A Joint Project for Psychology, Case-Based Reasoning, and Law

ULRIKE HAHN* and NICK CHATER
Department of Psychology, University of Warwick, Coventry, CV7 4AL, U.K.
(Current address: Department of Psychology, University of Wales, Cardiff, U.K.*
E-mail: ulrike.hahn@psy.ox.ac.uk)

Abstract. Case-based Reasoning (CBR) began as a theory of human cognition, but has attracted relatively little direct experimental or theoretical investigation in psychology. However, psychologists have developed a range of instance-based theories of cognition and have extensively studied how similarity to past cases can guide categorization of new cases. This paper considers the relation between CBR and psychological research, focussing on similarity in human and artificial case-based reasoning in law. We argue that CBR, psychology and legal theory have complementary contributions to understanding similarity, and describe what each offers. This allows us to establish criteria for assessing existing CBR systems in law and to establish what we consider to be the crucial goals for further research on similarity, both from a psychological and a CBR perspective.

Key words: Artificial Intelligence and Law, case-based reasoning, exemplar-models, legal reasoning, similarity

1. Introduction

In this paper, we bring together three fields: CBR,[1] psychology and law. These areas have considerable potential mutually to inform each other. Our focus on their relationship stems from an interest in the experimental and computational investigation of legal reasoning from both psychological and legal backgrounds. Our aim in this paper is to illustrate why these fields are particularly profitably linked, to outline what it is they currently have to say to each other, and, finally, to establish where it is we see most joint interest for future research.

CBR originated as a theory of human cognition, but has been the subject of little experimental or theoretical investigation in psychology. However, a wide range of psychological theories are based on the assumption that many

[1] We will use "case-based reasoning" to refer to case-based processes in human agents, whereas the acronym CBR will be reserved for the AI-endeavour of building systems whose reasoning is based on cases.

cognitive processes are best explained through generalization from a large number of stored instances. Generalization from past instances, whether in psychological theory or in CBR, is presumed to depend on the *similarity* between new and old instances. Hence, similarity is a central focus for any instance-based account of human or artificial reasoning.

Yet, similarity is also notoriously problematic. Skeptics argue that similarity is not a coherent notion at all, but a 'pretender, an imposter, a quack' (Goodman 1972). CBR and psychology both attempt to counter this skepticism by providing explicit accounts of similarity; psychology has also aimed to show that human similarity judgements are, in fact, highly constrained. These complementary attempts to understanding similarity, which are so fundamental to instance-based approaches to reasoning, are the focus of this paper.

1.1. *Why Law?*

Law is a rich domain in which to study reasoning from instances (or 'cases') for a number of reasons. First, law, at least in the Anglo-American tradition, represents the most sophisticated domain of human thought that is overtly concerned with reasoning from cases (Llewellyn 1930; Levi 1949). Second, both legal argument and decision making are conducted explicitly, making legal reasoning more accessible than most high-level reasoning tasks. In particular, both 'instances' (cases) and 'rules' (statutes) have explicit, verbal manifestations. This means that we can make non-arbitrary assumptions about crucial parts of the knowledge involved in the reasoning process. Third, the verbal nature of both the legal materials and many of the considerations evoked in the reasoning process make the comparison of a CBR system's performance with that of legal experts easier than in many other domains. Fourth, in law, – and this is a point we will return to later – cases and general rules are inevitably intertwined. Thus, law offers case-based and rule-based processes as well as numerous 'blends', making it an ideal domain in which to study various facets of the relation and interaction between cases and rules. Finally, legal reasoning is comparatively well-understood. Jurisprudence (legal philosophy) which aims, among other things, to make sense of the legal process itself, provides explicit meta-theory with a long tradition. Though lacking direct empirical support, it can greatly inform more detailed cognitive investigation; particularly helpful in this context is the ever growing jurisprudential discussion directly addressed at AI and Law (Susskind 1987; Bankowski, White and Hahn 1995, e.g.).

Any progress made in the understanding of case-based reasoning in law necessarily benefits jurisprudence, psychology, and CBR, and all three disciplines have particular contributions to make. Leaving the latter issue for the final section, Section 6, the respective benefits can be summarized as follows: for jurisprudence, psychological and computational work provides the currently sparse empirical backing. For psychology, this area is important as it is in the study of cognitively *high-level*, instance-based processes that experimental psychology is currently most lacking. Additionally, there is the promise that insights gained in law extend to other areas of cognition: reasoning in law as ‘an intermediately formalized field midway between logical or mathematical domains and common-sense reasoning’ (Ashley 1990) is full of connections and parallels in both directions (e.g., both legal and common-sense reasoning involve defeasible inference (Oaksford and Chater 1991)). From a CBR perspective, finally, the area has significance as legal CBR has been a major field of research in the past, producing some of the most sophisticated systems to date. The goal of CBR in the legal domain, however, has to be to model human legal judgements – to decide new cases as a lawyer would. In law, the principles upon which the artificial system determines similarity must be analogous to those underlying the *human* judgement of similarity. This sets law apart from application domains such as, for example, aluminium die casting or fault diagnosis on steam engines (Price and Pegler 1995; Georgin, Bordin and McDonald 1995). This means that the project of CBR in law and the project of understanding human legal judgements are intimately related. As with psychology, there is the additional hope that such an understanding will also contribute to other areas in which it is human similarity judgement that matters.

1.2. *Overview*

In this paper, we illustrate in detail how collaboration between psychology, CBR, and law can and should proceed. We first outline the psychological research concerned directly with reasoning from instances. Section 3, ‘Psychological Research on Similarity’, concentrates directly on general psychological research on similarity.[2] Section 4, ‘Similarity in Law’, illustrates the various ways in which legal cases can be ‘similar’, highlights the difficulties involved in matching legal cases, and pins down the role of similarity in case-based decision making. Additionally, two paradigmatic systems are discussed with respect to similarity assessment: HYPO and GREBE. Finally, the question of whether similarity in law can be captured without ‘factor-weights’ is addressed. Sections 5 and 6 draw these materials together to address whether

[2] With one important caveat – we more or less completely sidestep the analogical reasoning literature. This is not because we find it irrelevant (needless to say it is highly relevant) but because it is, in our opinion, an area in which research is truly interdisciplinary and where computational and experimental communities are well aware of each others work. Given only limited space, it seems justified to focus on areas less received in the CBR community.

'similarity' is a unitary construct (Section 5) and to outline future directions for integrated research between CBR, psychology and law, establishing the particular contribution each of these areas has to make. Here, we aim to show that psychology and CBR are complementary in that each can perform empirical investigations the other cannot and finish with a list of six issues for future investigation.

2. CBR and Psychology

There are two distinguishable areas of psychological research with relevance to CBR. The first is the direct investigation of instance-based approaches in various areas of cognition. The second is psychological modelling and experimentation. We begin by briefly reviewing the literature directly concerned with instance-based accounts.

Whether learning from experience involves abstracting general rules or storing detailed memories of specific experiences is a fundamental question in cognitive psychology. Independently of CBR, instance or 'exemplar-based' views have been advanced in many areas of psychology. In the literature on human concept formation, exemplar models (Medin and Schaffer 1978; Medin, Dewey and Murphy 1983; Nosofsky 1988, e.g.) are contrasted with rule-based definitional accounts (Rips 1975; Nosofsky 1994). Some (albeit few) of these models and procedures have found their way into CBR, e.g., Medin and Schaffer's "Context Model" which has been put to use in PROTOS (Porter, Bareiss and Holte 1990).

Furthermore, it has been proposed that apparently rule-based "schematic" aspects of memory can be understood as emerging automatically from generalizing large numbers of specific memory traces (Hintzman 1986). There has also been extensive debate concerning whether "implicit learning" involves the abstraction of rules, or can be explained by reliance on specific stored instances (Reber 1989; Shanks and John 1994; Redington and Chater 1995). The psychological study of analogy also focusses on how memory of examples can be employed in problem-solving and reasoning (Gentner 1989). Finally, a central theme in the recent psychology of reasoning has been the emergence of accounts that do not appeal to domain independent rules, but assume that reasoning is tied to specific instances (Griggs and Cox 1982).

Though intended as a general theory, the experimental materials involved in these traditions have almost exclusively been fairly low-level, artificial stimuli such as geometric shapes. CBR research, conversely, deals with high-level, 'real-world' tasks and materials (for overviews see Kolodner (1991, 1992, 1993) and Aamodt and Plaza (1994)). CBR and instance-based approaches in psychology are currently separated by a 'gulf in subject-matter'. Hence, the arguments for the cognitive reality of case-based reasoning rely predominantly on theoretical plausibility (e.g., (Schank 1982; Riesbeck and Schank 1989)) and have received very little empirical investigation (although see (Ross, Perkins and Tenpenny 1990; Ross 1989)). This suggests that 'bridging the gulf' between the two disciplines should be given high priority.

A further issue is that psychological research in this area is devoted almost exclusively to *distinguishing* rule-based from instance-based accounts. Virtually no work has been devoted to the *interaction* between instance and rule-based knowledge (Ross (1987) and Goldstone (1994a) are rare exceptions). In CBR, by contrast, the relationship between case- and rule-based approaches is frequently both implicitly or explicitly more relaxed. First, we find very generous usage of the term "case", as seen for instance in the work of Schank (Riesbeck and Schank 1989), which views cases and rules on a continuum, thus rejecting a stark distinction. Second, there is a growing body of research on hybrid case/rule-based systems. Given the fact that even fairly superficial consideration of reasoning in high-level cognition suggests that both individual experience *and* general knowledge structures (e.g., verbally communicated knowledge about regularities in the world or statutory legal rules) have a role to play, the investigation of the *interaction* of cases and rules seems imperative, both from the perspective of understanding cognition and of practical systems design.

To summarize the relevance of current exemplar-based research within psychology: experiments have provided important empirical support for exemplar-based accounts in a wide variety of tasks. However, the general focus on artificial, low-level materials and the fact that it almost exclusively pits case-based processes *against* rule-based processes, means that experimental work on instance-based reasoning, memory and learning currently has less *direct* relevance for CBR than is desirable and possible.

