

## **Discussion paper: how much of commonsense and legal reasoning is formalizable? A review of conceptual obstacles**

JAMES FRANKLIN

*School of Mathematics and Statistics, University of New South Wales, Australia*

[Received on 30 October 2011; revised on 9 March 2012; accepted on 19 March 2012]

Fifty years of effort in artificial intelligence (AI) and the formalization of legal reasoning have produced both successes and failures. Considerable success in organizing and displaying evidence and its inter-relationships has been accompanied by failure to achieve the original ambition of AI as applied to law: fully automated legal decision-making. The obstacles to formalizing legal reasoning have proved to be the same ones that make the formalization of commonsense reasoning so difficult, and are most evident where legal reasoning has to meld with the vast web of ordinary human knowledge of the world. Underlying many of the problems is the mismatch between the discreteness of symbol manipulation and the continuous nature of imprecise natural language, of degrees of similarity and analogy, and of probabilities.

The pioneers of artificial intelligence (AI) were gung-ho personalities, with little sense of ‘where angels fear to tread’, as pioneers generally are. They forged ahead vigorously, with extravagant promises based on a simple plan that they believed would formalize and hence automate all of human reasoning. Their strategy was based on two ideas:

- The representation of assertions about the world can be formalized by the predicate calculus or similar.
- Inference is the application of formal rules to the formal representations, plus search (through some space of possibilities).

Those were ideas for which there was some evidence. The predicate calculus of symbolic logic had had great success in formalizing mathematics; and it was possible for a computer program to learn how to play tic-tac-toe by searching through the space of possible moves.<sup>1</sup>

In some fields, those in 1960 who promised the moon in 10 years turned out to be right. But, in AI, it is 50 years later and the promises remain unfulfilled. Or rather, there is software such as Google, Excel and pathology test interpreters that perform many intelligent tasks that were not much considered in 1960, but the simplest commonsense and legal reasoning has proved intractable to formalization.

This article surveys what has been learned about the obstacles to formalization from five decades of experience. It will concentrate on representation more than on inference, since the main problems arise there. It will use mostly legal examples, though in general the difficult problems for legal reasoning have proved to be essentially the same as those of commonsense reasoning. That is because of the need

<sup>1</sup> Pamela McCorduck, *Machines Who Think* (2nd ed, Natick, MA: A.K. Peters, 2004); Marvin Minsky, Steps towards artificial intelligence, *Proceedings of the Institute of Radio Engineers* 49 (1961), 8–30.

for the law to interface with non-experts and to deal with the complex problems thrown up by real life. The close relation between legal and everyday reasoning was recognized from the early days of expert systems intended to formalize legal reasoning:

Because [legal problems] involve and describe our everyday activities they may require broad knowledge about the world and how things work. This need for commonsense reasoning over a broad range of activities and problems causes difficulties for expert system developers who are trying to keep the problem domain as narrow as possible and as free from the mysterious ‘commonsense reasoning’ and ‘world knowledge’ as possible.<sup>2</sup>

A list of problems that have come to light for formalization includes:

- The open-textured or fuzzy nature of language (and of legal concepts)
- Degrees of similarity and analogy
- The representation of context
- The symbol-grounding problem
- The representation of causation, conditionals and counterfactuals
- The balancing of reasons
- Probabilistic (or default or non-monotonic) reasoning (including problems of priors, the weight of evidence and reference classes)
- Issues of the discrete versus the continuous
- Understanding

These are all different problems, though as will appear there are connections and similarities. In each case the problems are formidable and far from completely solved, but significant progress has been made—progress that, even if it has not resulted in successful formalization, has at least deepened our understanding of legal and commonsense reasoning. There is also a reasonably clear separation between problems where AI has proved successful, such as organizing complex arguments and displaying the relation of the pieces of evidence in them, and those where it has proved unsuccessful, such as balancing the force of arguments.

### **Fuzziness, borderline cases and open texture**

The problem of fuzziness or vagueness or borderline cases is an easy one to appreciate. Many superficially classificatory concepts of ordinary language, like ‘tall’, ‘expensive’, ‘educated’, ‘reasonable’, in fact describe part of a continuum and are thus subject to borderline cases—a person may be unqualifiedly tall, or barely tall, or neither definitely tall nor short. That creates problems for automated reasoning as to the logical relation between, e.g. ‘The witness described X as quite tall’ and ‘X is of height 5 feet 8 inches’.

The basic idea of fuzzy logic is to represent borderline cases of concepts by a numerical function that expresses the degree to which, e.g. a person is tall. So someone unqualifiedly tall is said to be tall to degree 1, someone only slightly taller than short might be tall to degree 0.6, and so on. One should then

<sup>2</sup> Donald A. Waterman, Jody Paul and Mark Peterson, Expert systems for legal decision making, *Expert Systems* 3 (1986), 212–26, at 213–4; causal reasoning in law is especially reliant on commonsense methods, as described below.

translate vague linguistic terms into these ‘fuzzy membership functions’, compute with them, and if desired translate the results back into ordinary language.<sup>3</sup>

That is a satisfactory idea for representing vagueness formally, especially in the absence of any others, but it has proved very difficult to make it work in a way that at all adequately mimics how vague terms behave in ordinary reasoning. The reasons why it is difficult reflect wider problems with the formalization of ordinary reasoning.

First, if there is a fuzzy membership function for a natural-language concept like ‘tall’, it is very difficult to say what it actually is—what the numerical degree of tallness of a given individual is—e.g. by querying language users.<sup>4</sup> It is then hard to say how, e.g. the membership function for ‘tall’ should be modified to obtain the membership function for ‘very tall’, or how the membership function of ‘expensive’ for say real estate should relate to data on house prices. People are not conscious of having membership functions, and if there is an unconscious psychological reality to them, no recognized method has been discovered to extract it. Attempts to mimic logical relations like ‘if’ in terms of membership functions have not fared well either—e.g. to represent a statement like ‘If X is tall then X likely has high income’.<sup>5</sup>

A further problem is that membership functions are very sensitive to context: tall for a man is not the same as tall for a woman, a big mosquito is smaller than a small elephant, what is reasonable self-defense in one set of circumstances is an unreasonable preemptive first strike in another. Even quite young children can handle well-different contexts of comparison for the same word: a hat can be big as hats go, or big compared to a hat beside it, or big for the doll it is on.<sup>6</sup> The most commonly used and easily learned words tend to be relative words of this kind, and absolute and eternal concepts like ‘0.35 meters long’ come later. From the point of view of AI, it creates difficult problems, since representing context is itself a major problem (on which more later).

There is an obvious resemblance between fuzziness and the issues of the ‘open texture’ of legal concepts and the evaluation of the similarity of a case to precedents. It is hard to say whether an adequate formalization of fuzziness would solve or help with those problems.

The ‘open texture’ of legal concepts is certainly about the existence of borderline cases of concepts.<sup>7</sup> But it is not obvious that it is exactly the same problem as the fuzziness of natural language predicates.

<sup>3</sup> Introduction in George J. Klir and Tina A. Folger, *Fuzzy Sets, Uncertainty and Information* (Englewood Cliffs, NJ: Prentice Hall, 1988); legal applications in Jacky Legrand, Some guidelines for fuzzy sets application in legal reasoning, *Artificial Intelligence and Law* 7 (1999), 235–57.

<sup>4</sup> Review of methods in Taner Bilgic and I. Burhan Türksen, Measurement of membership functions: theoretical and empirical work, in D. Dubois and H. Prade, eds, *Fundamentals of Fuzzy Sets* (Norwell, Mass: Kluwer, 2000), ch. 3; introduction with a social science emphasis in Jay Verkuilen, Assigning membership in a fuzzy set analysis, *Sociological Methods and Research* 33 (2005), 462–96; the methods are complex and consistent results are hard to obtain. A rare, reasonably successful attempt to extract membership functions, but illustrative of the difficulties, in Thomas S. Wallsten, David V. Budescu, Amnon Rapoport, Rami Zwick, and Barbara Forsyth, Measuring the vague meanings of probability terms, *Journal of Experimental Psychology: General* 115 (1986), 348–65; the difficulties of extracting a membership function for ‘beyond reasonable doubt’ shown experimentally in Mandeep K. Dhami, On measuring quantitative interpretations of reasonable doubt, *Journal of Experimental Psychology: Applied* 14 (2008), 353–63.

<sup>5</sup> Satoro Fukami, Masaharu Mizumoto and Kokichi Tanaka, Some considerations on fuzzy conditional inference, *Fuzzy Sets and Systems* 4 (1980), 243–273–73; available formalisms for fuzzy implications described in Michał Baczyński and Balasubramaniam Jayaram, *Fuzzy Implications* (Berlin: Springer, 2008), but without much attention to agreement with natural language.

