

The representation of context: ideas from artificial intelligence

JAMES FRANKLIN

School of Mathematics, University of New South Wales, Sydney 2052, Australia

[Received on 2 April 2003; revised on 19 August 2003]

To move beyond vague platitudes about the importance of context in legal reasoning or natural language understanding, one must take account of ideas from artificial intelligence on how to represent context formally. Work on topics like prior probabilities, the theory-ladenness of observation, encyclopedic knowledge for disambiguation in language translation and pathology test diagnosis has produced a body of knowledge on how to represent context in artificial intelligence applications.

Keywords: context; artificial intelligence; open texture; legal reasoning; Cyc.

Everyone knows that ‘context is important’. ‘We are swaddled by an aether of context’, writes Doug Lenat.¹ In the hermeneutic circle, we understand a new part of a text only in the context of the whole, while the understanding of the new item feeds back into our understanding of the whole. And so on.

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 reasons for trying to do so.

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 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 neighbour’s fence, are snails *ferae naturae*? That is a joke, but the real problems that have arisen under the heading ‘What is an animal?’ are hardly less bizarre. Sheep are easy and wild lions are easy, but the behaviour 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.² The reason why *ferae naturae* is not a closed list of species that can easily be programmed into a system of automated legal reasoning is

¹ LENAT, D. The dimensions of context space, www.ai.mit.edu/people/phw/6xxx/lenat2.pdf

² JAMES, J. S. (ed.) 1986 *Stroud’s Judicial Dictionary of Words and Phrases*, 5th edn, London, articles ‘Animal’, ‘Domestic animal’, ‘Ferae naturae’.

that the biological characteristics of the animals have to be read against a context of their relevance to the legal situation of what kind of animal can sensibly be regarded as human property.

Legal theorists are well aware of this ‘open texture’ of legal concepts,³ but that phrase merely points to the problem, and perhaps suggests that the difficulty lies mostly in the existence of borderline cases of concepts. The question is what exactly has to be done to make an automated system ‘aware’ of the problem and able to give the right decision. It may be that we do not strictly need automated legal reasoning, which might lead to such undesirable effects as unemployment in the legal profession. But if we cannot automate it, that is evidence that we do not understand it in the first place.

There are two perspectives on the subject of context, as on so many others: a humanist and a scientific one. The humanist is sensitive to the infinite complexities of the topic and all the possible difficulties of approaching it, while the scientist is crude, simplistic, optimistic and gets something done. Both perspectives are necessary, if any progress is to be made.

The most relevant science is Artificial Intelligence, where the discipline of programming fiercely exposes anything that has been overlooked, or covered up in the waving of hands.

Let us briefly survey several areas, mostly well-known, where the problem of context raises its ugly head and eats the naïve theorist for breakfast. We will keep to simple illustrative examples. The ideal question to keep in mind while taking in these examples is not so much the standard humanist one ‘Is this really different from something I know about already?’ but the scientific one ‘How could I start programming software that would deal with this?’

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.’

(The mathematician George Polya calls this the ‘fundamental inductive pattern’.⁴)

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, for example, the theory that the butler did it was virtually ruled out by other evidence, the finding of the knife behind the sofa would not make it a credible theory.

So where does one find an initial assignment of degrees of belief? ‘Choose any one’ is not correct, since if one begins with a dogmatic assignment—with some beliefs assigned

³ HART, H. L. A. 1961 *The Concept of Law*, Oxford, pp. 120–124; see MARGALIT, A. 1979 Open texture. (A. Margalit ed.), *Meaning and Use*, Dordrecht, pp. 141–152; BIX, B. 1991 H. L. A. Hart and the ‘open texture’ of legal language. *Law and Philosophy*, **10**, 51–72.

⁴ POLYA, G. 1969 *Patterns of Plausible Inference*, Princeton, pp. 3–4.

probability zero or one, or almost, then no amount of evidence will dig one out. 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.