3. Psychological Research on Similarity

There is, however, a large body of research in experimental psychology concerned with similarity itself which has had little acknowledgement in AI. This research has two main threads. First, there is an established tradition with psychophysical origins and emphasis which is dominated by geometric models. These geometric models have a fairly long and successful history of fitting (predominantly perceptual) data (see e.g., Shepard (1980, 1987) for overviews). Second, the more recent set-theoretical approach by Tversky (1977) challenges some of the basic assumptions of the geometric view, based again on a wealth of experimental results.

As will become clear, neither model is the last word. However, experimental work on similarity has been developed with reference to these paradigms and their possible shortcomings. Accordingly, we briefly introduce both.

3.1. *Geometric Models of Similarity*

Geometric models of similarity represent items as points in a coordinate space where the metric distances between two points reflect the observed similarity. Most frequently, the space is Euclidean and the aim is to represent the objects in question in a space of minimum dimensionality. Geometrical models of similarity have a methodology enabling one to proceed from experimentally obtained similarity measures to an underlying *similarity space*. This is achieved through a set of statistical procedures referred to as "multidimensional scaling" (MDS) (Shepard 1980, 1987) that generate spatial representations from metric or ordinal "proximity data". This is any kind of data providing information about relatedness of pairs of objects, such as explicit pairwise similarity judgements, recognition data or confusability data.

Formally, a traditional MDS-based model is given by:

$$d_{ij} = [\sum \mid x_{im} - x_{jm} \mid^{r}]^{1/r}$$

where x_{im} is the *psychological value* of exemplar *i* on dimension *m*; the value of *r* defines the distance metric. A value of $r = 2$ defines the metric as Euclidean; other values and thus metrics are possible, e.g., $r = 1$ which specifies the 'city-block' metric. In general, it seems to depend both on the nature of the particular items and subject's strategy which value of *r* better fits the data (Nosofsky 1988; Goldstone 1994a).

These spatial representations can be viewed in two ways: merely as a convenient way of describing, summarizing and displaying similarity data or as a psychological model of mental representation and perceived similarity (Tversky and Gati 1982). To obtain a full psychological model for any cognitive task, this internal space must be supplemented by a specification of how these similarities are used. A prominent example is Nosofsky's *Generalized Context Model*, which extends Medin and Schaffer's exemplar account of categorization (Nosofsky 1984; Medin and Schaffer 1978). It accurately fits subject's performance on recognition, identification, and categorization tasks for a stimulus set of schematic faces (Nosofsky 1988).

3.2. *Set-Theoretic Models of Similarity*

The other dominant approach aims to specify similarity in set-theoretic terms. The most influential model is Tversky's *Contrast Model*, which defines similarity between objects as the weighted combination of their common and distinctive features (Tversky 1977). Specifically,

$$Sim(i,j) = af(I \cap J) - bf(I - J) - cf(J - I)$$

I and *J* are the feature sets of entities *i* and *j*; *a, b, c* are non-negative weight parameters; *f* is an interval scale and $f(I)$, accordingly, is the scale value associated with stimulus *i*. The parameters *a, b* and *c* characterize the task. Due to these parameters, the contrast model provides not one single unique index of similarity but a family of indices. The scale *f* reflects the salience or prominence of the various features, thus measuring the contribution of each feature to the overall similarity. The scale value $f(I)$ associated with stimulus (object) *i* is therefore a measure of the overall salience of *i*, which might depend, for instance, on intensity, frequency, familiarity or informational content (Tversky 1977; Tversky and Gati 1978).

3.3. *Dimensions vs. Features*

Geometric and contrast model both purport to be general and both are potential options for use in computational systems. The contrast model, for example, has been suggested for CBR by Janetzko, Melis and Wess (1993) and geometric similarity measures have been used in a spatial CBR task (Haigh and Shewchuk 1994). The discussion surrounding the underlying assumptions of these models is thus of both theoretical interest in psychology and practical interest for CBR. The psychological debate has taken the form of a critique of geometric models whose 'limitations' supply the support for feature-based accounts.

It has been argued that it is more appropriate to represent some properties (e.g., countries or personality) in terms of *qualitative* features rather than in terms of *quantitative* dimensions (Tversky 1977). However, it is important to point out that MDS, and with it geometric models, does not necessarily require *continuous* dimensions; discrete dimensions are possible and the representation of 'features' as present or not present does not automatically present a difficulty (Nosofsky 1990). Furthermore, conceptual stimuli might often be structured in a way that gives rise to hierarchical featural groupings or clusters and thus 'pseudo-dimensions' (Garner 1978).[3] This general point is illustrated by Ashley's (1990) HYPO system, which is discussed in Section 4. Problems do arise, however, with nominal variables admitting of several

[3] To provide an example of Rosch's (Rosch and Lloyd 1978): the presence of an automatic transmission can be treated as a binary feature; once it is decided that the relevant set of objects are cars and that cars must have a transmission, 'automatic' and 'standard' become two levels on the pseudo-dimension 'transmission'.

values such as 'eye-colour', where the values ('brown', 'blue', 'green') have no meaningful serial ordering, a constraint inherent in the notion of dimension (Hahn and Chater 1997).

A stronger criticism is that the metric axioms, which underly the idea of similarity as a distance, are empirically violated. These are:

1. Minimality: $\delta(a,b) \geq \delta(a,a) = 0$
2. Symmetry: $\delta(a,b) = \delta(b,a)$
3. Triangle Inequality: $\delta(a,b) + \delta(b,c) \geq \delta(a,c)$

Empirical results have been gathered (particularly by Tversky) that seem to conflict with each of these assumptions.

Minimality is challenged by certain types of recognition data where an object is identified as another object more frequently than it is identified as itself (Tversky 1977). All this conclusively establishes, however, is that some data is not unproblematically treated as proximity data: if identification probability violates minimality when interpreted as an immediate measure of similarity, this simply demonstrates that this straightforward identification is not possible. It does not make it meaningful to introduce a notion of self-similarity which can be greater or lesser.

The triangle inequality is elusive; it cannot be formulated in purely ordinal terms, and is difficult to refute even with interval data (Tversky 1977). Nevertheless, the case of a violation is reported by Tversky and Gati (1982), but it remains difficult to judge how widespread such violations are.

Far more convincingly established is, at present, the existence of asymmetries. That the similarity of *a* to *b* need not equal that of *b* to *a* is illustrated intuitively by similes such as "butchers are like surgeons" and "surgeons are like butchers", which differ in meaning with respect to whom they compliment or criticize (example from Medin, Goldstone and Gentner (1993)). A number of experiments have demonstrated that this effect is not specific to similes, but occurs with similarity statements ('*a* is similar to *b*') and directional similarity judgements (Tversky 1977; Tversky and Gati 1978; Rosch and Lloyd 1978). However, these asymmetries might arise from *processing*, not from the internal space itself; as mentioned above, the cognitive model of performance on any given task requires a supplementary specification as to how these similarities are used. In fact, enhanced geometric models which allow flexible attention weights on dimensions (Nosofsky 1988; Aha and Goldstone 1992) can deal with asymmetries in a way analogous to the contrast model. The model becomes:

$$d_{ij} = [\sum w_m \mid x_{im} - x_{jm} \mid^r]^{1/r}$$

with weights w_m.

Asymmetries are captured by assuming that in directional comparisons such as ‘c is like d’ (as opposed to non-directional assessment of similarity such as ‘how similar *are* c and d’) one naturally focusses on the subject (a). For the contrast model this means that the features of the subject are weighted more heavily than the features of the referent (thus: $a > b$, above), with the consequence that similarity is reduced more by the distinctive features of the subject than by the distinctive features of the referent. For geometric models, one can assume that dimension weightings are selected by focussing on the properties of the subject of the comparison, and, hence, need not be the same in both directions (a, b vs. b, a), leading to different values for each comparison.

From a CBR perspective, it is worth noting that asymmetries arise naturally in systems where matching is dependent on the direction of the match. Where a metric is applied uniformly, however, asymmetries could present a problem. So far the magnitude of experimentally observed asymmetries has been relatively small. Further work needs to be done: the size of these effects with legal materials, for instance, has yet to be determined. Only then can it be judged whether it is necessary to allow for asymmetry in a system or whether, in fact, assuming symmetry, though false, remains a sufficiently close approximation.

In the meantime, despite heavy attacks on geometric models, whether as a cognitive model or as a practical procedure, both types of model, geometric and set-theoretic, are still widely viewed as viable options for psychology (for a more detailed discussion than possible here see Hahn and Chater (1997)).

3.4. *Selecting Features and Dimensions – What Both Approaches Leave Out*

Finally, psychological literature contains a wealth of recent materials on an aspect of similarity neglected by both models: the problem of determining *which* factors enter the comparison and how they are assigned weights. MDS generates a spatial representation from proximity data, that is, based on what subjects actually considered relevant. What gives MDS its power is the very fact that these dimensions need not be known in advance. Feature-based models such as Tversky’s, on the other hand, presuppose an appropriate feature set. Both fall short, in that they have no story on how the relevant properties (featural or dimensional) are initially selected.

Excluding the process of feature selection brings both theoretical and practical limitations. In law, determining the relevant dimensions of comparison is a central part of the problem. Ideally, CBR systems should be able to cope with this aspect of similarity assessment as well. But it is here that existing systems are most oversimplified or heavily dependent on preprocessing.

On the theoretical side, omitting feature selection threatens to make one's model almost vacuous. Goodman (1972), in his philosophical critique of similarity, pointed out that all entities share an infinite set of properties – a plum and a lawnmower both share the characteristic that they weigh less than 100 kilos (and less than 101 kilos etc.). Hence, in this sense, everything is similar to everything else. The notion of similarity, he concludes, is only meaningful as similar *in a certain respect*, but then it is the respect(s) which does all the (explanatory) work, for Goodman reason enough generally to avoid the notion.

In terms of cognitive agents, 'similar in a given respect' can be equivalently reformulated as *similar in a given representation*. The represented features *are* the relevant respects. Accordingly, similarity is then not viewed as an objective relation between two entities, but as a relation between representations, i.e., determined by the agent. Given a particular (finite) representation, similarity assessment *can* be given an algorithmic account ranging from simple feature counts to complex, multi-parameter accounts like the contrast model. A large part of the explanatory burden, however, now rests on the appropriate representations.