<sup>6</sup> Karen S. Ebeling and Susan A. Gelman, Children’s use of context in interpreting ‘big’ and ‘little’, *Child Development* 65 (1994), 1178–92.

<sup>7</sup> Herbert L.A. Hart, *The Concept of Law* (Oxford: Clarendon, 1961), 120–4; see Avishai Margalit, Open texture, in A. Margalit, ed., *Meaning and Use* (Dordrecht: Reidel, 1979), 141–52; Brian Bix, H.L.A. Hart and the ‘open texture’ of legal language, *Law and Philosophy* 10 (1991), 51–72.

In Hart's famous example, an ordinance 'No vehicles are allowed in the park' is doubtfully applicable to motorcycles or roller skates. The concept 'vehicle' is said to have a 'penumbra' of meaning. That could suggest a membership function for the concept 'vehicle': 1 for cars, maybe 0.5 for motorcycles, say 0.1 for roller skates. That may be so, as far as it goes, but it is not very helpful in itself for resolving problems that involve reasoning with open-textured concepts. Typically, there is some 'reason' why vehicles are forbidden in the park, a basis or intuitive *ratio decidendi* that will be to some degree mentally present to someone making a decision as to whether the rule applies to a case (e.g. the risk of vehicles injuring people may be relevant to their banning in the park, a risk that is low for roller skates). That reason may be much more relevant to the decision than the degree of membership of the borderline cases in the abstract—or perhaps one should say rather that the degree of membership is sensitive to the context of the reason. Either way, the problem of reasoning with open-textured concepts is not solved simply by appealing to fuzzy membership functions.

### Degrees of similarity and analogy

Similar problems arise with trying to compute the similarity of a case to precedents, if that is to be used in an effort to decide cases by formal reasoning. While similarity in general—say between colors—is a fuzzy concept, it may be that the similarity of cases to precedents is driven more by reasons, and typically the best reason defeats all the others on an all-or-nothing basis.

Similarity to precedent has been much discussed in the law-and-AI literature with reference to the case of *Popov v Hayashi*.<sup>8</sup> In 2001 a valuable home run ball was hit into the stands by a celebrated baseball player who hit a record-breaking 73rd home run. According to the custom of baseball, a fan who catches such a ball owns it. Alex Popov, a fan in the stands, partially caught the ball in his mitt, but before he could secure it in his hand, he was thrown to the ground by a mob of fans also trying to get the ball. The ball rolled loose and was picked up by Patrick Hayashi, who was not one of the mob that had knocked Popov down. Ownership of the ball was disputed between Popov and Hayashi.

The case obviously turns on possession, as it applies in cases where possession is almost or partly secured and then lost. The precedents considered in the case were mostly about hunting: in one case, merely hunting a fox with hounds was held to be insufficient to confer any right of possession, against someone else who killed the fox that was on the run. In another, a whale that was harpooned and washed ashore was found to be the property of the man who had harpooned it, not of the man who found it on the shore, account being taken of the customs and practices of whalers. There was a precedent with fish and another with ducks. The question is, how is a formalization of legal reasoning to recognize and represent these cases concerning wildlife as relevantly similar to the case involving a baseball? Plainly baseballs are not in themselves similar to foxes, so there must be some theory of relevant and irrelevant dimensions to a case.<sup>9</sup>

<sup>8</sup> *Popov v Hayashi*, 2002 WL 31833731 (Ca. Sup. Ct. 2002); legal commentary in Paul Finkelman, Fugitive baseballs and abandoned property: who owns the homerun ball? *Cardozo Law Review* 23 (2002), 1609–33, section V; Patrick Stoklas, *Popov v Hayashi*, a modern-day *Pierson v. Post*: a comment on what the court should have done with the seventy-third home run baseball hit by Barry Bonds, *Loyola University Chicago Law Journal* 34 (2002–3), 901–43; Michael Pastrick, When a day at the ballpark turns a 'can of corn' into a can of worms: *Popov v Hayashi*, *Buffalo Law Review* 51 (2003), 905–35.

<sup>9</sup> Discussed in T.J.M. Bench-Capon, Representing *Popov v Hayashi* with dimensions and factors, *Artificial Intelligence and Law*, 20 (2012), 15–35; Douglas Walton, Similarity, precedent and argument from analogy, *Artificial Intelligence and Law* 18 (2010), 217–46.

One suggestion for the form of such arguments from similarity or analogy, which usefully brings out what has to be done about dimensions of similarity, is:

Premise 1:  $a$  has features  $f_1, f_2, \dots, f_n$ .

Premise 2:  $b$  has features  $f_1, f_2, \dots, f_n$ .

Premise 3:  $a$  is  $X$  in virtue of  $f_1, f_2, \dots, f_n$ .

Premise 4:  $a$  and  $b$  should be treated or classified in the same way with respect to  $f_1, f_2, \dots, f_n$ .

Conclusion:  $b$  is  $X$ .<sup>10</sup>

Where is the information to come from to support Premises 3 and 4, i.e. the information that a feature  $f_i$  is relevant to a property  $X$ ? Must that be coded by humans who understand it, or is there a prospect of an AI system finding it for itself? The scope of the ambition of the AI-and-law project depends greatly on the answer to that question.

### The representation of context

Everyone knows that ‘context is important’. ‘We are swaddled by an aether of context,’ writes Doug Lenat, whose Cyc project was one of the most far-reaching in the formalization of commonsense reasoning.<sup>11</sup> In the hermeneutic circle, we understand a new part of a text that we are reading only in the context of the whole, while the understanding of the new item feeds back into our understanding of the whole.<sup>12</sup>

Once we have nodded sagely in agreement with those platitudes, then what? It has proved hard to go beyond vague generalities in discussing such an amorphous notion as ‘context’. But there are good reasons for trying to do so, if the AI project is to make progress.<sup>13</sup>

For example: How can an automated legal reasoning system classify animals?

The classification of animals in a legal context is quite different from that in a biological context. The main legal division is into tame or domestic animals (which have human owners who are liable for damage they cause) and wild ones (*ferae naturae*). The question arises in one of A.P. Herbert’s misleading cases, when someone throws snails over the neighbor’s fence, are snails *ferae naturae*?<sup>14</sup> That is a joke, but the real legal problems that have arisen over classifying animals are hardly less bizarre. Sheep are easy and wild lions are easy, but the behavior of some species means that whether they are wild depends on the context in which they are encountered. Bees are *ferae naturae*; when hived they become the qualified property of the person who hives them, but become *ferae naturae* again when they swarm. Parrots may become, but young unacclimatized parrots are not, ‘domestic animals’. A performing bear is not a domestic animal, nor is a caged lion or a tame seagull used in a photographer’s studio. The phrase ‘bird, beast or other animal, ordinarily kept in a state of confinement’ includes a ferret.<sup>15</sup> The reason why *ferae naturae* is not a closed list of species that can easily be

<sup>10</sup> Marcello Guarini, A defense of non-deductive reconstructions of analogical arguments, *Informal Logic* 24 (2004), 153–68, at 162; further in Federico Picinali, Structuring inferential reasoning in criminal fact finding: an analogical theory, *Law, Probability and Risk*, this issue.

<sup>11</sup> Douglas Lenat, The dimensions of context space, [www.cyc.com/context-space.doc](http://www.cyc.com/context-space.doc)

<sup>12</sup> Samuel W.K. Chan and James Franklin, Dynamic context generation for natural language understanding: a multifaceted knowledge approach, *IEEE Transactions on Systems A: Systems and Humans* 33 (2003), 23–41.

<sup>13</sup> James Franklin, The representation of context: ideas from artificial intelligence, *Law, Probability and Risk* 2 (2003), 191–9.

<sup>14</sup> Alan Patrick Herbert, *More Misleading Cases* (London: Methuen, 1930), ch. X.

<sup>15</sup> John S. James, ed., *Stroud’s Judicial Dictionary of Words and Phrases* (4th ed, London: Sweet and Maxwell, 1972), articles ‘Animal’, ‘Domestic animal’, ‘Ferae naturae’.

programmed into a system of automated legal reasoning is that the biological characteristics of the animals have to be read against a context of their relevance to the legal question of what kinds of animals can sensibly be regarded as human property.

Many of the hard lessons about the difficulty of representing context have been learned in attempts at translating natural languages.<sup>16</sup> Those difficulties transfer to other areas like representing the background knowledge needed to interpret commonsense and legal statements.