One of the principal uses of prior probabilities as contexts has been in the defence objective Bayesians have made to the objection from alleged paradoxes. Objective Bayesians, following Keynes, hold that there is one and only one correct answer to the degree to which a given body of evidence supports a hypothesis, and that this is a matter of pure logic. The main objection to this thesis has come from alleged paradoxes of prior probabilities: for example, should the prior probability that a crow is white be a half (because it could be white or not white) or a third (because it could be white, black or neither), or something else? Objective Bayesians have been ingenious in arguing that if we work with the prior beliefs we actually have, instead of with the infinity of prior beliefs that we might have had, we can reach a unique answer.⁵

The thesis of the *theory-ladenness of observation* greatly troubled the philosophy of science a few decades ago. Against inductivists like Carl Hempel who argued for a bottom-up building of scientific theory out of neutral observational data, Norwood Russell Hanson argued that observations could not be seen at all except in the context of some theory—without a theory, they were just junk and not observations of anything in particular. The debate then moved to cognitive science, where it seemed that the experience of visual illusions showed that the visual system forced interpretations on some data, contrary to what the data (so to speak) wished to say.⁶ The way the debate worked was that those emphasizing the theory-ladenness of observation were in favour of relativism about scientific knowledge while the inductivists were realists, but it is not so clear in retrospect that that is the right way to see the issues. A realist view of science could be happy with theory-laden observation, and even praise its advantages, such as speed in decision-making, provided that observation can really bear on theories and speak in favour of the true ones.

Plainly, there are opportunities for relativists of all kinds to make hay out of abstract talk about ‘contexts’ and how we cannot escape them. Such thoughts will all be versions of what the Australian philosopher David Stove identified as ‘the worst argument in the world’:

We can know things only

- as they are related to us,
- under our forms of perception and understanding,
- insofar as they fall under our conceptual schemes,
- etc.

So,

We cannot know things as they are in themselves.

This argument is invalid, so no more needs to be said.⁷

⁵ JAYNES, E. T. 2003 *Probability Theory: The Logic of Science*, Cambridge, Chapter 15.

⁶ ESTANY, A. 2001 The thesis of theory-laden observation in the light of cognitive psychology. *Philosophy of Science*, **68**, 203–217.

⁷ FRANKLIN, J. 2002 Stove’s discovery of the worst argument in the world. *Philosophy* **77**, 615–624; cf.

Linguistic context is the central and best known kind of context. Well-studied examples include:

- Indexicals: the meaning of ‘I’ depends on extra-linguistic context identifying the speaker. (‘The meaning of a word like “I” is a function that takes us from the context of utterance to the semantic value of the word in that context.’⁸) The interpretation of words like ‘yes’ or gestures like a wink across the room obviously depends heavily on context.

- Anaphora resolution, or the reference of pronouns: in

‘Clariss will concentrate on *its* Filemaker Pro database software, changing *its* name to Filemaker’.

the different referents of the two ‘its’ are inferred from the context, as well as some outside knowledge about the likelihood of different changes of name.

- The problem of disambiguation in general. The ambiguity in

‘The chickens are ready to eat’

could be resolved by a hint in the previous sentence, such as ‘The oven bell has sounded.’

- The problem of relative words: not only is a big mosquito smaller than a small elephant, there are different possible contexts that have an effect. 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.⁹ The most commonly used and easily learned words tend to be relative words of this kind, and absolute and eternal concepts like ‘0.35 metres long’ come later. From the point of view of AI, it is most annoying.

- The problem of how much text to take in before reaching a decision on its meaning. One can very easily make mistakes from taking a small piece of text ‘out of context’, as when the British comedy *Not the Nine O’Clock News* had a Tory policy speech quote the Bible thus:

‘It is easier for a rich man to enter the kingdom of heaven than for a camel to.’

The evils of fundamentalism in scriptural interpretation and legal interpretation are well known. Fundamentalism is a determination to do local interpretation with inadequate attention to the global context of what the whole text is aiming at.

- The question of how the *beginning* of a text allows inference of a likely context. For example, if a dialogue begins:

DEMBSKI, W. A. 1994 The fallacy of contextualism. *Princeton Theological Review*, Oct., (www.arn.org/docs/dembski/wd_contextism.htm)

⁸ RECANATI, F. 1993 *Direct Inference: From Language to Thought*, Oxford, p. 235.

⁹ EBELING, K. S. & GELMAN, S. A. 1994 Children’s use of context in interpreting ‘big’ and ‘little’. *Child Development*, 65, 1178–1192.

‘Waiter!’

‘Sir?’

one will infer a context including two speakers a few feet apart, speaking the same language, taking place in a historical time period after the invention of restaurants, both males with the second probably aged at least 15; the next sentence is expected to be an order or complaint, and so on. (There are similarities with the legal problem of implied rights in constitutions, where the text would not make sense but for certain assumptions that are not explicitly stated in it; anyone who makes the assumptions explicit can expect complaints from fundamentalists.¹⁰)

One could go on, but enough of the obvious has already been stated.