This is all the more an issue as the most pervasive cognitive feature of 'similarity' seems to be its sensitivity to goals, intents and context. This implies *representational* flexibility on the part of the reasoner, whether human or computer. This flexibility must depend – if our current intuitions on similarity are at all correct – on the selection of things to be represented (features, dimensions, whatever) and their respective weights.

Hence, Goodman's original criticism, to some extent, carries through. A model of similarity with no contribution on representation is avoiding a vital factor. On the other hand, a model of similarity should not be expected to provide a full theory of mental representation. It is, however, possible to distinguish a class of issues which can and should be addressed. We establish these in the following.

There are two classes of factors determining representational content and the assignment of weights: *knowledge-based* and formal, non-knowledge-based factors, which arise from the process of similarity comparison itself and which we will refer to as *process principles*. The nature of the knowledge affecting similarity in law will be investigated at length in Section 4. We focus on process principles here.

Relevant experimental results are found in the 'diagnosticity principle', which addresses the effects of context (Tversky 1977), the 'focussing hypothesis' (Tversky 1977), investigations into independence or non-independence of features in similarity judgements (Gati and Tversky 1984; Goldstone, Medin and Gentner 1991), recent studies on the varying impact of relational vs. featural attributes, depending on their relative (quantitative) dominance in the pair of entities under comparison (Goldstone et al. 1991) and 'structural alignment' (Goldstone 1994b), to name the most prominent examples. Additionally, their implications for computational modelling are being explored (Goldstone 1994b). Here, we focus on the two process principles most relevant to CBR: first, the diagnosticity principle and second, what is known as the "MAX-hypothesis".

In studying the effects of context, we find a duality between knowledge-based factors and process principles. As seen in law, the weight of a factor can change in a knowledge-based fashion with the addition of a factor which alters the interpretation of the situation in question: the fact that a cheque has been given provides little evidential support of an intent to fulfill a contract if it also becomes known that the check book is stolen. Context, however, also affects weighting in a formal manner. According to Tversky's (1977) "diagnosticity principle", the classificatory significance of a particular factor (and thus its weight) is dependent, in part, on the set of objects under investigation. For example, the feature 'real' has little diagnostic value in the set of 'mammals' since it is common to all. As it cannot be used to sub-classify them, its weight is low. If the object set is extended to include Pegasus, unicorns, and mermaids, however, its diagnostic value, and thus weight, is considerably increased. This indicates a systematic relationship between feature weights and sub-clusters of objects within a set.

The "MAX-hypothesis" likewise describes systematic shifts in a factor's weight. Goldstone et al. (1991) found that experimental subjects seemed to treat attributes and relations as cognitively distinct groups, with the size of the group affecting the weight attached to any given property. If, for instance, shared relations dominate between two entities, the weight of a single one of these relations will be enhanced. Hence, they found that the impact of changing the same attribute of a geometric figure affected judged similarity more if the stimuli already shared many common attributes (as opposed to relations) whereas it had less impact if common relations were dominant. Goldstone et al. (1991) refer to the cognitively distinct groups of relations and attributes as 'pools'. Succinctly put, the "MAX-hypothesis" states that attributional and relational similarities are pooled separately, and shared properties affect similarity more if the pool they are in is already relatively large. It is possible that there is a general phenomenon at stake here according to which similarities *of the same type* are pooled together, and similarities within a pool influence each other more (i.e., increase their relative weightings) than similarities across pools (Goldstone et al. 1991). This is supported by recent research which found pooling according to thematic criteria (e.g., 'size-related-properties'), again with simple geometric stimuli

(Hahn, Chater and Henley 1996). Further experimental investigation will be necessary to determine the exact conditions under which pooling occurs.

What practical advantage might be achieved by incorporating process principles such as these into CBR systems remains to be seen. A crucial factor will be whether the magnitude of their effects on weights is significant in comparison to knowledge-based influences, an issue likely to be dependent on specifics of domain, task, size of case base and other variable factors. Regardless of practical utility, however, process principles should be part of any theoretically adequate model of similarity. Resuming the above thread, process principles are aspects determining representational content, and thus similarity, which are specific to similarity assessment. As such, they must be covered by any model of similarity which claims to be fully explanatory. Here, research has only just begun.

4. Similarity in Law

4.1. *Representing Cases*

In this section, we consider the sorts of representational flexibility required in law by both cognitive agents and artificial systems, and investigate their legal implications. Finally, we establish more precisely the role of similarity in legal reasoning.

It is well-known that the strength of case-based reasoning lies in the fact that problem solutions can be offered even in domains which lack *strong models*. In these, there is either no theory from which an answer can be deduced, the theory is only partial, or alternative theories compete (Ashley 1990; Porter et al. 1990). Given the fact that in law there is typically "no one right answer" at least in "hard cases",[4] it seems that the lack of strong models here is not merely due to our current insufficient understanding of the domain but rather its very nature.

In short, it is an unavoidable aspect of law, and cases, accordingly, will always have *some* role to play. Nevertheless, there do seem to be significant differences between 'well-understood' and less 'well-understood' areas of law. Legal areas in which we have long experience, that is, where we have seen *many* cases, consolidate into forms of *general* non-instance-based knowledge.

[4] See e.g. (Ashley 1990, p. 233). This is the view taken by almost all legal practitioners and theoreticians. It is doubted by Dworkin (1977) but even for him applies only if time and resource constraints on judicial decisions were to be removed. Additionally, it might be held that *no* case has a unique right answer (and in this sense there are no 'clear cases') but that it is merely a question of the time and creativity devoted to finding viable alternative arguments and solutions which is limited in practice.

For example, a branch of English law such as contract law has evolved a set of clear rules or principles which can be applied in a statutory fashion (Stone 1994). Case-law in such well established domains is relegated to a more supplementary status, governing only finer points of the law or clarifying the exact scope of a legal predicate. In contrast, in ill-understood, rapidly developing domains such as some areas of "tort" (delict) , no adequate set of general principles has yet been found and cases remain the predominant source of information in deciding a novel case (Weir 1988). These considerations have immediate implications for both the demand in CBR systems and the requirements which such a system must meet.

In deciding on the representation of cases, the systems designer faces a trade-off: the more one knows exactly *a priori* which features (i.e., dimensions) to select and how best to represent them, the more one knows about the domain in question, and, hence, the less one needs cases at all! The less one has a domain theory, the less one knows what to select, the more one has to be able to *reason* about what is relevant and, hence, the more representational flexibility one has to provide. In law, this involves various types of semantic knowledge, legal knowledge, purposes, and more – a formidable task. We now identify these various sources more precisely by investigating the different ways in which (real) legal cases might "match".

There are many ways in which two, ultimately similar, legal cases might mismatch. For example, each might contain features which have no equivalent in the other. This might take the form of differing *values* for a (possibly binary) feature. Or, the difference might be at the *representational level* itself, that is, the representation of one case does not mention this feature at all. Often, the presence or absence of the missing aspect may be inferred, leading to a refined representation exhibiting a greater degree of match (such inference is one of the main aspects of the legal CBR system GREBE discussed in Section 4.2). Another possibility is that two cases exhibit a superficial mismatch, which is resolvable because the features in question are equivalent with respect to some classification and it is only membership in this category that matters. For example, two legal cases – one involving a motorcycle and one involving a car – might, regardless of the mismatch 'motorcycle/car', exhibit a high degree of similarity due to the fact that both concern 'vehicles'. Such classificatory equivalence may or may not be reflected in the fact that they share an established super-ordinate category.

Conversely, dissimilar cases might superficially match. Cases might correspond from a featural perspective but clash from a relational one, for example, where the same type of entity is present in both cases but in a radically different role. Here, one can imagine two cases about traffic accidents, one in which a car drives into a wall, and a second in which a cyclist drives into a parked car. The fact that both cases include cars will not form a relevant commonality.

This set of possibilities implies various types of knowledge – e.g., semantic, common-sense, and domain specific heuristics and factual knowledge – in legal similarity assessment. It does not, however, apply to law in any unique way, and there are many domains in which the same difficulties may arise. In the following, we concentrate on two difficulties of matching, which, though general, also illustrate important peculiarities of the legal domain. The first reflects the problems in selecting the right, relevant subset from the infinite set of features with which to represent the case set. The second relates to the level of abstraction at which cases are represented.

4.1.1. *Feature Selection, Defeasibility, and the "New Feature" Problem*

The representation of real legal cases (i.e., a description of the case in a law report, etc.) is not *uniform* in the sense that they are not presented in terms of an a priori determined *fixed* set of factors (dimensions, features or relations). As mentioned, this is largely due to the fact that we generally *don't know* definitively what possible aspects might be relevant as this depends, for one, on understanding what cases might arise in the future.

In deciding on a particular set of features to represent a precedent, (in 'real-life' or in a computational system) one can 'err' in two complementary ways: by including irrelevant features, or by excluding relevant ones. While the selection of the features with which to represent a precedent is not definitive as is the formulation of a rule, case-based reasoning nevertheless requires a decision on the *importance* of the unmatched features in precedent or new case for classification.[5]

Providing a decision even where a new case lacks features of the precedent is part of what CBR is all about. In this sense, the presence of irrelevant, 'extra' features in the precedent touches on what CBR is good at. However, the representation of superfluous features in the precedent will present a problem if they occur in larger numbers or will lead to misclassification of new cases which lack only these features. Clearly, this can be amended through *learning*. The practical utility of CBR systems in law, however, does demand that this learning process does not lag too far behind the solutions to be found by a human reasoner – equipped with massive amounts of knowledge – at any given time.

The reverse problem, where it turns out that not all potentially relevant features were represented in the precedent, i.e., where a new case has a *novel*

[5] To clarify, the issue here is not of inferring the presence or absence of the non-represented feature but rather that of deciding on the role this feature has in the classification of the case.

but relevant feature,[6] relates to a special instance of *defeasibility* in law. Defeasibility in law is both ubiquitous and multi-faceted. Its centrality makes it worthy of pause for consideration. We outline its various forms in law in the familiar guise of defeasible inference from rules, examining how this translates to CBR subsequently.