Here is a brief cartoon history of automatic language translation, indicating a few lessons that have been learned about context. In the 1950s, it seemed easy. The system looked up in a dictionary the words of the input language, say Russian, rearranged according to the grammatical rules of English, and printed the result. The results were farcical. An early example that gave optimists pause was ‘The pen is in the box’ versus ‘The box is in the pen’. These could both be reasonable assertions, but only if ‘pen’ means ‘writing instrument’ in the first sentence and ‘enclosure’ (as in ‘sheep pen’) in the second.<sup>17</sup> It was realized that disambiguation in such cases would require encyclopedic world knowledge, in this case knowledge of the typical relative sizes of objects. The problem is not rare, as (on average) the shorter the word, the more possible meanings it has. Some researchers despaired, others applied for huge grants. Now encyclopedic knowledge is in a loose sense ‘context’, in being outside information that has to be imported to a text to help in interpreting it. But it is not ‘context’ in the narrower sense in which we talk of ‘different contexts’. World knowledge is fixed, while the point of the contexts that handle anaphora resolution or indexicals is that they are volatile: the reason ‘I’ means me now and you later is that the immediate context has changed. To understand language well enough to translate it (e.g. or to make sense of witness statements), a system must call on not only shared and fixed background knowledge, but a moving ‘microtheory’ of ‘facts currently in play’.

Research has proceeded for decades, and language translation software is at the stage of being some use. But even now, the software that translates web pages on Google is weak exactly in the area of context. It often works, when it does work, simply by supplying enough correct lookups of words and phrases to allow the human reader to infer context and override the mistakes and gaps.<sup>18</sup>

Those examples suggest more questions than answers. In the abstract, there are several problems about representing context and inferring from and to it that are resistant to solution:

First, there is a ‘meshing’ problem: context and incoming information should work together, without one having all the voting strength. If new information is to cause major revision to a context, it needs to be strong to some definite degree. There must be a correct tuning to ensure the right winner when there is a standoff between an entrenched context and a challenge from a conflicting piece of data. (For example, if language understanding software is to deal correctly with ‘My sister’s kids are at the store again. Those three boys sure like candy,’ it must carry a context forward from the first sentence to the second, in order to identify ‘boys’ in the second with ‘kids’ in the first; but if the second sentence were ‘Those kids are at school’, there would be conflict between the context and the new information that would need resolution.<sup>19</sup>)

<sup>16</sup> One of the leaders of AI writes: ‘Almost all previous discussion of context has been in connection with natural language... However, I believe the main AI uses of formalized context will not be in connection with communication but in connection with reasoning about the effects of actions directed to achieving goals. It’s just that natural language examples come to mind more readily.’ (John McCarthy, *Artificial intelligence, logic and formalizing common sense*, in R.H. Thomason, ed., *Philosophical Logic and Artificial Intelligence* (Dordrecht: Kluwer, 1989), 161–90, at 180.)

<sup>17</sup> Yehoshua Bar-Hillel, The present status of automatic translation of languages, in *Advances in Computers*, vol. 1, ed. F.L. Alt (New York: Academic Press, 1960).

<sup>18</sup> E.g. Lee Gomes, Google Translate tangles with computer learning, *Forbes Magazine*, Aug 9, 2010.

<sup>19</sup> Chan and Franklin, *op. cit.* n. 12.



Second, there is the problem of ‘knowing what context’ one is in: the information to decide that is in some context, leading to an infinite regress problem. (For example, if a dialogue begins, ‘Waiter!’ ‘Sir?’, it is easy for a human to infer a likely context which will make a subsequent sentence like ‘There’s a fly in my soup’ expected and ‘We ride at dawn’ unexpected, but the substantial knowledge needed to infer that context must be imported.)

A problem with ‘multiple contexts’ for a single proposition, some proximate, some remote, some possibly in opposition to one another, some nested in one another. How is one to get hold of them all (or of the relevant ones), and then how can one make them all bear (each to its correct degree) on the task of dealing with incoming information?

The answers to those questions are unclear and the problem of context can appear intractable.<sup>20</sup> The AI community is not easily discouraged, and there has been considerable work on initial approaches.<sup>21</sup> An example from one of the most determined attempts to make progress shows how hard the problem is.

Doug Lenat’s Cyc project has taken to heart the lesson that commonsense reasoning depends on very extensive knowledge—about the world, what to expect of humans, popular culture, the different kinds of things there are, and so on. His plan is to identify all that knowledge, and have his large team of assistants type it in. ‘Nursing is what nurses do’—it will be there somewhere, or able to be quickly inferred from what is there.<sup>22</sup> Context is one branch of his vast plan, and—while it has not actually been made to work—the effort to implement it has made genuine progress and certainly revealed something about what it would take to complete the project.

Lenat first makes the point that a typical spoken sentence will make many assumptions about its context, and it is not productive to try to write down this open-ended list. For example, the instruction ‘If it’s raining outside, carry an umbrella’ implicitly assumes:

- The performer is a human being.
- The performer is not a baby, quadriplegic, dead and so on.
- The performer is about to go outside soon.
- The performer is not too poor to own an umbrella.
- We are not talking acid rain on Venus or Noah’s flood-sized rain.
- The performer is not hydrophobic, hydrophilic, Gene Kelly in love, etc.

(Legal examples on implicit reasonable exceptions to rules could easily be supplied.) Obviously, a human listening to the sentence does not have these assumptions in mind, and a computer system should not waste its time computing them either. Any one of them should be generated only if some question requires it. The question should stimulate some inference from some much simpler activated representation of context.

<sup>20</sup> Ben-Ami Sharfstein, *The Dilemma of Context* (New York: NYU Press, 1989).

<sup>21</sup> Introduction in Varol Akman and Mehmet Surav, Steps toward formalizing context, *AI Magazine* 17 (3) (Fall, 1996), 55–72; surveys in Patrick Brézillon, Context in artificial intelligence, *Computers and Artificial Intelligence* 18 (1999), 321–40 and 425–46; Paolo Bouquet, Chiara Ghidini, Fausto Giunchiglia and Enrico Blanzieri, Theories and uses of context in knowledge representation and reasoning, *Journal of Pragmatics* 35 (2003), 455–84; a bibliography at <http://www.eecs.umich.edu/~rthomaso/bibs/context.html>.

<sup>22</sup> Douglas Lenat and R.V. Guha, *Building Large Knowledge-Based Systems: Representation and Inference in the Cyc Project* (Reading, Mass: Addison-Wesley, 1990); a recent report in Douglas Lenat et al, Harnessing Cyc to answer clinical researchers’ ad hoc queries, *AI Magazine* 31 (3) (Fall, 2010), 13–32.

That active context, Lenat argues, is something like an imagined scenario.<sup>23</sup> It is not a fully fledged picture like a movie, swarming with extras and special effects. It is more like a choice of a point in a small-dimensional space (he suggests 12 dimensions) which is enough, by being in contact with one's permanent background world knowledge, to point to and call up when necessary many default assumptions about the context. Examples of these dimensions are time, type of time (as in 'just after eating'), culture ('idle-rich thirtysomething urban'), level of sophistication ('technical/scientific' versus 'two-year-old'), epistemology (who would know the facts in question) and so on. If some new information suggests further facts, they can be retrieved from background knowledge and added into the assumptions of the context, on a need-to-know basis. For example, if umbrellas are mentioned, facts about how they protect from rain are available, if but only if there is a call for them in building the scenario.<sup>24</sup>

Lenat's plan is heavily dependent on all commonsense knowledge about the world being already coded in the system. That is a serious problem for the Cyc project, since that has not yet been done despite the person-centuries of effort devoted to it.

The jury is still out on whether any AI system will be able to equal the performance of the reasonable man, whose role as a repository of commonsense facts is an essential resource of the law.<sup>25</sup> If so, it will not be soon.

### The symbol grounding problem

Plainly, questions of context tend to ramify indefinitely, since any fact 'in' a context has its own context. There is a danger that an excursus into such murky regions will prove to be an expedition from which one does not return.

An infinite regress threatens: If everything has to be understood in a context, how do we understand the context itself? The terms mentioned in the context: what gives them meaning? Do we need to represent the web of associations of those concepts, and so on indefinitely? If there is an outermost context, how is it grounded in reality? It is easy to find oneself proving that babies can never learn anything since they have nothing to start from; nevertheless they do learn (and faster than academics).<sup>26</sup> The problem comes close to Hubert Dreyfus's argument against the possibility of AI, to the effect that it is impossible to behave or reason in a human-like way without having the experience of growing up in a human body.<sup>27</sup>

<sup>23</sup> Lenat's work here thus connects with extensive research in cognitive psychology under the names of *mental models* (Philip N. Johnson-Laird, *Mental Models: Towards a cognitive science of language, inference and consciousness*, Cambridge: Cambridge University Press, 1983) and *bounded rationality* (Gerd Gigerenzer and Reinhard Selten, eds, *Bounded Rationality: The adaptive toolbox* (Cambridge, Mass: MIT Press, 2001) according to which most normal human inference does not follow the plan of symbolic logic and traditional AI of shuffling symbols by rules, but involves context-specific and relatively simple mental models.