So in a computer translation system, how can context in such problems be represented, that is, given the form of some code that can contain the knowledge of the context so as to bear on the following stream of incoming words?¹¹

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 that was that. 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.¹² It was realized that disambiguation in such cases would require encyclopedic world knowledge, for example of the expected 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, 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. Success is hourly expected. But even now, the software that translates web pages on google is weak exactly in the area of context. It works by supplying enough correct lookups of words and phrases to allow the human reader to infer context and override the mistakes and gaps.

AI researchers tend to believe that language understanding, though it is a good test bed and source of examples, is not quite the main game. A leader in the field

¹⁰ CLAUS, L. 1995 Implication and the concept of a constitution. *Australian Law Journal*, **69**, 887–904; CRAVEN, G. 1999 The High Court of Australia: A study in the abuse of power. *University of New South Wales Law Journal*, **22**, 216–242.

¹¹ CHAN, S. W. K. & FRANKLIN, J. 2003 Dynamic context generation for natural language understanding: a multifaceted knowledge approach. *IEEE Trans. Syst. A.*, **33**, 23–41.

¹² BAR-HILLEL, Y. 1960 The present status of automatic translation of languages. *Advances in Computers*, (F. L. Alt ed.), New York.

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.'¹³

As everyone knows, many *jokes* work by building up and then upsetting a context of expectation.

EXAMPLE Charlie Chaplin was asked how to film a fat lady slipping on a banana peel. He said 'You show the fat lady approaching; then you show the banana peel; then you show the fat lady and the banana peel together; then she steps over the banana peel and disappears down a manhole.'¹⁴

If there were joke-understanding software, it would need to represent those contexts and understand how the punchline was in conflict with them.

In cognitive psychology, there has been work under the names of *mental models*¹⁵ and *bounded rationality*.¹⁶ It is maintained that 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.

EXAMPLE To work out how to be in the right place to catch a ball, it is not normally necessary to use the laws of mechanics to calculate the trajectory of the ball and the path of oneself. All it needs is to follow the special-purpose heuristic: run so as to keep the angle of gaze constant.¹⁷ This heuristic is wrong or meaningless outside this special context, but inside its context, it delivers good performance very fast.

The examples above suggest more questions than answers. In the abstract, there are several problems about representing context and inferring from and to it that seem quite unsolvable:

- An *infinite regress* problem. If everything has to be understood in a context, how do we understand the context? 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 a lot faster than academics).¹⁸ In the linguistic case, this comes down to the *symbol-grounding problem*:¹⁹ we cannot just define words by other words, for the same reason as 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/be triggered by our experiences of cats (and

¹³ MCCARTHY, J. 1989 Artificial intelligence, logic and formalizing common sense. (R. H. Thomason ed.) *Philosophical Logic and Artificial Intelligence*, Dordrecht, pp. 161–190, p. 180.

¹⁴ www.geocities.com/Athens/Delphi/9910/chaplin.html

¹⁵ JOHNSON-LAIRD, P. N. 1983 *Mental Models: Towards a Cognitive Science of Language, Inference and Consciousness*, Cambridge, MS.

¹⁶ GIGERENZER, G. & SELTEN, R. (eds) 2001 *Bounded Rationality: The Adaptive Toolbox*, Cambridge, MA.

¹⁷ GIGERENZER, G. & SELTEN, R., *Bounded Rationality*, p. 7.

¹⁸ GOPNIK, A., MELTZOFF, A. N. & KUHL, P. K. 1999 *The Scientist in the Crib: Minds, Brains and how Children Learn*, New York.

¹⁹ HARNAD, S. 1990 The symbol grounding problem. *Physica D*, **42**, 223–246, also at www.ecs.soton.ac.uk/~harnad/Papers/Harnad/harnad90.sgproblem.html

only cats). Perhaps we understand how a thermostat does it, but human cognition and language is not much like a thermostat.

- A *meshing* problem: context and new information should work together, without one having all the voting strength. If new information is to cause regime change 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 an upstart piece of data. The reason is that a probability 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 sixteenth 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).²⁰ 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.
- A problem of *knowing what context* one is in: the information to decide that is in some context, leading to another infinite regress problem.
- A problem with *multiple contexts*, some proximate, some remote, some possibly in opposition to one another, some nested in one another. How is one to grab hold of them all, and then how can one make them all bear (each to its correct degree) on the task of dealing with incoming information?