First, defeasibility arises as a result of the fact that we can never entirely foresee all possibilities in legislating or establishing common law rules. This gives rise to "implied exceptions", policy-based exceptions and other restrictions of the application of a rule, even though its antecedent is fulfilled. A prominent example of defeasibility of this kind is exhibited in the case of a murderer who is the heir of his victim. Here, it was held that the heir was not entitled to the inheritance (by invoking the simultaneously 'discovered', general 'principle', that one should not benefit from one's wrongs) even though literally the rules about inheritance applied.

The structure of this sort of defeasibility is described by the notion of a rule which 'perfectly selects' the case in question in the sense that all the conditions defined in the rule are met. Nevertheless, the case exhibits an *additional* feature, not mentioned in the rule, which necessitates a decision contrary to this rule. We will thus refer to this type as *"the new feature problem"*. This problem cannot be dealt with through more circumspect rule definition. It demands the ability to perform default reasoning, since these exceptions are not enumerable in advance, simply because we have no way of knowing them, given both finite cognitive resources and an unpredictable future.

Second, we find defeasibility of a *pragmatic* sort throughout modern legal systems; in order to make the proof of certain legal facts easier, to protect certain legal positions or strike a balance between the parties, the law frequently distinguishes between *constitutive facts*, which have to be established in order for a legal consequence to hold, and *impedient facts*, which hinder this consequence *if* established, but do not require that contrary proof be given (i.e., proof that they do *not* exist for the consequence to be effective). Examples of this kind are 'rebuttable presumptions of fact'[7] or exceptions to norms which are left to the party seeking them.[8] The pragmatic distinction involved here (i.e., who has to establish what in order to derive a conclusion) necessitates

[6] Again, this is to be distinguished from the case where a feature is represented but is not present. The latter corresponds to *value* mismatch (binary or other).

[7] That is, the assumption of a certain fact unless *disproved*. This amounts to a reversal of the burden of proof between parties, generally made in cases where certain sensitive rights require recognition of the fact that the circumstances are particularly difficult to ascertain by the holder of the right, thus threatening to undermine it in practice.

[8] Examples in German law are self-defence in civil matters or the fact that a claim is time-barred.

some form of default- or non-monotonic reasoning. Again, defeasibility cannot be avoided by 'better feature selection' which collects all possibly relevant features into a single rule. The combination of the conditions of a norm with the negation of all possible impedient facts in the antecedent of a rule, though possible, would violate these pragmatic distinctions, as it fails to reflect that the consequence in question *does not* require proof of these negative facts (for detailed investigation from both a legal and a logical perspective see Sartor (1991, 1995).

Finally, the explicit 'rule plus exception' structures found in law in general can be viewed as 'defeasibility' of a lesser sort. Statutory materials frequently employ general rules plus more specific rules covering exceptions: the German civil code, for instance, contains general rules governing contract as well as sets of more specific rules, covering special types of contracts such as sales of goods or rent, which replace or modify these general rules. The general rules, in this case, can be applied only if the more specific ones are inapplicable. This is defeasibility in a sense, as it can be likened to birds, penguins and the capacity to fly. The fact that these exceptions, however, are both explicit and finite makes a difference. A representation scheme which compounded these rules plus exceptions into one conjunct *would* be both possible and adequate. Non-monotonic or default reasoning is thus not necessary.

All three types of defeasibility can have analogues in legal areas governed not by rules, but by cases. The "new feature problem" variant merely requires a new case with an unexpected and, hence, in the previous case(s), unrepresented feature. Pragmatic defeasibility arises if the legal systems provides defeating conditions whose establishment is left to the interested party. Such conditions need not be founded in statutory rules, but could equally be introduced in case law. Finally, the 'rule plus exception' situation corresponds simply to cases having matches of different degree. Here, the new case merely matches the exceptional precedent(s) more closely (Branting 1989). Hence, this last form of defeasibility is dealt with in a straightforward manner. Given representational care, pragmatic defeasibility seems achievable as well. CBR, finally, is in the same boat as rule-based approaches when it comes to the "new feature problem": all legally relevant features of a precedent are present in the new case, but additionally this new case contains an aspect which demands an alternative decision. The problem is determining the influence of this new factor on the outcome of the case in question. This problem is the same, whether it arises in the context of cases or of rules. In summary, this leaves CBR at par with rules on defeasibility.

The "new feature problem" is *hard*. Seemingly, it can only be dealt with in a knowledge-based fashion. The system – rule or case-based – must have general knowledge in order to decide whether unmatched features of the new case merit a different decision. It is thus not surprising that, to the best of our knowledge, all CBR systems in the legal domain with the exception of CHIRON (Sanders 1994) *ignore* the new feature problem, most frequently by not even allowing for such new features to be represented. In legal CBR, this is the case with HYPO where cases can be matched on the pre-specified dimensions. Alternatively, whilst more flexible in their representation scheme, other CBR systems discard unmatched features of the new case from the similarity matching process (e.g., GREBE (Branting 1991a), PROTOS (Porter et al. 1990)). Finally, CHIRON can only make use of unmatched features as part of a HYPO-style (see below) 3-ply argument generation; it cannot establish the relevance or irrelevance of a 'new feature' nor does it aim to decide. Nevertheless, in practical terms, CHIRON might at least suggest something to a human lawyer by pointing out these distinctions.

4.1.2. *Levels of Abstraction and the Role of Similarity*

The second matching difficulty with particular relevance in law, level of abstraction, is less complex. Whether or not two cases match directly often depends on the level of abstraction chosen to represent them. Imagine, for instance, the infamous case Donoghue vs. Stevenson: Mrs. Donoghue raised an action of reparation against a manufacturer of aerated waters on the ground that she had drunk some of the contents of a bottle of Stevenson's ginger beer, bought for her in a pub, before discovering in the remainder the remnants of a decomposing snail. She subsequently suffered gastro-entritis and a nervous shock. This case can be described as being about 'a ginger beer in an opaque bottle containing a decomposing snail' or as '. . . containing a decomposing animal'; these versions differ in degree of match with a novel case involving decomposing worms. Dealing with (different) levels of abstraction is a pervasive feature of law. Legal rules are typically more abstract than the description chosen for the cases they subsume; this is what makes them general. It also leads to the typical legal classification problems of whether, for example, a particular behaviour can be described as displaying 'reasonable care'. In legal theory the term 'open-texture' (of predicates) is used to refer to the uncertainty introduced by this gap in generality. CBR has been championed as a solution here because the facts of the precedent and the new case are expressed at the same low level of generality (Branting 1989). CBR, on its own, however, inherits the opposite problem: to what degree two cases should be matched because they fall into the same higher-level category despite a surface mismatch.

It is vital to stress here that enabling a CBR system to appropriately match at an inferred, higher level of abstraction is not simply a matter of supplying semantic knowledge in form of taxonomic relations. One reason for this is that higher-level matches may go beyond existing linguistic categories. Human reasoners 'see' abstract commonalities between cases that often forms the basis of a novel legal category and thus a legal term. More importantly, though, choosing the right level of abstraction at which to match is at the heart of the actual legal problem. Legal decision making strives for generality because our very notion of justice demands that like cases be treated alike. Unless a new case displays a *relevant* difference, it must be treated like the precedent. Deciding whether a difference is significant enough to merit a different decision is a core task of legal reasoning. Matching at a higher level of abstraction obliterates differences, and thus touches exactly on this issue. In legal parlance, the decision about which degree of abstraction to adopt in matching two cases is the decision about the *ratio decidendi* of a case (Cross 1977; Branting 1993). The problem is whether Donoghue vs. Stevenson is a case about 'dead snails in opaque bottles of beverage', 'snails in bottles', 'unpleasant foreign objects in chattels for human consumption', 'foreign bodies in objects', or any possible combination of these (MacCormick 1987). This is a decision about what *'the law' is*, not just about similarity.

What distinguishes law from many classification problems in other domains is that there is *no prespecified answer*. Indeed, there is generally no one 'right' answer. Case-based reasoning in law is not a method of getting at an underlying, independent category structure, as case-based medical diagnosis can be seen to be. This means that legal reasoning based on precedent frequently exhibits a strong element of *choice*. This choice is not itself made through case-based reasoning. It is a constrained choice – constrained through precedent and hypothetical cases for which we have clear intuitions – as we cannot depart from these on the basis of arbitrary, irrelevant distinctions. However, whether or not a distinction is arbitrary or legally conclusive *coincides* with the question of the breadth of cases to which a particular legal consequence should be applied. The supporting considerations in both cases are the same.

This has a number of implications for the role of similarity, and hence case-based reasoning in law. Case-based reasoning in law presents itself as a two-stage process. In the first stage, the precedents with some bearing on the case at hand are retrieved. The second stage is a decision of the new case. Because there is no perfect match, this will often involve complex considerations which narrow down the aspects of the case deemed relevant to a few, possibly one or two, crucial aspects on which the decision is based. Similarity, in the psychological sense we have been discussing, governs only the first stage – the selection of relevant cases. Here, we see similarity judgement in a way analogous to the wide variety of other tasks studied in psychology. In contrast, the actual *decision* of the new case, though strongly influenced by the degree of similarity found to retrieved precedents, typically contains degrees of freedom which are not resolved on the basis of similarity alone, but through deliberate considerations of purposes, goals, and the desired shape of the law; this stage can involve economic, social, legal, and plain common-sense considerations. In terms of similarity, the decision of the case through a legal institution authoritatively *establishes* the similarity or dissimilarity to the precedent by virtue of the adopted classification. This final similarity is resultant, however; only rarely will it coincide with the starting point.

In general, then, similarity assessment in law is hard. Even the ‘retrieval phase’ of finding meaningful possible comparisons requires semantic knowledge, common-sense rules, and legal knowledge. Hence, detailed consideration both of which area of law a practical system can meaningfully operate in and what it can realistically be expected to do is required.