<sup>24</sup> Lenat, The dimensions of context space, *op. cit.* n. 10, section 3.

<sup>25</sup> In A.P. Herbert's only slightly exaggerated caricature: 'No matter what may be the particular department of human life which falls to be considered in these Courts, sooner or later we have to face the question: Was this or was it not the conduct of a reasonable man? Did the defendant take such care to avoid shooting the plaintiff in the stomach as might reasonably be expected of a reasonable man? (*Moocat v. Radley* (1883) 2 Q.B.) Did the plaintiff take such precautions to inform himself of the circumstances as any reasonable man would expect of an ordinary person having the ordinary knowledge of an ordinary person of the habits of wild bulls when goaded with garden-forks and the persistent agitation of red flags? (*Williams v. Dogbody* (1841) 2 A.C.) I need not multiply examples . . .' Alan Patrick Herbert, *Uncommon Law* (London: Methuen, 1935), 2.

<sup>26</sup> Alison Gopnik, Andrew N. Meltzoff and Patricia K. Kuhl, *The Scientist in the Crib: Minds, Brains and How Children Learn* (New York: HarperCollins, 1999).

<sup>27</sup> Hubert Dreyfus, *What Computers Still Can't Do: A Critique of Artificial Reason* (Cambridge, Mass: MIT Press, 1992), ch. 7.



In the case of the meaning of words, this comes down to the symbol-grounding problem:<sup>28</sup> we cannot just define words by other words, for the same reason that we cannot learn the meaning of Chinese words using only a Chinese–Chinese dictionary. Somewhere, our use of the word ‘cat’ has to attach to, be learnt from and be triggered by our experiences of cats (and only cats). The symbol-grounding problem is usually posed in terms of robots, which need to tie their internal symbols to the input from sensors and the output to effectors such as arms; in that case the need for the connection between experience and symbols is clear. For expert systems, such as those for legal decision-making, the problem is not quite so obvious but is still present. A legal expert system contains technical terms, such as ‘possession’ or ‘probable cause’. The meaning of those terms should be the same as in a legal textbook, i.e. given by the writers and readers of the text (as the philosophers say, ‘parasitic on meanings in the head’). But the terms must also be tied to the system’s database of cases. (For example, the case *Popov v. Hayashi* must be tied to the term ‘partial possession’.) How are those connections to be made, and made correctly? Should they be made by hand or automatically, and if the latter, where can one start?

To what extent a solution to those intractable problems is needed depends on what ambition is chosen for a formal system of legal reasoning. It might be hoped to make legal reasoning a somewhat self-contained system and prevent it ramifying indefinitely into problems of commonsense reasoning. It is not easy to do that, simply because the point of law is to decide on problems that arise extra-legally, in ordinary life, and information about how they fit into ordinary life is essential to the legal outcome.

### Cause and conditionals

It is well known to philosophers, AI experts and legal theorists alike that reasoning involving causes, conditionals and counterfactuals is a minefield, in both commonsense and legal reasoning.<sup>29</sup> Let us take just a few examples of the problems they cause for formalization. Typical of where they arise in law is in negligence in torts: if the defendant negligently starts a fire which burns down the plaintiff’s house, the defendant is taken to have caused the loss, and the counterfactual is crucial: if the defendant had not started the fire (or ‘but for’ the fire), the plaintiff would not have suffered loss. But then, what if the fire started by the defendant merges with another fire caused by someone else, or by a lightning strike? What is the liability of the defendant, given that the house would have burned down anyway?<sup>30</sup>

The difficulties of formalizing such reasoning lie, before any legal complications begin, in representing straightforward counterfactuals. The standard theory of them involves possible worlds: ‘if the fire had not started, the house would not have burned down’ makes reference to a possible world which is like the actual one, but lacking the fire; in that world, nature takes its ordinary course and the house does not burn down. That requires some formalization of the ‘ordinary course of nature’ and ‘ordinary

<sup>28</sup> Stevan Harnad, The symbol grounding problem, *Physica D* 42 (1990), 223–46; Mariarosaria Taddeo and Luciano Floridi, Solving the symbol grounding problem: a critical review of 15 years of research, *Journal of Experimental and Theoretical Artificial Intelligence* 17 (2005), 419–45; philosophical perspectives in F. Adams and K. Aizawa, Causal theories of mental content, *Stanford Encyclopedia of Philosophy* 2010, <http://plato.stanford.edu/entries/content-causal/>.

<sup>29</sup> Conceptual overview in Antony Honoré, Causation in the law, *Stanford Encyclopedia of Philosophy*, 2001, revised 2010 <http://plato.stanford.edu/entries/causation-law/>; Peter Menzies, Counterfactual theories of causation, *Stanford Encyclopedia of Philosophy*, 2001, revised 2008 (<http://plato.stanford.edu/entries/causation-counterfactual/>); AI perspectives in Jos Lehmann, Joost Breuker and Bob Brouwer, Causation in AI and law, *Artificial Intelligence and Law* 12 (2004), 279–315; Rinke Hoekstra and Joost Breuker, Commonsense causal explanation in a legal domain, *Artificial Intelligence and Law* 15 (2007), 281–99.

<sup>30</sup> *Anderson v. Minneapolis, St. P. & S. St. R.R. Co.*, 146 Minn. 430, 179 N.W. 45 (1920), discussed in Fleming James and Roger F. Perry, Legal cause, *Yale Law Journal* 60 (1951), 761–811.

course of human affairs', which is difficult enough. But there is a further major difficulty with specifying how exactly the possible world without the fire differs from the actual world: it must differ from it, in some sense, minimally. David Lewis explains:

*'If kangaroos had no tails, they would topple over'* is true (or false, as the case may be) at our world, quite without regard to those possible worlds where kangaroos walk around on crutches, and stay upright that way. Those worlds are too far away from ours. What is meant by the counterfactual is that, things being pretty much as they are—the scarcity of crutches for kangaroos being pretty much as it actually is, the kangaroos' inability to use crutches being pretty much as it actually is, and so on—if kangaroos had no tails they would topple over.

We might think it best to confine our attention to worlds where kangaroos have no tails and *everything* else is as it actually is; but there are no such worlds. Are we to suppose that kangaroos have no tails but that their tracks in the sand are still as they actually are? Then we shall have to suppose that these tracks are produced in a way quite different to the actual way. Are we to suppose that kangaroos have no tails but that their genetic makeup is as it actually is? Then we shall have to suppose that genes control growth in a way quite different from the actual way (or else that there is something, unlike anything there actually is, that removes the tails). And so it goes; respects of similarity and difference trade off. If we try too hard for exact similarity to the actual world in one respect, we will get excessive differences in some other respect.<sup>31</sup>

Let us take another example, the 'pen vignette' considered in recent philosophy of causality, where the notion of cause is seen to involve a contrast with a background of normalcy:

The receptionist in the philosophy department keeps her desk stocked with pens. The administrative assistants are allowed to take pens, but faculty members are supposed to buy their own. The administrative assistants typically do take the pens. Unfortunately, so do the faculty members. The receptionist repeatedly emailed them reminders that only administrators are allowed to take the pens.

On Monday morning, one of the administrative assistants encounters Professor Smith walking past the receptionist's desk. Both take pens. Later that day, the receptionist needs to take an important message . . . but she has a problem. There are no pens left on her desk.

Experimental subjects are given this vignette and asked whether they agree with these statements:

- Professor Smith caused the problem.
- The administrative assistant caused the problem.

Overall, subjects agreed with the first and disagreed with the second.<sup>32</sup> The point of the example is that intuitions about cause, of the kind on which ethical and legal thought relies, are relative to a background of what is normal. In this case the normality is normative, since the example was set up so that the problem could have been avoided just as easily by either the administrative assistant or Professor

<sup>31</sup> David Lewis, *Counterfactuals* (Oxford: Blackwell, 1973), 8–9.

<sup>32</sup> Joshua Knobe and Ben Fraser, Causal judgment and moral judgment: two experiments, in W. Sinnott-Armstrong, ed., *Moral Psychology, vol 2: The Cognitive Science of Morality* (Cambridge, Mass: MIT Press), 441–8; discussion in Christopher Hitchcock and Joshua Knobe, Cause and norm, *Journal of Philosophy* 106 (2009), 587–612.