That has let the humanists wring their hands to their hearts’ content. Obviously, the problem is quite intractable.²¹ Now, can the scientists come up with something simple that does something?²²

Here are two interesting examples from the artificial intelligence lab. Plan A is a frontal assault, designed to do everything, and do it the way humans do it. Plan B is something much simpler, which performs a specific task in a way that makes the interpretation depend on a simple proxy for context.

Plan A. Doug Lenat’s Cyc project has taken to heart the lesson that commonsense reasoning depends on a lot of 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 in there somewhere, or able to be quickly inferred from what is in there. 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.

²⁰ FRANKLIN, J. 2001 *The Science of Conjecture: Evidence and Probability before Pascal*, Baltimore, pp. 76–79; modern theory in JAYNES *Probability Theory*, Chapter 18.

²¹ SHARFSTEIN, B.-A. 1989 *The Dilemma of Context*, New York.

²² Surveys in BRÉZILLON, P. 1999 Context in artificial intelligence. *Computers and Artificial Intelligence*, **18**, 321–340 and 425–446; BOUQUET, P. *et al.* 2003 Theories and uses of context in knowledge representation and reasoning. *J. Pragmatics*, **35**, 455–484; Introduction in AKMAN, V. & SURAV, M., 1996 Steps towards formalizing context. *AI Magazine* **17**, Fall, 55–72; a bibliography at www.eecs.umich.edu/~rhomasto/bibs/context.html; links at context.umcs.maine.edu

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, 'If it's raining outside, carry an umbrella' implicitly assumes

- the performer is a human being,
- the performer is not a baby, quadriplegic, dead . . . ,
- 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 ark sized rain,
- the performer is not hydrophobic, hydrophilic, Gene Kelly in love, etc.

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.

That active context, he argues, is something like an imagined scenario. 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.²³

The 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 effort devoted to it. It is not so much a problem, though, if we are using Cyc as a model of how humans handle context. For humans do have commonsense knowledge available. The module of shared commonsense world knowledge is what is called in legal circles 'the reasonable man', of whom A. P. Herbert writes:

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.)

²³ LENAT *op cit*, Section 3.

I need not multiply examples. It is impossible to travel anywhere or to travel for long in that confusing forest of learned judgements which constitutes the Common Law of England without encountering the Reasonable Man. He is at every turn, an ever-present help in time of trouble, and his apparitions mark the road to equity and right. There has never been a problem, however difficult, which His Majesty's judges have not in the end been able to resolve by asking themselves the simple question, 'Was this or was it not the conduct of a reasonable man?' and leaving that question to be answered by the jury.²⁴

The jury is still out on whether any AI system will be able to equal the performance of the reasonable man. If so, it will not be soon.

Plan B. The PEIRS system for diagnosis of pathology results of thyroid tests uses reports on patients that list their sex, age, etc. and the results of several tests. Initially, there were a large number of trial cases, each assigned a correct diagnosis by an expert. The system then grew a classification tree using these cases—there are various standard methods of doing that. The aim of the tree is to generalize from the cases, in that a new case with slightly different test readings from any of the original cases should be classified correctly, as it follows down the classification tree and is diagnosed as 'similar' to one of the original cases. Naturally, the system sometimes made mistakes, and the special feature of the system was its method of handling 'knowledge maintenance'—how the tree grew and 'learned' from its mistakes. Whenever an expert pathologist found the system had made a mistake, the differences between the new case and the old case that had led to its wrong diagnosis were displayed. The expert decided which was the crucial difference, and supplied the correct diagnosis. The system then automatically added a new leaf to the tree, which would distinguish the new case (and ones like it) from the old one. The new leaf does not disturb any of the old diagnoses for which the system gave the right answer: the reason is that the new leaf only operates in a very specific context, namely, the context defined by the tree down to that point. Only a small number of cases, satisfying known constraints, get down to the point where the criterion of the leaf may or may not apply to them.²⁵ In due course, the system attained performance comparable to human experts.

Artificial intelligence researchers have clarified some possible starting places in understanding—really understanding—context. If their progress has been less than hoped for by others and much less than expected by themselves, their excuse is that a small amount of working code is worth a lot of hot air.

²⁴ HERBERT, A. P. 1935 *Uncommon Law*, London, p. 2.

²⁵ EDWARDS, G. *et al.* 1993 PEIRS: a pathologist maintained expert system for the interpretation of chemical pathology reports. *Pathology*, **25**, 27–34, further in RICHARDS, D. & COMPTON, P. 1998 Taking up the situated cognition challenge with ripple down rules. *Int. J. Human-Computer Studies*, **49**, 895–926.