4.2. *Similarity in Legal CBR: Assessing Similarity in GREBE and HYPO*

With its ever-growing number of systems, CBR has provided a wide range of ways in which to compute similarity, and commercially available shells such as ReMind (Cognitive Systems Inc.) typically offer a selection. General purpose measures (for an overview see Herbig and Wess (1992)) are frequent, as are task and representation specific solutions. Legal CBR has largely adopted the latter approach. Rather than attempting a general survey of the wealth of legal CBR systems, we provide a detailed analysis of two of the most widely known systems – Branting’s GREBE and Ashley’s HYPO – which, in many ways, provide two different, paradigmatic approaches, complementing each other in the difficulties they try to tackle and those which are avoided by simplification. GREBE’s case representation is less rigid and less dependent on a pre-existing analysis of the domain. Hence, similarity matching is more complex, but also more true to life. HYPO’s simplifications here, on the other hand, enable the system to display considerable sophistication in the generation of arguments based on similarity to precedents. Space considerations prevent detailed general overviews of these systems – for these the reader is referred to the literature (Branting 1989, 1991; Ashley 1990). Here, we focus directly on the aspects relevant to our discussion.

4.2.1. *GREBE*

GREBE aims to reason directly from a factual level (i.e., working out the legal significance of the represented factors is part of the program’s task). Branting considers this to be crucial both because working out the *legally relevant* aspects of a case is often the largest part of the decision making problem and because the flexibility of case-law as a source for multiple, possibly competing, arguments about the legal consequences of a set of facts requires ‘a representation that is free of bias toward any particular analysis’. Any unbiased representation, however, must be of a ‘finer granularity’ than legally relevant features (Branting 1989). This makes the system interesting from the point of view of the types of knowledge relevant to similarity in law as discussed above (Section 4.1). Branting aims to take account of the fact that determining the relevant similarities between precedents and a new case with respect to a legal category requires knowledge of the *explanation* of the precedent’s membership in that category, a view for which he draws support both from the psychology of concepts (Murphy and Medin 1985) and legal theory (Raz, 1979; Murray 1982). The role of explanations in improving similarity matching becomes apparent from our discussion of levels of abstraction above in 4.1. Explanations can inform the decision on which level of abstraction to adopt in matching two cases. Explanations, like rules, thus *constrain* the process of similarity assessment.

In GREBE, explanations take the form of *precedent explanations* (later termed “precedent constituent” (Branting 1991b)) where “explanation” is a collection of reasoning steps that relate the facts of the case to the solution of a problem in the case.[9] The precedent explanations are flanked by both legal and common-sense rules, the latter consisting both of inferences such as ‘if an activity is a duty of employment then each necessary step of that activity is a duty of employment as well’ and semantic rules expressing taxonomic relations among predicates. Given a new case and a proposition about that case, GREBE aims to construct an explanation for that proposition using any combination of rules and precedent explanations. Because cases in GREBE are represented in a semantic network, matching is done through structure mapping, thus focussing on relations rather than features. GREBE’s knowledge is used to minimize the degree of unexplained match between two cases where possible. For example, facts which are not explicitly stated can be inferred, as can a match between distinct features that are members of the same more abstract category. Ultimately, one or more explanations of the supplied propositions are provided as output.

Finally, it is interesting how Branting discerns two distinct stages:[10] *(1)* determining relevant similarities and differences between a new case and each applicable precedent, and *(2)* using knowledge of similarities and differences

[9] Another interesting feature of GREBE in this context is that the precedent explanations (or ‘constituents’) can be put to use outside the context of the entire case from which they stem. The use of *parts* of precedents, Branting points out, is often more important than the degree of match to any whole precedent (Branting 1991b).

[10] In the particular context these distinctions are made with respect to the case-based evaluation of an open-textured predicate which has been subject to relevant case law. We find, however, that this description pertains to case law in general, whether in the context of clarifying open-textured predicates or not.

to determine the ‘controlling’ precedent, and hence outcome of the case (or construct conflicting arguments, if no unique controlling precedent can be found). This distinction corresponds to the two stages we identified above, stage (1) being the domain of case-based reasoning, and stage (2) comprising a complex, multi-layered decision process. GREBE, with its knowledge-based refinement of similarity matching, illustrates well that even the first stage in law relies on knowledge and inference.

4.2.2. *HYPO*

HYPO, in contrast, has a different emphasis, focussing less on classification than on modelling various aspects of precedent in legal *argument*. This corresponds to the second of our stage decision making process. HYPO does not aim to provide a unique decision to a new case. Rather, it generates arguments, rebuttals and counter-arguments on the basis of the relevant precedents it discovers. The system explicates how the relevance of the set of factors in a legal case depends on how they are used in legal argument (i.e., in which sides argument and in what role, citing or rebuttal) and the other cases in the set. In this sense, it is an investigation of various specific aspects of context, in particular of purposes or goals.

These strategic aspects of argument and their influence on relevance are, as belonging to stage (2), outside of the basic step of assessing similarity as it concerns us. It must thus suffice to say that HYPO displays considerable sophistication in these respects, making an important contribution to our understanding of legal argument. It is also clear, however, that this sophistication is bought at the expense of considerable simplification in the basic assessment of similarity performed.

Cases are represented through a fixed set of *dimensions*, a knowledge representation construct for the representation of *‘factors’*, that is, collections of stereotypical facts that strengthen or weaken the plaintiff or defendant’s position on a particular legal claim. These ‘factors’ have magnitudes, reflecting that a particular case may be a more or less extreme example with respect to this factor.

While the dimensional nature of HYPO’s representation means that it can be likened to spatial representations of similarity presented in Section 3, HYPO does not use a similarity measure as such, and, in particular, not Euclidean distance. Instead, HYPO orders the retrieved precedents in terms of ‘degree of on-pointness’ (defined as degree of overlap of dimensions applicable to both the new case and a given precedent) in a directed acyclic graph (tree) referred to as the ‘claim lattice’. This can be viewed as providing an implied *rank order* measure of a gross similarity; magnitude of factors is not considered here. This ordering of cases on the claim lattice is subsequently one of the factors governing the selection of which case best to cite in argument. Additionally, the role of the client seeking advice from the system and the exact argument role (initial citing of the case, rebuttal, counter-argument) are taken into account. Only here does the system incorporate the magnitudes of factors.

4.2.3. *Feature Selection in HYPO and GREBE*

Given that the set of dimensions is fixed, HYPO in no way deals with the problem of determining the relevant dimensions themselves, i.e., with the feature selection process. As we have noted, this limitation in scope can be criticized on the grounds that frequently we do not *know* in advance exactly what these dimensions ought to be. We sketched the problem above that the less one has an adequate domain theory, the less one knows precisely which features are relevant. This suggests that not all areas of law are equally amenable to the sort of preprocessing into relevant dimensions on which HYPO's success is based. One area where a 'dimensional analysis' in the spirit of HYPO may have failed is in the domain of government appeals in criminal cases (Mendelson 1989). Assuming a continuum from strictly rule-governed to strictly fact-governed areas of law (here, no generalizations can be made, all cases are decided on their facts), Mendelson characterizes 'government appeals' as highly fact-governed. It is an area which is very unsettled, fact-sensitive and conflicting, and thus one in which dimensional analysis cannot fully capture what is going on. HYPO's relative success is attributed to the fact that the legal area in which it operates, trade secrets misappropriation, already lies significantly closer to the rule governed end of the spectrum.[11]

With GREBE the situation is slightly different. Again, the feature selection and weighting problem is not addressed. The system assumes that the facts justifying each conclusion reached in a precedent consist of all and only those facts relevant to the conclusion. Furthermore, it is exclusively the precedent that determines the factors along which to match, since non-matching features of the new case are disregarded (see the 'new feature problem' above). But GREBE is far less rigid in its representational requirements, meaning that the system need not be presented an *explicit* value on each of the relevant dimensions, nor are *all* cases assessed along the *same* uniform set. Rather, the dimensionality of the space is determined through the particular cases selected.

[11] It seems possible that Mendelson's conclusions are in part based on a mistaken impression about HYPO's similarity measure; the general point, however, that a domain must exhibit relative uniformity to allow a fixed set of dimensions to be extracted nevertheless holds.

4.3. *Similarity Without Weights?*

Finally, having drawn together various strands of our discussion of similarity such as knowledge influences, representational flexibility and the feature selection problem, there is one element which featured prominently in the psychological accounts but is notably absent in the two systems just outlined: feature (or dimension) *weights*. Neither HYPO nor GREBE make use of weights. Indeed, Ashley has strongly argued that numerical weights for legal factors are inappropriate. He advocates a 'least commitment approach' on weights which deals only symbolically with relevance, taking into account which side a case is being cited for and how this effects the importance of a factor for the argument (Ashley and Rissland 1988). Factors themselves, however, are principally of equal weight. This is in stark contrast to all psychological models presented above, as even the basic spatial models without explicit, flexible weights can account for the relative importance -weighting- of dimensions through the *scale* of the dimension. Hence, the issue merits a detailed discussion.

Ashley and Rissland give four reasons why experts in domains such as law *postpone* any commitment to weights as long as possible:

1. "Such a commitment might cut off certain possibly fruitful lines of reasoning and thereby limit their problem solving performance."
2. "Reduction to numerical weights, in particular, makes it difficult to recover symbolic information needed for certain reasoning methods like case-based justification and contrast-and-compare discussion of alternatives."
3. "Assigning actual 'weights' and predicting interactions among the factors is highly problematic and dependent on individual problem situations."
4. "Experts in domains like the law simply do not reason in terms of weighting schemes. In fact in the legal domain, any reasoner that based an opinion or course of action upon a purely numerical scheme would be highly suspect."

More generally they hold that "the theory doesn't have a numerical method for assigning weights to factors but then neither do attorneys. Legal experts may believe that a certain factor is generally more important than another factor but they seldomly assign numerical weights or probabilities to express the difference and they are always aware of the possibility that in some combinations of factors and magnitudes, the opposite might be true, the usually minor factor may be the more significant" (Ashley and Rissland 1988).

These considerations can be summarized into three distinct issues: weights are inappropriate because lawyers (must) maintain a certain degree of *'open-*

mindedness' about what is important in a case, because numerical weights are simply neither *available* nor *introspectively accessible*, and lastly, because weights are highly *context-dependent* (an objection also endorsed by (Branting 1989)). We address these in turn.