Smith having left the pen, but it was Professor Smith who should not have taken it. Again, the problem for formalization is that representing normalcy is an enormous task.<sup>33</sup>

### The balancing of reasons

The problems discussed so far have concerned the representation of facts and their relationships. Even if those problems were solved, there remain difficulties in formalizing ‘reasoning’ from those facts. While a good deal of reasoning has been successfully formalized, a crucial remaining problem concerns the ‘balancing’ of reasons (or cases, or reasons and cases). Suppose there are substantial reasons for a decision in a case, and a precedent that suggests an opposite decision. The judge may decide either ‘an ingenious argument has been put to me by Mr. Haddock for the defendant, which no doubt would produce substantial justice in this and many other cases. But he cited no authority which went the length for which he contended, and my duty is not to make new law but . . .’ OR ‘It is true, as counsel for the plaintiff pointed out, that no reported case can be found which precisely supports Mr. Haddock’s able argument for the defendant. But I think the principle of *Williams v. Dogsbody* is wide enough to cover this case . . .’<sup>34</sup> The metaphors of ‘length’ and ‘wide enough’ suggest that balancing is taking place on a continuous scale. If so, three questions arise:

- Who is to string this scale from case to principle? (or case to case, or . . .)
- How is distance on the scale to be measured and represented?
- Where is the information to come from that allows balancing to be done (according to the old legal maxim, ‘Arguments are to be weighed, not counted’) so as to calculate the position of the present case on the scale?

‘Judicial balancing’ has been discussed particularly in constitutional law, where judges find it natural to describe what they are doing as ‘balancing’ the (to some degree extra-legal) conflicting requirements of individual rights and state interests. ‘When rights and state interests, each with their claim to legitimacy, are perceived to be in collision, we are compelled toward “weighing” the “strength” of state interests against the “degree” of intrusion on individual rights. “Balancing” becomes the principle technique of judicial decision’.<sup>35</sup> Balancing is even harder if the judge is expected to compare extra-legal, semi-technical but unquantified considerations such as the desirability of different uses of public property<sup>36</sup> or the value of endangered species versus economic benefits.<sup>37</sup>

<sup>33</sup> Something on the complexity of how humans do it in Daniel Kahneman and Dale T. Miller, Norm theory: comparing reality to its alternatives, *Psychological Review* 93 (1986), 136–53.

<sup>34</sup> E.G. Coppel, The judicial development of the law, *Medico-Legal Society Proceedings* (Medico-Legal Society of Victoria), 1938, (<http://mlsv.org.au/files/1937-1938/The%20Judicial%20Development%20of%20The%20Law.pdf>) 151–75, at 165.

<sup>35</sup> Richard H. Pildes, Avoiding balancing: the role of exclusionary reasons in constitutional law, *Hastings Law Journal* 45 (1993–4), 711–51, at 711 (Pildes himself prefers a ‘ratio decidendi’ approach); explicit references to terror cases in Michel Rosenfeld, Judicial balancing in times of stress: comparing the American, British and Israeli approaches to the War on Terror, *Cardozo Law Review* 27 (2005–6), 2079–150; a judicial perspective on the psychology of balancing in Frank M. Coffin, Judicial balancing: the Protean scales of justice, *New York University Law Review* 16 (1988), 16–42; warnings on cultural differences in how balancing is done in Jacco Bomhoff, Balancing, the global and the local: judicial balancing as a problematic topic in comparative (constitutional) law, *Hastings International and Comparative Law Review* 31 (2008), 555–86.

<sup>36</sup> Joris Naiman, Judicial balancing of uses for public property: the paramount public use doctrine, *Boston College Environmental Affairs Law Review* 17 (1989–90), 893–920.

<sup>37</sup> Federico Cheever, Butterflies, cave spiders, milk-vetch, bunchgrass, sedges, lilies, checker-mallows and why the prohibition against judicial balancing of harm under the Endangered Species Act is a good idea, *William and Mary Environmental Policy Law and Policy Review* 22 (1997–8), 313–52.

The balancing of reasons is the topic that perhaps makes most obvious the wisdom of those who have insisted on the ‘tacit, ingrained and subterranean knowledge’<sup>38</sup> that lawyers and other humans rely on to reach conclusions. Oliver Wendell Holmes wrote long ago:

The language of judicial decision is mainly the language of logic . . . Behind the logical form lies a judgment as to the relative worth and importance of competing legislative grounds, often an inarticulate and unconscious judgment, it is true, and yet the very root and nerve of the whole proceeding. You can give any conclusion a logical form. You always can imply a condition in a contract. But why do you imply it? It is because of some belief as to the practice of the community or of a class, or because of some opinion as to policy, or, in short, because of some attitude of yours upon a matter not capable of exact quantitative measurement, and therefore not capable of founding exact logical conclusions . . .<sup>39</sup>

Leaving aside for the moment Holmes’ views about what matters exactly should be taken into account, it is undeniable that balancing of arguments leads inevitably into extra-logical considerations buried in the mind of the reasoner. Extracting and representing them remains one of the most serious obstacles to the formalization of commonsense and legal reasoning.

### Probabilistic and default reasoning

There is indeed one kind of balancing of reasons where a formalization has had great success. It is probabilistic reasoning concerning matters of fact, where the formalization based on calculating with numerical probabilities, using Bayes’ theorem, is standard, successful in many real cases, and logically well justified.<sup>40</sup> Bayes’ theorem concerns the updating of degrees of belief and is intended to give a complete recipe for belief revision. It has a complicated formula, but its simplest corollary gives its main message:

‘The verification of a (non-trivial) consequence renders a theory more probable’.<sup>41</sup>

For example: The detective reasons that if the butler did it, the knife will be behind the sofa. The knife is found behind the sofa, so the theory that the butler did it is better supported than it was before.

The reference to the ‘before’ state of belief is essential: the theorem explains only how to update a previous state. If, e.g. the theory that the butler did it is virtually ruled out by other evidence, the finding of the knife behind the sofa will not make it a credible theory.

While Bayesian theory is very sound, three major problems stand in the way of applying it formally and exactly to real arguments, especially in the messy areas of commonsense and legal reasoning. They are the specification of prior or background beliefs, the weight of evidence problem, and reference class problems.

<sup>38</sup> Peter Tillers, The structure and logic of proof in trials, *Law, Probability and Risk* 10 (2011), 1–6; the classic text is Michael Polanyi, *The Tacit Dimension* (London: Routledge, 1966).

<sup>39</sup> Oliver Wendell Holmes, The path of the law, *Harvard Law Review* 10 (1897), 457–78.

<sup>40</sup> James Franklin, The objective Bayesian conceptualisation of proof and reference class problems, *Sydney Law Review* 33 (2011), 545–61.

<sup>41</sup> The mathematician George Polya called this the ‘fundamental inductive pattern’: G. Polya, *Patterns of Plausible Inference* (Princeton: Princeton University Press, 1968), 3–4.

## Priors

First, where does one find the initial assignment of degrees of belief, to be updated with Bayes theorem? ('Initial' may mean either absolutely without assumptions,<sup>42</sup> or, more likely in real cases, the background beliefs to be assumed before consideration of the actual case starts.) 'Choose any one' is not correct, since if one begins with a dogmatic assignment—with some beliefs assigned probability zero or one, or almost—then no amount of evidence will dig one out of those prejudices. In short, to decide the credibility of a belief, the belief must be in a context of other relevant beliefs, which have been assigned degrees of belief that are not grossly unreasonable; the circularity is hard to transcend, as a matter of principle.

## The weight of evidence problem

A limitation of the Bayesian view is that the probability of a hypothesis on evidence is a measure of the 'balance' of reasons for and against a proposition, and hence does not distinguish between a balance of few reasons and a balance of many. As Suarez pointed out in the 16th century, there is a difference between 'negative doubt' (when there are no reasons for or against an opinion) and 'positive doubt' (when there are substantial reasons for and against, but they balance).<sup>43</sup> Positive doubt is more robust to new evidence; negative doubt is merely presumptive, as the scales can be tipped by any significant piece of evidence that comes to hand. The weight of evidence—the total amount of evidence, whether positive or negative—is not captured by probabilities, and there is no established method of quantifying it either (though there are some suggestions).<sup>44</sup>

Civil cases are decided on the balance of probabilities, and if there is very little evidence to balance, that is unfortunate but perhaps unavoidable.<sup>45</sup> But the use of probabilities of low weight in criminal cases is more worrying.<sup>46</sup> A probability of guilt of 0.9 reached through balancing a small amount of evidence is different from a probability of 0.9 based on a mass of evidence, because the chance discovery of a new minor piece of evidence could well reduce the first to 0.7 but is unlikely to do so for the second. One might therefore be rationally less willing to condemn a defendant to a heavy sentence on a probability of 0.9 of low weight than on a probability of 0.9 of high weight. The purely qualitative language of 'beyond reasonable doubt' could be argued to mean a doubt that is both large enough in probability and of sufficient weight to rely on.

<sup>42</sup> Leading to some classic problems about whether priors can be derived from symmetry arguments; Bayesian replies in Edwin T. Jaynes, *Probability Theory: The Logic of Science* (Cambridge: Cambridge University Press, 2003), ch. 15 and James Franklin, Resurrecting logical probability, *Erkenntnis* 55 (2001), 277–305. However, absolute priors are rarely relevant in legal reasoning, where the problem is more the richness of background information than the lack of it.