First, the issue of 'open-mindedness' crucially relates to our distinction between two different stages in legal case-based decision making. We hold that in the retrieval of the relevant precedents – the phase we see governed by similarity – weights *do* play a role. This process is highly context dependent, informed by what it is you are looking for, i.e., the novel case at hand. Depending on this, different aspects of previous cases will be considered relevant or irrelevant, while nevertheless always being *present* both in the explicit (written) representation and our recollection of it. We see no way other than through variable weights, in which this particular kind of flexibility could be achieved. However, we agree with Ashley and Rissland that, in the decision phase, the commitment to any particular final 'weighting scheme' is postponed as this is tantamount to making the decision itself. In this second stage, 'weights' continuously shift, as does the selection of what are ultimately the 'crucial' or 'decisive' facts of the case, until this process is terminated through a decision.

Second, on weights not being accessible to legal experts: weights are no more inaccessible in law than in other domains in which the psychological models described in Section 3 successfully capture human performance. Subjects *typically* cannot say more than that one factor is 'more important' than another. There are two independent explanations for this: (1) more precise weighting schemes are being employed but are not consciously accessible; (2), there really are no 'numerical' values in operation, and hence, available. In psychological experiments, subjects generally supply only *rank-order* information (i.e., an ordering, for example, into tallest, medium size, smallest). Nevertheless, the similarity measure provided by the model (e.g., distance) and assumed to underlie performance may be of a far finer grain. Psychologists do not lose sleep over this discrepancy; most aspects of psychological theory are not available to conscious access. It is also possible that, for many purposes, the factor weights employed might make use of only very coarse numbers (possibly some analogue of 'big' and 'small'). Relevant literature illustrating the feasibility of this in the context of reasoning is to be found both in Pearl's approach to default logic (Pearl 1988) and Glymour's work on causal induction (Spirtes, Glymour and Scheines 1993), both of which make use of non-numerical networks. These qualitative approaches can be viewed as operating with very "ill-defined" numbers.

Finally, context dependency can only be marshalled against the use of weights for *practical* reasons in systems design. The flexibility of 'weights' according to context rules out fixed weighting schemes, but not weights per se, as the psychological models above illustrate. The need for flexible weights clearly presents a difficulty for computational modelling. Unfortunately, this does not support the conclusion that the cognitive system operates without them. Hence, from a cognitive perspective, none of these arguments against weights are compelling.

Arguing positively for weights, psychologists generally think of similarity as quantitative, a fact documented by both models described in Section 3. Given this premise, weights seem necessary. Establishing similarity involves the integration of various bits/types of “information” (features, relevance, etc.). How all this could be non-numerical and lead to a quantifiable property seems rather a mystery. Furthermore, experimental data seems to require flexible weights when performance with a single set of stimuli across a variety of tasks is related as in (Nosofsky 1988). None of these considerations lead to watertight conclusions, of course. There might be as yet unknown alternatives for integration, and alternative psychological models might be developed. But current approaches make weights seem central to similarity.

The key question of *where* numerical weights come from, needs much working out. What must be developed is an implementable notion of flexible ‘self assignment of weights’ since there is no external agent to assign them in the cognitive system. At present, we can only speculate dimly as to what this might look like. A general architectural demand might be that one has representations which *interact* instead of discrete, static, knowledge structures which need an explicitly specified control mechanism to determine how they relate. That is, we suspect that a solution should be sought along the lines of effects which arise out of non-centrally controlled interactions of representations. Whilst schemes like interactive-activation models have *some* of these qualities (and also illustrate that factor weights may ultimately correspond more to something like activation levels rather than to weighted connections), it cannot be stressed enough that all current computational paradigms fall rather short of the mark. In summary, the entire topic of weighting in similarity assessment is, in our opinion, a central matter for further investigation, and one which needs the collaboration of theoretical, experimental and computational work.

5. Similarity: Is it Really a Unitary Construct?

The working assumption behind this article is that similarity is at least to some extent a unitary construct. This assumption underlies the psychologist’s search for formal models of similarity which hold across a wide range of tasks and domains. Furthermore, we have continuously been advocating that

CBR and psychology can and should jointly attempt to understand this one phenomenon, similarity.

Our working hypothesis is that perception of similarity is influenced by two distinct factors: knowledge-based factors and process principles. The process principles, we assume, are universal. The specific influences of knowledge, obviously, are not. It is thus through the varying impact of prior knowledge that differences in similarity manifest themselves.

We view this claim as a useful hypothesis, founded on the wish for parsimony and generality. It also respects the intuition, promoted by our ubiquitous use of the term "similarity" widely across contexts in everyday life, that there is some single thing, or at least a closely knit set of phenomena, which is common to all these uses.

We assume the existence of some highly general, universal formula, such as geometrical models or the contrast model, which holds across the many instances of similarity we find. The hard work is being done by a prior feature selection and weighting process, a process which again contains universal, formal aspects (process principles), which, hence, need to be incorporated in the model, as well as domain and task specific influences. But is this perspective correct? Is similarity a unitary construct at all?

5.1. *Stages of Similarity Assessment: An Alternative*

The first alternative is to view similarity assessment as a multi-stage process of a different sort, comprising a computationally cheap retrieval stage and subsequent filtering of the retrieved case-set in a series of one or more, more detailed, similarity matching steps. In contrast, our description (Section 4.1 above) contains a single matching step which is preceded by a non-trivial feature selection and weighting process. The "cheap retrieval plus refinement" approach, however, is common to most CBR systems and is an approach which has been subjected to theoretical refinement, for instance, by Janetzko et al. (1993). These authors propose three stages. First, a 'syntactic' measure of similarity (e.g., Tversky's contrast model) is employed, using only those features of the cases which are explicitly represented. In the second stage, pragmatic relevance (i.e., concerning goals and purposes) is taken into account. This is the first stage that makes use of domain theory other than that implicit in the representation of the cases. The last stage provides a consistency check, employing domain knowledge to decide whether a retrieved case is, for some reason, definitely 'dissimilar' (Janetzko et al. 1993).

Similarly, computationally cheap retrieval mechanisms have been proposed to capture psychological data on analogical reasoning (Gentner and Forbus 1992). Here, the aim is to reconcile both the idea that analogy is dominated by higher-order relational structure and the fact that human retrieval of analogues and remindings is dominated primarily by surface matches (i.e attributional matches plus a limited number of shared first-order relations).

In law, however, the separation of similarity assessment into a computationally cheap ‘syntactic’ step with only subsequent pragmatic and knowledge-based refinement seems problematic. As illustrated above in Section 4.1, legal cases typically match in highly complex ways. Even determining the original set of somehow generally “relevant” precedents requires inference, completion, and the deployment of legal, common-sense and semantic knowledge.

How much of this lawyers perform in “reminding” and whether there are computationally cheap but acceptable approximations to this level of match is an important question, and one on which empirical work has not really begun.

However, for practical systems design, it is crucial to bear in mind that such “remindings” are by far not the only or even the most important ‘retrieval mechanism’ in legal practice. Textbooks, commentaries and case-libraries are consulted extensively, i.e., psychological constraints on memory access do not apply. It is this quality of search that a legal CBR system must match if it is to compete as a source of legal advice.

5.2. *Different Measures for Different Tasks?*

A second possibility is that different *tasks* make use of different types of similarity assessment. For example, low-level and high-level cognition (e.g., comparing geometric shapes vs. legal cases) may differ in the way similarity is computed. Furthermore, even within what amounts to high-level tasks, the existence of a generic similarity measure appropriate for all CBR domains has been doubted (Bartsch-Spörl 1992). Diagnostic domains, Bartsch-Spörl suggests, are well-characterized through perceptible, measurable, frequently continuous features, suggesting the use of continuous similarity measures. For other domains such as configuration, a useful and a useless solution may lie very close together. In this case, similarity in terms of structural equivalence or subset-relationships is suggested. This is reminiscent of the general (psychological) features vs. dimensions debate reviewed in Section 3. It remains to be seen whether the more precise classification boundaries required in configuration cannot be accommodated through knowledge and thresholds, both for practical purposes and psychological modelling. In any case, further research in CBR should reveal the degree to which different tasks require different similarity measures, thus providing important guidelines for psychology.

In summary, whether or not similarity is a unitary construct is a central but currently very much open empirical question, demanding further research. In the following section, we draw together the various strands in this article in order to establish in summary what we see as the main issues for future research on similarity.

6. Directions for Future Research

Both psychology and CBR face the problem of *"scaling up"* , that is making the problems they address more 'real-world'. This involves difficulties for both disciplines. Fortunately, their difficulties are *different*, thus making the two enterprises highly complementary.

Modelling legal reasoning in any discipline is clearly very difficult. Both CBR and psychological approaches can therefore study only scaled-down versions of the full problem. What makes the two approaches complementary is that they 'scale down' in different ways. In legal CBR, one of the largest problems is the amount of knowledge involved, both legal and otherwise. For psychology, it is the practical demands of experimentation. For each, there is thus particular interest in the aspects only the other can provide.

6.1. *The Role of Psychology: Contribution and Limitation*

The greatest commodity experimental psychology has is actual human performance data. These data are crucial as empirical findings are often counterintuitive. Nonetheless, psychological methods are only feasible in limited contexts.

- Psychological theories of similarity and categorization can only be matched to simple numerical data. This means that many important aspects of legal reasoning, e.g., the construction of justifications for particular legal judgements, cannot be dealt with because they provide data which is too complex for quantitative analysis.
- Obtaining a sufficiently large set of data for empirical analysis and theoretical modelling is likely to be extremely time-consuming for all but the simplest of tasks. For example, consider testing the MDS models of similarity judgements which were devised in the context of simple perceptual materials in a legal context. These experiments require a full (or almost full) matrix of pairwise similarity judgements, usually of the order of a hundred or more. However, when such judgements concern legal materials the time required to make a single judgement may be considerable, given that passages describing legal cases must be read, understood and evaluated. Hence, it is inevitable that cases are grossly simplified (e.g., in our own experimental studies (Hahn and Chater 1994), legal cases are simplified to a few lines of text).