<sup>43</sup> James Franklin, *The Science of Conjecture: Evidence and Probability before Pascal* (Baltimore: Johns Hopkins University Press, 2001), 76–9.

<sup>44</sup> J. Maynard Keynes, *Treatise on Probability* (London: Macmillan, 1921), ch. 6; L. Jonathan Cohen, Twelve questions about Keynes' concept of weight, *British Journal for the Philosophy of Science*, 37 (1985), 263–78; Jaynes, *Probability Theory*, *op. cit.* n. 42, ch. 18; recent work in Picinali, *op. cit.* n. 10 and David Hamer, Probability, anti-resilience and the weight of expectation, *Law, Probability and Risk*, to appear. Dempster-Shafer theory also addresses the issue, by measuring separately the evidence for a proposition and the evidence against: Glenn Shafer, *A Mathematical Theory of Evidence* (Princeton: Princeton University Press, 1976).

<sup>45</sup> E.g. the reliance on symmetry in the absence of evidence in *T.N.T. Management v. Brooks* (1979) 23 A.L.R. 345, discussed in Richard Eggleston, *Evidence, Proof and Probability* (2nd ed, London: Weidenfeld and Nicolson, 1983), 184.

<sup>46</sup> Barbara Davidson and Robert Pargetter, Guilt beyond reasonable doubt, *Australasian Journal of Philosophy*, 65 (1987), 182–97; James Franklin, Case comment: United States v. Copeland, 369 F. Supp. 2d 275 (E.D.N.Y. 2005): quantification of the 'proof beyond reasonable doubt' standard, *Law, Probability and Risk* 5 (2006), 159–65.

Given the legal relevance of weight, the inability to measure it remains an obstacle to the formalization of the strength of legal arguments.

### Reference class problems

A recurring problem of probabilistic reasoning concerns the choice of the right reference class for a generalization. An example is the much-discussed *Shonubi* case. In *United States v. Shonubi*,<sup>47</sup> sentencing guidelines required an estimate of how much heroin Charles Shonubi, a Nigerian drug smuggler, had carried through New York's John F. Kennedy Airport (JFK) on seven previous trips during which he had been undetected. The estimate was based on the average amounts of heroin found on Nigerian drug smugglers caught at JFK airport in the time period. Why should that be used as the reference class relevant to the case, rather than, say, George Washington Bridge tollbooth collectors (Shonubi's day job)?

The problem is not confined to odd corners of the law but is common in all areas where general facts have to be taken into account. A more typical example might involve valuation: Valuing a house for sale involves estimating its price from the sale records for 'similar' houses. No other house is exactly similar to the given one, so how widely or narrowly should one choose the reference class of 'similar' houses, and on what criteria? Number of bathrooms? Age? Street number? Ethnicity of owner?

But even that gives little sense of the ubiquity of the problem. Legal reasoning relies heavily on generalizations, or in an older language, rebuttable presumptions of fact. Sworn witnesses generally tell the truth, notarized documents can be taken as valid unless there is evidence to the contrary...<sup>48</sup> All such generalizations are subject to a reference class problem: a particular case is in many reference classes, so how is one to know which are the relevant classes, and what should one do when some of them suggest different conclusions about the case? Without a solution, there is potential for endless argument between counsel on the relevance of different classes in any case involving statistical evidence or commonsense generalizations.<sup>49</sup>

John Venn introduced the problem in the nineteenth century,<sup>50</sup> and Hans Reichenbach gave it the name 'reference class problem', arguing:

If we are asked to find the probability holding for an individual future event, we must first incorporate the event into a suitable reference class. An individual thing or event may be incorporated in many reference classes, from which different probabilities will result. This ambiguity has been called the *problem of the reference class*.<sup>51</sup>

Plainly, the problem will appear in any case where there may be doubt as to what class containing an instance is most relevant to determining its probability of having some attribute. Philosophers, as is their way, have written at length on the ubiquity and difficulty of the problem, without offering a solution.<sup>52</sup> There is some consensus in philosophy that the reference class problem is inherently

<sup>47</sup> 895 F. Supp. 460 (E.D.N.Y. 1995), discussed in Peter Tillers, *If wishes were horses: discursive comments on attempts to prevent individuals from being unfairly burdened by their reference classes*, *Law, Probability and Risk* 4 (2005), 33–49.

<sup>48</sup> David A. Schum, *The Evidential Foundations of Probabilistic Reasoning* (New York: Wiley, 1994), 81–3.

<sup>49</sup> See articles in *International J. of Evidence and Proof*, special issue on the reference class problem, 11 (2007) issue 1.

<sup>50</sup> John Venn, *The Logic of Chance* (London: Macmillan, 1866), 176.

<sup>51</sup> Hans Reichenbach, *The Theory of Probability* (Berkeley and Los Angeles: University of California Press, 1949), 374.

<sup>52</sup> Henry E. Kyburg, *The reference class*, *Philosophy of Science* 50 (1983), 374–97; Alan Hájek, *The reference class problem is your problem too*, *Synthese* 156 (2007), 563–85; Mark Colyvan, Helen M. Regan and Scott Ferson, *Is it a crime to belong to a reference class?*, *Journal of Political Philosophy* 9 (2001), 168–81.



unsolvable, that ‘there is no principled way to establish the relevance of a reference class’.<sup>53</sup> AI researchers on commonsense reasoning have also come across the problem, and have also been nonplussed.<sup>54</sup>

It has been argued elsewhere that the problem is not quite as intractable as that, and in principle it has a solution in terms of the relevance of attributes, where relevance may be measured by covariance, the degree to which one variable ‘goes with’ another.<sup>55</sup> Nevertheless, applying any such ideas in a system of formalized legal or commonsense reasoning would be an enormous task, as it involves understanding the intuitive statistical evidence for a huge range of commonsense generalizations.

A number of different problems for formalization of reasoning have now been surveyed. Certain common themes emerge. There are two issues common to many of the problems: discrete versus continuous, and human understanding.

### Discrete versus continuous issues

Behind a number of the problems described above is a basic issue about the discrete versus the continuous. The contrast between the two is one of the great themes of mathematics.<sup>56</sup> The discrete and the continuous appeal to different kinds of minds. The continuous mind prefers the calculus, differential equations, smooth flows, gradual transitions, bell curves—the mathematics of the seventeenth to nineteenth centuries. The discrete mind prefers whole numbers, cryptography, symbolic logic, discrete approximations, computer code—the mathematics of the 20th century.

The ‘artificial intelligentsia’ are almost all discrete-oriented. With a background in symbolic logic, linguistics and computing, they tend to regard it as axiomatic that to formalize reasoning means to translate it into discrete symbols manipulated by formal rules. From that point of view, the continuous is at best an inconvenience and at worst something to be swept under the carpet. Lawyers, with their linguistic (though non-numerical) orientation, tend to share that bias. That is unfortunate when it comes to formalizing intuitive reasoning, since there are many aspects of reasoning to which continuity is essential. We have already seen fuzziness, where there is a smooth transition from ‘tall’ to ‘short’, with no firm cutoff and no clear way to divide the continuum with labels like ‘fairly tall’, ‘neither tall nor short but more tall than short’. Continuity was an essential feature of *Popov v Hayashi*: there is a spectrum of degrees of gaining possession of something, with hot pursuit of a fox being hardly at all in possession, having harpooned a whale being close to full possession, and Popov’s partially catching the baseball being about half way (if we accept the decision in the case, which was that Popov and Hayashi had equal interests in the ball.) We saw also probability, which varies continuously with evidence. Other examples of reasoning where continuity is essential include slippery slope arguments in ethics, counterfactual arguments involving closeness of possible worlds (such as the causation cases above), extrapolation arguments and John Stuart Mill’s ‘method of concomitant variation’.

<sup>53</sup> Mark Colyvan and Helen M. Regan, Legal decisions and the reference class problem, *International Journal of Evidence and Proof* 11 (2007), 274–85, at 275.

<sup>54</sup> Raymond Reiter and Giovanni Criscuolo, On interacting defaults, *Proceedings of the 7th International Joint Conference on Artificial Intelligence*, (1981), 270–6; Pei Wang, Reference classes and multiple inheritances, *International J. of Uncertainty, Fuzziness & Knowledge-Based Systems* 3 (1995), 79–91.

<sup>55</sup> James Franklin, Feature selection methods for solving the reference class problem: Comment on Edward K. Cheng, ‘A practical solution to the reference class problem’, *Columbia Law Review Sidebar* 110 (2010) 12–23.

<sup>56</sup> James Franklin, *What Science Knows: and How It Knows It* (New York: Encounter Books, 2009), 118–22.

Can reasoning involving continuity be formalized? It is certainly possible. Just use numbers. As in fuzzy logic and probability, the continuum can be represented by numerals (i.e. decimals), which measure the degrees on the continuum. In principle that is easy, and there are certainly areas where it works well, such as computer graphics—coordinates are calculated and the resulting display on the screen looks right. But there are many challenges when it comes to using numbers in reasoning where continuous variation is involved (some of them rehearsed in the long debate in legal scholarship over ‘Trial by mathematics’,<sup>57</sup> but applicable more generally):

- First, there is the simple challenge of forcing theorists to keep to the continuous without following the siren call of discrete language and backsliding into discrete shortcuts. For example, decades of work in AI on ‘non-monotonic logic’ and ‘default reasoning’ hoped to do probabilistic reasoning by discrete ‘default’ assumptions arranged in hierarchies (similar to legal presumptions, in being taken to be true unless there is evidence otherwise); it is finally largely agreed that there was truth in the legal maxim that ‘arguments are to be weighed, not counted’ and that there is only one right way to do probabilistic reasoning: Bayesian with numbers.<sup>58</sup>
- Secondly, if numbers are to be used in principle, where are the actual numbers in any case to come from? Who is to elicit where ‘half-tall’ is, in inches, and how? In the face of human resistance to attaching a number to the degree to which evidence in a given criminal trial supports the hypothesis of guilt, where is the number to come from, and who is to take responsibility for it? We saw above the difficulties for fuzzy membership functions, which ought to be no harder than any other case.
- Thirdly, although it is ‘possible’ to represent a continuum by decimal numbers and there is no adequate alternative, it is an inherently awkward way to do it. It is not how the brain does it and it is a poor solution for the kind of imprecise continua that are relevant to reasoning. The main problem concerns accuracy. Numbers are precise, but degrees of tallness, probabilities on evidence, closeness of possible worlds and so on are in their nature imprecise. It is ridiculous to identify the standard of proof beyond reasonable doubt with a probability of 0.937, and any other precise number is equally incredible. If one tries to represent the imprecision of numbers by some further numbers (for example by representing imprecise probabilities by intervals<sup>59</sup>) then matters are likely to deteriorate, both conceptually and computationally.

Only the human (and sometimes animal) brain has solved the problem of efficient calculation with the intuitive and the imprecise. How it does so is still unknown.

## Understanding

The AI project was premised on imitating one thing—human understanding—by something quite different—blind manipulation of uninterpreted symbols according to rules. Despite the inherent

<sup>57</sup> Laurence H. Tribe, Trial by mathematics: precision and ritual in the legal process, *Harvard Law Review* 84 (1971), 1329–93; Peter Tillers, Trial by mathematics – reconsidered, *Law, Probability and Risk* 10 (2011), 167–73.

<sup>58</sup> Works in the 1980s on the nonmonotonic logic approach included M. Ginsberg, ed., *Readings in Nonmonotonic Reasoning* (Los Altos: Morgan Kaufmann, 1987); the change was marked by such works as Judea Pearl, *Probabilistic Reasoning in Intelligent Systems* (San Francisco: Morgan Kaufmann, 1988); David J. Spiegelhalter, A. Philip Dawid, Steffen L. Lauritzen and Robert G. Cowell, Bayesian analysis in expert systems, *Statistical Science* 8 (1993), 219–47.

<sup>59</sup> Peter Walley, *Statistical Reasoning with Imprecise Probabilities* (London: Chapman and Hall, 1991); legal reasons for calculations with the imprecise, with an attempt to use Dempster-Shafer belief functions for the purpose, in Kevin M. Clermont, Death of paradox: the killer logic beneath the standards of proof, Cornell Legal Studies Research Paper No. 12-6 (Feb 2012). Available at SSRN: <http://ssrn.com/abstract=1986346>.

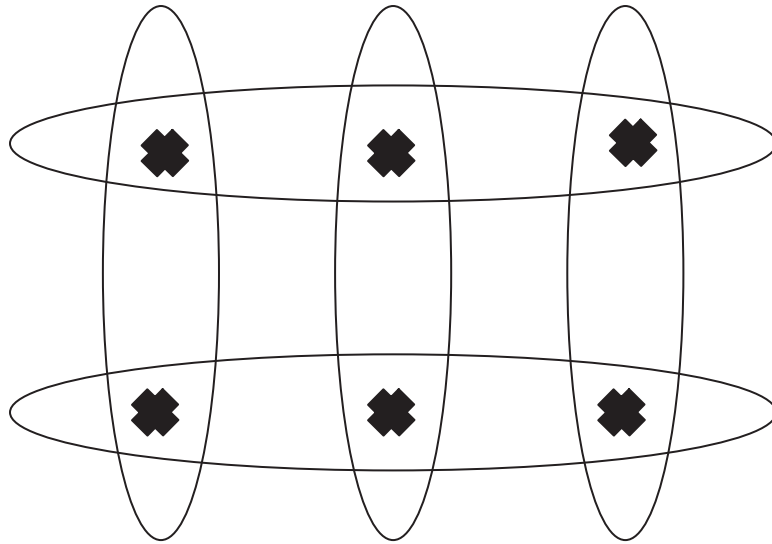


Fig. 1. Why  $2 \times 3 = 3 \times 2$ .

unlikelihood of success, there was reason to think it might be possible to do that always, because it was possible sometimes: namely in mathematical calculation. Addition, the preserve in ancient times of the most intelligent of the tribe, proved possible to teach to any seven-year-old with an abacus. Who could predict the limits of computation with rules?

After 50 years of experience, it is time to face the possibility that understanding is essentially quite unlike rule-following. We now have a great wealth of experience on what can be done with AI (in all its various forms) and the more experience we have, the more unique human understanding looks.<sup>60</sup> The mental act of really grasping meaning, and the necessary connections between one idea and another, is something that keeps escaping the kind of things that AI does.

Let us take a very simple example from mathematics, because that is the home ground of pure understanding (just as much as it is the home ground of calculation).

Why is  $2 \times 3 = 3 \times 2$ ? It is because two rows of three things are exactly the same things as three columns of two (Fig. 1). They are just the same things, considered differently. So we not only know ‘that’  $2 \times 3 = 3 \times 2$  but we ‘understand’ why it ‘must’ be so.

The prospects for putting genuine understanding of that kind into software or hardware are as close to zero as they were in 1950.<sup>61</sup>

<sup>60</sup> Linda Zagzebski, Recovering understanding, in M. Steup, ed, *Knowledge, Truth and Duty: Essays on Epistemic Justification, Responsibility and Virtue* (New York: Oxford University Press, 2001), 235–58.

<sup>61</sup> Which raises the problem of how the brain does it. Theories include that of Aristotle and the scholastics that the intellect has an immaterial power of grasping universals (Zdzislaw Kuksewicz, The potential and the agent intellect, ch. 29 of *The Cambridge History of Later Medieval Philosophy*, ed. N. Kretzmann et al., Cambridge, 1988), and Augustine’s theory of divine illumination (Robert Pasnau, Divine illumination, *Stanford Encyclopedia of Philosophy*, revised 2006, <http://plato.stanford.edu/entries/illumination/>), but theories suitable for a post-Enlightenment mentality are hard to find.

Understanding is equally important in legal and commonsense reasoning. In *Popov v Hayashi*, the description of the facts does not include the legal truth that the problem concerns possession and partial possession. There is a fundamental need to understand that, and then to understand that partial possession comes on a continuum, and then to understand that the actual case is about half-way on that continuum. In such cases, the human mind is able to perform the mysterious process of ‘abstraction’: of taking particulars and ‘separating in thought’ relevant properties of them. Of course it would be wonderful if a computer system could do that. But there is no plan on how it might do so.

The fact that software cannot do something essential to human reasoning does not mean that there will be no useful outcome of the formalization project. That is because there are different possible objectives for the project. In planning how to produce software that does something usefully intelligent, one may either:

- Ask the software to do all the work itself, including data interpretation, reasoning and communication (as in the traditional AI project), or
- Have humans input a great deal of knowledge, and have the software do the reasoning (as in Lenat’s Cyc), or
- Have humans do the reasoning, but use the software as a decision support tool by helping to organize and display the data (as in Excel or systems for diagramming complex arguments).
- Have the software harvest the results of intelligent human decisions (as in Google and Wikipedia).

The third and fourth plans have been the most successful to date. In AI as applied in law, most effort has been put into the third plan. A successful case study will show both the prospects, and how the obstacles described above impose limitations.