– Psychological experiments must be carefully controlled. Only a small number of variables can be manipulated in any study. However, real legal reasoning on actual legal materials will vary on a vast number of different dimensions; it will be impossible experimentally to establish which variables are important in determining behaviour. In all areas of psychology, the requirement for careful experimental control has generally led psychologists to use highly restricted tasks and uniform materials. Additional problems of control are raised in tasks which depend on knowledge (whether general or specialized – e.g., legal – knowledge). Such knowledge may vary considerably between subjects, and the knowledge deployed in tackling a task may even vary within subjects across trials. This has again led psychologists to favour abstract materials, where variations in knowledge are assumed to be less important.

None of these considerations rules out an experimental investigation of legal reasoning but they do place restrictions on which aspects can be studied. The types of models used, as well as the need for rigid experimental control and practical constraints such as task difficulty and duration, mean that only very restricted tasks can be employed. In general, working on tasks requiring what, from the computing perspective, is a vast amount of knowledge, poses no problem to the experimental psychologist. This same prior knowledge does, however, create difficulties if it is the nature of knowledge used in a particular task itself that is of interest.

6.2. *The Role of CBR: Contribution and Limitation*

From the point of view of gaining better insights into legal reasoning and cognition in general, CBR is particularly useful where it aims to model the full complexity of reasoning even if this can only be done using an extremely limited area of expertise. Most theoretically sophisticated CBR systems deal with a small number of cases. They are nonetheless important because they require the designer to formulate explicit accounts of the types of knowledge and inference involved in real legal reasoning. These issues can be addressed even with a system involving only a very few cases. Reasoning in general, and legal reasoning in particular, depends, as we have seen, on a vast amount of knowledge. It is currently not realistic to expect that the entire knowledge involved in real legal judgements can be formalized in any artificial reasoning system for any reasonably sized domain. While this knowledge bottleneck presents a major drawback for *practical* systems building and a major difficulty for cognitive modelling, it has its advantage from the psychological perspective. In contrast to humans, one has complete control over the knowledge a system has, thus allowing the modeller to pinpoint exactly the types of knowledge which are necessary for the task in a way which is well outside the scope of psychological investigation.

6.3. *The Issues*

We now consider the implications of our discussion for future research.

1. An important area is the systematic investigation of what we have called process principles. In particular, further experimental investigation with high-level, contentful materials, rather than the low-level, abstract items typically used in psychological studies, is required. This is an area we are addressing with an experimental program (Hahn and Chater 1994). CBR can supplement this work by providing insights into the impact and importance of process principles such as MAX (Goldstone et al. 1991) even where it disregards these aspects of similarity assessment, since resultant discrepancies to human judgements (if any) give an idea of the magnitude and relevance of these factors. Also, computational modelling can provide an important impetus for the theoretical work necessary to integrate these phenomena into more encompassing formal (and thus computational) models of similarity.
2. A second area is further investigation of the knowledge-based factors affecting similarity. This is primarily a task for the computational enterprise. As stated, building a system to achieve a task naturally provides insights into the knowledge necessary for this task in ways which psychology cannot. In carefully controlled domains, however, psychology can complement this research by experimental inquiry into what knowledge people actually use.
3. Closely related is the issue of weights, which demands a significant amount of further work, both experimental and computational. Most importantly, we need *some* concept of how various types of knowledge might interact to provide appropriate weights. Currently, we have little clue as to how this might happen. In providing a mechanistic account of the weighting process, it strikes us as likely that non-symbolic computational paradigms, or at least aspects thereof, will also need to be integrated.
4. Also related to the types of knowledge influencing similarity is the topic of the interaction of cases and rules, which we have been able to address only in passing. Here, we see an opening both for further experimental (complementing e.g., the work of Ross (1987) and Berry and Broadbent (1984)) and CBR research.
5. The distinction of legal case-based decision-making into two distinct stages needs further empirical support, and each stage requires clarification. As concerns the first stage, experimental investigation can determine

whether any currently available model can fit this stage even in simplified, experimentally controlled conditions. Additionally, inquiry into the factors affecting similarity in this stage seems both feasible and necessary, though restricted to simplified tasks. Regarding the second phase, progress seems both possible through the integration and extension of psychological research on decision making and through more detailed computational modelling, exploring both the types of inference present here and the nature and role of legal argument (e.g., (Gordon 1993; Rissland, Skalak and Friedman 1993; Sartor 1993).

6. Finally, there remains the question of whether similarity assessment is uniform throughout cognition. Answering this question, as much of cognitive science, is complicated by the intimate relation between representation and process which typically permits either representational or processing tweaks as desired to keep the model of a cognitive process afloat (a good case in point is the debate between exemplar and prototype models of categorization – see e.g., (Nosofsky 1992; Barsalou 1989), or the recently revived comparison of both these approaches with rule+exception models (Nosofsky 1994)). Nevertheless, particular models of similarity can be discredited. This can proceed by establishing that the representational requirements of the model are unsuitable or implausible for a particular domain, as illustrated in the debate between spatial models and featural accounts. In this context, the legal domain is particularly useful because which factors play a role is clearer here than elsewhere. And again, systems building, in general, sheds light on the knowledge necessary to achieve adequate performance on a given task. Furthermore, psychology can greatly be supported in its own project of testing models of similarity on a wide variety of different data types by the experiences of CBR with particular measures across domains.

This list reflects our interest in cognitive modelling. But progress on any of these issues can inform the task of building practical, useful AI systems. We cannot yet know whether the contribution will be extended models of similarity which can serve as all-purpose measures, an account of weighting, or a more detailed understanding of knowledge and processes to be captured in law. At the very least it will assist the choice of *realistic* practical projects.

At present, what we know about similarity and its role in law points toward systems operating in domains involving large numbers of cases of relative factual uniformity, with CBR limited to retrieval. A legal CBR system which exemplifies this well is KICS (Yang, Robertson and Lee 1993a; Yang, Robertson and Less 1993b). This system uses CBR of considerable sophistication to retrieve relevant cases in the domain of building regulations. The ultimate decision is left to the user, but given that the domain in question generates relevant cases with alarming frequency, this nevertheless seems useful. Avoiding an ultimate decision, is, we think, demanded by the fact that similarity governs only the retrieval phase of legal decision making. Given the complexity of the decision phase, adequate systems can be expected only in very special domains, the nature of which is likely to make them highly amenable to rule-based approaches (for an example see (Dayal, Harmer, Johnson and Mead 1993)). Relative factual uniformity of the domain is demanded by both the cost of supporting flexible representations (Branting 1989) and the current lack of an adequate model of weighting. This points toward legal applications in areas which, though 'easy' for humans, are tedious due to the number of cases to be taken into account. It is our hope, however, that the collaboration of psychology, legal theory, and CBR can extend these boundaries in the future.

Acknowledgements

We would like to thank Karl Branting and Kate Sanders for valuable comments on their systems. Remaining errors are, of course, our responsibility. Thanks are also due to David Aha and two anonymous reviewers who provided detailed comments on an earlier version of this paper. Ulrike Hahn is supported by ESRC grant No. R004293341442.

References

Aamodt, A. & Plaza, E. (1994). Case-based reasoning: foundational issues, methodological variations, and system approaches. *AI Communications* **7**: 39–59.

Aha, D. W. & Goldstone, R. L. (1992). Concept learning and flexible weighting. In *Proceedings of the Fourteenth Annual Conference of the Cognitive Science Society*. Erlbaum.

Ashley, K. (1990). *Modeling Legal Argument-Reasoning with Cases and Hypotheticals*. MIT press.

Ashley, K. & Rissland, E. (1988). Waiting on Weighting: a Symbolic Least Commitment Approach. In *Proceedings: AAAI – 88*. American Association for Artificial Intelligence.

Bankowski, Z., White, I. & Hahn, U. (eds.). (1995). *Informatics and the Foundations of Legal Reasoning*. Kluwer.

Barsalou, L. (1989). Intraconcept Similarity and its Implications for Interconcept Similarity. In Vosniadou, S. & Ortony, A. (eds.), *Similarity and Analogical Reasoning*, chap. 3. Cambridge University Press.

Bartsch-Spörl, B. (1992). Warum es für die Bestimmung von Ähnlichkeit von Fällen beim fallbasierten Schliessen noch keine Patentrezepte gibt und auch keine geben wird . . . In Althoff, K., Wess, S., Bartsch-Spörl, B. & Janetzko, D. (eds.), *Workshop: Ähnlichkeit von Fällen beim fallbasierten Schliessen*. SEKI Working Paper SWP-92-11 (SFB).

Berry, D. & Broadbent, D. (1984). On the Relationship between Task Performance and Associated Verbalizable Knowledge. *The Quarterly Journal of Experimental Psychology* **36a**: 209–231.

Branting, K. (1989). Representing and Reusing Explanations of Legal Precedents. In *Proceedings of the Second International Conference on Artificial Intelligence and Law*. ACM.

Branting, K. (1991a). *Integrating Rules and Precedents for Classification and Explanation*. Ph.D. thesis, University of Texas at Austin.

Branting, K. (1991b). Reasoning with Portions of Precedents. In *Proceedings of the Third International Conference on Artificial Intelligence and Law*. ACM.

Branting, K. (1993). A reduction-graph model of ratio decidendi. In *Proceedings of the Fourth International Conference on Artificial Intelligence and Law*. ACM.

Cross, R. (1977). *Precedent in English Law* (3rd edition). Clarendon Press, Oxford.

Dayal, S., Harmer, M., Johnson, P. & Mead, D. (1993). Beyond Knowledge Representation: Commercial Uses for Legal Knowledge Bases. In *Proceedings of the Fourth International Conference on Artificial Intelligence and Law*. ACM.

Dworkin, R. (1977). No Right Answer?. In Hacker, P. & Raz, J. (eds.), *Law, Morality and Society: Essays in Honour of H.L.A. Hart*. Clarendon Press, Oxford.

Garner, W. (1978). Aspects of a Stimulus: Features, Dimensions, and Configurations. In Rosch, E. & Lloyd, B. (eds.), *Cognition and Categorization*. Erlbaum.

Gati, I. & Tversky, A. (1984). Weighting Common and Distinctive Features in Perceptual and Conceptual Judgements. *Cognitive Psychology* **16**: 341–370.

Gentner, D. (1989). The mechanisms of analogical learning. In Vosniadou, S. & Ortony, A. (eds.), *Similarity and Analogical Reasoning*, chap. 7. Cambridge University Press.

Gentner, D. & Forbus, K. D. (1991). MAC/FAC: A Model of Similarity-based Retrieval. In *Proceedings of the Fifteenth Annual Meeting of the Cognitive Science Society*. Erlbaum.

Georgin, E., Bordin, F. & McDonald, J. (1995). CBR applied to fault diagnosis on steam turbines. In Watson, I., Marir, F. & Perera, S. (eds.), *First United Kingdom Workshop on Case-Based Reasoning*.

Goldstone, R. (1994a). The Role of Similarity in Categorization: Providing a Groundwork. *Cognition* **52**: 125–157.

Goldstone, R. (1994b). Similarity, Interactive Activation, and Mapping. *Journal of Experimental Psychology: Learning, Memory, and Cognition* **20**: 3–28.

Goldstone, R., Medin, D. & Gentner, D. (1991). Relational Similarity and the Nonindependance of Features in Similarity Judgements. *Cognitive Psychology* **23**: 222–262.

Goodman, N. (1972). *Problems and Projects*, chap. Seven Strictures on Similarity. Bobbs Merill Comp.

Gordon, T. (1993). The Pleadings Game – Formalizing Procedural Justice. In *Proceedings of the Fourth International Conference on Artificial Intelligence and Law*. ACM.

Griggs, R. A. & Cox, J. R. (1982). The elusive thematic-materials effect in Wason's selection task. *British Journal of Psychology* **73**: 407–420.

Hahn, U. & Chater, N. (1994). Similarity in Legal Reasoning – an Experimental Approach. Unpublished manuscript.

Hahn, U. & Chater, N. (1997). Concepts and Similarity. In Lamberts, K. & Shanks, D. (eds.), *Knowledge, Concepts, and Categories* (*forthcoming*). Psychology Press/MIT Press.

Hahn, U., Chater, N. & Henley, R. (1996). Weighting in Similarity Judgements: Investigating the MAX Hypothesis. In *Proceedings of the 18th Annual Meeting of the Cognitive Science Society*.

Haigh, K. & Shewchuk, J. (1994). Geometric Similarity Metrics for Case-Based Reasoning. In *Proceedings: 1994 AAAI workshop on case-based reasoning*. American Association for Artificial Intelligence.

Herbig, B. & Wess, S. (1992). Ähnlichkeit und Ähnlichkeitsmasse. In *Fall-basiertes Schliessen – Eine Übersicht*. SEKI Working papers SWP-92-08, University of Kaiserslautern, Germany.

Hintzman, D. (1986). "Schema abstraction" in a multiple-trace memory model. *Psychological Review* **93**: 411–428.

Janetzko, D., Melis, E. & Wess, S. (1993). System and Processing View in Similarity Assessment. In Richter, M., Wess, S., Althoff, K. & Maurer, F. (ed.), *Presentations and Posters: First European Workshop on Case-Based Reasoning*. SEKI Report SR-93-12 (SFB 314).

Kolodner, J. (1991). Improving Human Decision Making through Case-Based Decision Aiding. *AI Magazine*, 52–68.

Kolodner, J. (1992). An Introduction to Case-Based Reasoning. *Artificial Intelligence Review* **6**: 3–34.

Kolodner, J. (1993). *Case-based reasoning*. Morgan Kaufmann.

Levi, E. H. (1949). *An Introduction To Legal Reasoning*. University of Chicago Press.

Llewellyn, K. N. (1930). *The Bramble Bush: On Our Law and Its Study* (1960 edition). Oceana Publications.

MacCormick, N. (1987). *Legal Reasoning and Legal Theory*. Clarendon.

Medin, D., Dewey, G. & Murphy, T. (1983). Relationships Between Item and Category Learning: Evidence that Abstraction Is Not Automatic. *Journal of Experimental Psychology* **9**: 607–625.

Medin, D., Goldstone, R. & Gentner, D. (1993). Respects for Similarity. *Psychological Review* **100**: 254–278.

Medin, D. & Schaffer, M. (1978). Context theory of classification learning. *Psychological Review* **85**: 207–238.

Mendelson, S. (1989). An attempted dimensional analysis of the law governing government appeals. In *Proceedings of the Second International Conference on Artificial Intelligence and Law*. ACM.

Murphy, G. & Medin, D. (1985). The role of theories in conceptual coherence. *Psychological Review* **92**: 289–316.

Murray, J. L. (1982). The Role of Analogy in Legal Reasoning. *UCLA Law Review* **29**: 833–871.

Nosofsky, R. (1984). Choice, Similarity and the Context Theory of Classification. *Journal of Experimental Psychology: Learning, Memory and Cognition* **10**: 104–114.

Nosofsky, R. (1988). Exemplar-based Accounts of Relations between Classification, Recognition, and Typicality. *Journal of Experimental Psychology: Learning, Memory and Cognition* **14**: 700–708.

Nosofsky, R. (1990). Relations between Exemplar-Similarity and Likelihood Models of Classification. *Journal of Mathematical Psychology* **34**.

Nosofsky, R. (1992). Exemplars, Prototypes, and Similarity Rules. In Healy, A., Kosslyn, S. & Shiffrin, R. (eds.), *From Learning Theory to Connectionist Theory: Essays in Honor of William K. Estes*. Erlbaum.

Nosofsky, R. (1994). Rule-Plus-Exception Model of Classification Learning. *Psychological Review* **101**: 53–79.

Oaksford, M. & Chater, N. (1991). Against Logicist Cognitive Science. *Mind and Language* **6**: 1–38.

Pearl, J. (1988). *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*. Morgan Kaufmann.

Porter, B., Bareiss, R. & Holte, R. (1990). Concept Learning and Heuristic Classification. *Artificial Intelligence* **45**: 229–263.

Price, C. & Pegler, I. (1995). Deciding Parameter Values with Case-Based Reasoning. In Watson, I., Marir, F. & Perera, S. (eds.), *First United Kingdom Workshop on Case-Based Reasoning*.

Raz, J. (1979). *The Authority of Law*. Clarendon Press, Oxford.

Reber, A. S. (1989). Implicit Learning and Tacit Knowledge. *Journal of Experimental Psychology: General* **118**.

Redington, M. & Chater, N. (1995). Transfer in Artificial Grammar Learning: A Re-evaluation. *Journal of Experimental Psychology: General* forthcoming.

Riesbeck, C. & Schank, R. (1989). *Inside Case-based Reasoning*. Erlbaum.

Rips, L. (1975). Inductive Judgements about Natural Categories. *Journal of Verbal Learning and Verbal Behaviour* **14**: 665–681.

Rissland, E., Skalak, D. & Friedman, M. (1993). BankXX: A program to generate argument through case-based search. In *Proceedings of the Fourth International Conference on Artificial Intelligence and Law*. ACM.

Rosch, E. & Lloyd, B. (1978). Process. In Rosch, E. & Lloyd, B. (eds.), *Cognition and Categorization*. Erlbaum.

Ross, B. (1987). This Is Like That: The Use of Earlier Problems and the Separation of Similarity Effects. *Journal of Experimental Psychology: Learning, Memory and Cognition* **13**: 629–637.

Ross, B. (1989). Some Psychological Results on Case-Based Reasoning. In *Proceedings of the workshop on Case-Based Reasoning*, pp. 144–148.

Ross, B., Perkins, S. & Tenpenny, P. (1990). Reminding-based Category Learning. *Cognitive Psychology* **22**: 460–492.

Sanders, K. (1994). *CHIRON: planning in an open-textured domain*. Ph.D. thesis, Department of Computer Science, Brown University, Providence, R.I.

Sartor, G. (1991). The structure of norm conditions and non-monotonic reasoning in law. In *Proceedings of the Third International Conference on Artificial Intelligence and Law*. ACM.

Sartor, G. (1993). A simple computational model for non-monotonic and adversarial legal reasoning. In *Proceedings of the Fourth International Conference on Artificial Intelligence and Law*. ACM.

Sartor, G. (1995). Defeasibility in Legal Reasoning. In Bankowski, Z., White, I. & Hahn, U. (eds.), *Informatics and the Foundations of Legal Reasoning*. Kluwer Academic Publishers.

Schank, R. (1982). *Dynamic Memory: A Theory of Learning in Computers and People*. Cambridge University Press.

Shanks, D. & John, M. S. (1994). Characteristics of dissociable human learning systems. *Behavioral and Brain Sciences* **17**: 367–395.

Shepard, R. (1980). Multidimensional Scaling, Tree-fitting, and Clustering. *Science* **210**: 390–399.

Shepard, R. (1987). Toward a Universal Law of Generalization for Psychological Science. *Science* **237**: 1317–1323.

Spirtes, P., Glymour, C. & Scheines, R. (1993). *Causation, Prediction, and Search.* Springer Verlag.

Stone, R. (1994). *Contract Law*. Cavendish.

Susskind, R. (1987). *Expert Systems in Law*. Oxford University Press.

Tversky, A. (1977). Features of Similarity. *Psychological Review* **84**: 327–352.

Tversky, A. & Gati, I. (1978). Studies of Similarity. In Rosch, E. & Lloyd, B. (eds.), *Cognition and Categorization*. Erlbaum.

Tversky, A. & Gati, I. (1982). Similarity, separability and the triangle inequality. *Psychological Review* **89**: 123–154.

Weir, T. (1988). *A Casebook on Tort* (6 edition). Sweet and Maxwell.

Yang, S., Robertson, D. & Lee, J. (1993a). KICS: a knowledge-intensive case-based reasoning system for statutory building regulations and case histories. In *Proceedings of the Fourth International Conference on Artificial Intelligence and Law*. ACM.

Yang, S., Robertson, D. & Less, J. (1993b). Use of Case Based Reasoning in the Domain of Building Regulations. In Richter, M., Wess, S., Althoff, K. & Maurer, F. (eds.), *Presentations and Posters: First European Workshop on Case-Based Reasoning*. SEKI Report SR-93-12 (SFB 314).