### **Case study: Prakken’s formalization of a Dutch legal case**

A great deal of work has been undertaken on the formalization of legal reasoning, with more modest ambitions but greater real success than in pioneering days. We examine one state-of-the-art example of such work, in order to see how the conceptual obstacles described above relate to what has actually been accomplished in formalization. Without an examination of a serious possible counterexample, the above account of obstacles to formalization risks comparison with the (apocryphal) proof that it is aerodynamically impossible for the bumblebee to fly. Furthermore, a case study makes it possible to see where the line lies between what has and what has not been accomplished by formalization.

We choose Prakken’s formalization of the reasoning in a routine Dutch case of disputed possession.<sup>62</sup> Some detail is necessary to appreciate the issues. A certain van der Velde owned a large tent at a camp site. Nieborg, the plaintiff, and his wife were interested in buying the tent but could not afford it. Van der Velde made the tent available to them, and in return the Nieborgs did work around van der Velde’s house. Nieborg claimed after some time that the work done was enough to pay for the tent. Van der Velde was angry and demanded the tent back, saying there was no agreement for its sale. When Nieborg refused to return the tent, van der Velde and others seized it forcibly. Van der Velde then sold it to a third party. Ownership of the tent is disputed between Nieborg and the third party.

<sup>62</sup> Henry Prakken, Formalising ordinary legal disputes: a case study, *Artificial Intelligence and Law*, 16 (2008), 333–59.

Several explicit rules of Dutch law are relevant:

- Possession of a good ‘in good faith’ is sufficient for ownership.
- As an exception to that, if some other person was the owner of the good less than three years ago and involuntarily lost possession (e.g. by theft) then that person is still the owner. (As far as proof goes, that means that there is a presumption that possession confers ownership, and the burden of proof is on someone who claims the exception: that possession was lost involuntarily.)
- Holding a good (meaning not physical holding but holding ‘as if one were the owner’) creates a presumption of possession.
- As an exception to that, if the holding started as ‘holding for someone else’, then it does not create a presumption of possession.

The main legal issue was whether van der Velde had sold the tent to Nieborg (in which case Nieborg would have been owner when forcibly dispossessed).

Prakken models the legal rules by defeasible implication rules such as:

$$\begin{aligned} & \text{PossessesInGoodFaith}(x,y,t) \Rightarrow \text{Owner}(x,y,t) \\ & \text{PossessesInGoodFaith}(x,y,t) \wedge \text{Owner}(z,y,t') \wedge (t - t') < 3\text{years} \\ & \wedge \text{InvoluntaryLoss}(z,y,t') \Rightarrow \neg\text{Owner}(z,y,t')^{63} \end{aligned}$$

The symbol  $\Rightarrow$  indicates a defeasible implication, i.e. a rule that applies unless some other rule is also applicable. In case of conflict, which rule applies is decided by a separate system of priority rules—e.g. in the case of the two rules just given, the second states an exception to the first so the second takes priority.

Much of Prakken’s work models the turn-taking and decision rules of the case—e.g. modeling the structure of the legal argument back and forth so that it becomes clear which statement replies to which, and how the judge rules on the evidence and argumentation to date. That modeling is very successful in formalizing the overall structure of the case.

After those legal aspects have been considered (both in the original case in court, and in Prakken’s formalization), it becomes clear that the main issue to be decided is one of (mental) fact: whether van der Velde had intended to sell the tent to Nieborg in consideration of the work of Nieborg and his wife (and had informed Nieborg of that intention). On this issue, the case moved to consider the evidence of witnesses. The witnesses were van der Velde and Nieborg, who said the opposite to each other, and two further witnesses who reported things that Nieborg had told them. In summary, the witness evidence was:

- Van der Velde said that Nieborg expressed his gratitude (for the loan of the tent); that he offered to do work after accepting the tent; that he (van der Velde) would have offered the tent without any work being done; and that he had been angry when Nieborg claimed he had done enough work to pay for the tent.
- Nieborg claimed that his and his wife’s work was in payment for the tent.
- The two other witnesses, Gjaltema and van der Sluis, reported that Nieborg had expressed his gratitude towards van der Velde; van der Sluis added that Nieborg had said that the tent made it possible for him to go on holiday in summer when he could not otherwise afford it.

<sup>63</sup> Note the potential for symbol-grounding problems in the common AI practice of using atomic names like *PossessesInGoodFaith* which have meaning to humans but not to the computer system.

- Nieborg attempted to undermine the sincerity of the three hostile witnesses, saying that as an expression of gratitude, the amount of work done was excessive, and that a travel ticket showed that the Nieborgs were able to afford holidays earlier, undermining the witnesses' claims that they could not afford a holiday. Nieborg said also that the true reason for van der Velde's anger was dissatisfaction with the quality of Nieborg's painting work, but this claim was not backed by any evidence.

In his decision, the judge summed up his view of the evidence as:

On the basis of the three witness testimonies of van der Velde, Gjaltema and van der Sluis, when considered jointly and in their mutual relations, the court regards as proven that Nieborg had at 5 July 1974 obtained the tent on loan . . . [Although the word 'loan' was not used] the court regards as decisive that the witnesses speak of 'to make use of' . . . 'use' . . . and 'to give in use'.

That the use was free is proven by the testimony of van der Velde, who in this context explicitly uses the term 'free', combined with the gratitude shown by Nieborg as mentioned by all three witnesses and his remark to the witnesses Gjaltema and van der Sluis that receiving the tent made it possible for him and his wife to go on holiday that year.

Prakken offers a formalization of the arguments concerning the witness evidence. On witness evidence in general, he lays down this defeasible rule:

Witness W says that  $\varphi$ ; therefore (presumably)  $\varphi$

and has the following critical questions for the rule's applicability:

- (1) Was W in the position to know about  $\varphi$ ?
- (2) Is W sincere?
- (3) Did W's senses function properly when observing  $\varphi$ ?
- (4) Did W's memory function properly when testifying that  $\varphi$ ?

In the case at hand, the main question is whether Van der Velde and Nieborg were sincere. Prakken is again successful in listing discrete items of evidence, and explaining which of them bear negatively and positively on others in the dialogue between the two sides.

However, Prakken makes no attempt to model the crucial matter of the strength of the support which the testimony of Gjaltema and van der Sluis gives to van der Velde's sincerity (and hence to his overall account); nor of the degree to which Nieborg's failure to respond to this evidence weakens his case. Prakken comments that at this point 'the judge's reasoning seems clearly flawed or at least incomplete'. It is true that the judge's reasoning is incomplete, but that is because the reasoning depends on an intuitive grasp of how strong the support is, in the light of the judge's understanding of human nature. A judgement of whether the Nieborgs' work is 'excessive' as an expression of gratitude is an intuitive judgement on a continuous scale (of an inferred psychological quantity). Whether the three witnesses' evidence 'combines with' the strength of the Nieborgs' gratitude to provide strong evidence of the truth of van der Velde's claim that the tent was merely loaned is again an intuitive judgment on a continuous scale (of probability).<sup>64</sup> Those are the crucial judgments on which the decision in the case

<sup>64</sup> Prakken's views on the 'accrual' or combining of several reasons for the same conclusion in Henry Prakken, A study of accrual of arguments, with applications to evidential reasoning, in *Proceedings of the Tenth International Conference on Artificial Intelligence and Law (ICAIL'05)*, Bologna, 2005, 85–94.



rests. The fact that they remain beyond the reach of formalization shows why the formalization of legal reasoning in real cases is very far from being complete.

Prakken comments, on the incompleteness of his formalization:

An important representation issue is that of unexpressed premises. . . . civil procedure allows the adversaries to leave the applicable law and commonsense knowledge implicit, and requires the judge to complete such incomplete arguments. However, the present combination of dialogue system and logic does not allow for logically incomplete arguments, and therefore this feature of the dispute cannot be modelled.

That is so, but the problem is not merely the incompleteness of the arguments in the sense that premises are left implicit. Even when the premises are identified (as in ‘the Nieborgs’ gratitude was excessive’) the problem remains that intuitive judgments on a continuous scale, made in the light of common knowledge of typical human nature, are resistant to formalization.

### Conclusion

Fifty years of experience with AI, and examination of working systems like Prakken’s, suggest two conclusions on the prospects for AI in law, one positive and one negative:

First, legal argumentation does contain a discrete structure (of items of evidence, relevance of one to another, legal rules, turn-taking of parties in a case, judicial determinations and so on), which formal methods can usefully represent, order and display.<sup>65</sup>

Second, once that has been done, the further step of evaluating the strength of those arguments, in the light of common human knowledge, involves intuitive judgments of degree based on genuine understanding; those tasks are, on present evidence, beyond the methods of AI.

<sup>65</sup> This conclusion thus supports the use of diagrammatic methods to organize evidence and legal reasoning, as they allow the discrete structure of propositions and their relations to be displayed in ways easy for humans to grasp. See for example Tim van Gelder, *The rationale for Rationale, Law, Probability and Risk* 6 (2007), 23–42, and other papers in the same special issue.