The History of the Frame Problem

M.Kamermans and T.Schmits
Faculty of Artificial Intelligence
University of Amsterdam

June 2004

Abstract: This paper deals with the history of the Frame Problem, covering the period from its first formulation by McCarthy and Hayes in their 1969 paper "*Some philosophical problems from the standpoint of artificial intelligence*" up to Thielscher's description of the fluent calculus in 2001. In this paper we sketch the history in terms of the problem itself and proposed solutions over a thirty year period. In addition, this paper discusses some problems that are related to the original Frame Problem. We conclude that the Frame Problem as it was originally formulated has been solved with Shanahan's and Thielscher's approaches and that at least the logical chapter of the Frame Problem has been closed.

1 – An introduction into the Frame Problem

Every researcher in the field of AI will eventually encounter the Frame Problem, often described as the problem of 'knowing what stays the same as actions occur in a changing world'. This first encounter with the frame problem is often met with an underestimation of the problem, followed by many hours of trying to find different ways to get around it, only to find it is still quite present, and very hard to successfully get rid of. The most prevalent question concerning the frame problem is of course "What exactly is the Frame Problem?" probably followed by "Is there a solution?"

To answer the first question, the Frame Problem is really not much more than the problem of finding a logical representation that adequately describes what doesn't change when actions take place in a dynamical world. A model in which you can look at your watch, that doesn't specify looking at your watch doesn't change the colour of your socks, technically does not preclude the possibility of your socks changing colour at every occurrence of you checking the time. This sounds like a reasonably trivial problem, which has the most obvious solution of simply explicitly stating "checking your watch doesn't alter the colour of socks". However, this very quickly leads to needing to write down for nearly every action that it doesn't change the colour of your socks, eyes, hair, cat, carpet, the air, robins during spring time, or even the colour of colourless ideas that happen to dream furiously.

This very problem, having to say which actions *don't* change certain objects would result in a theoretically endless list, but at least as long as the number of actions you wanted to model times the number of different properties objects might have, was first noted by McCarthy and Hayes in 1969 when they discussed a form of logical reasoning about the world, later to be called situation calculus. In the paper that kicked it all off, "*Some philosophical problems form the standpoint of artificial intelligence*", they first formulated this exact problem and opened up a discussion that would take 30 years before a conclusion could be reached.

The researcher that runs into the Frame Problem might also imagine that they could simply say "Well, if an action is performed, and it doesn't affect a particular property, then that property simply doesn't change. Problem solved." This is called the commonsense law of inertia, and it was this idea that stood at the basis of the modern day solutions to the frame problem. The idea was not new, though. In the early days of the Frame Problem discussion, the idea was that it was a fact that nothing changed when no actions explicitly modified objects, but this led to some interesting anomalies inherent to theory – these anomalies have been avoided in modern solutions by a subtle change, namely that the commonsense law of inertia is interpreted as "Noting everything which is explicitly changed, you ignore everything else, *assuming* that it stayed the same". Due to this difference, the modern day approach effectively solves the Frame Problem, answering the second question.

In its time, the Frame Problem also attracted the attention of the philosophers, seeing in it a deeper problem related to the entire cognitive process, and an interesting problem that might have farreaching consequences because it essentially underlined that it could not be the case that for thought in, for instance, humans vast numbers of exclusion rules resided somewhere in the brain, stating what wasn't changed when certain actions were performed, giving the argument that perhaps the logical approach that demonstrated the problem was flawed as a whole. Some saw in the Frame Problem the demise of AI, or at least AI through logic, and some even went so far as to declare the Frame Problem as so intrinsically connected with things like cognition, the induction problem and the general problem of knowing what is relevant in situations, that there was no way AI would be able to find a solution on its own for it. With the logical solution to the Frame Problem finally found, it is interesting to look at, and try to answer by looking at its history, the following two questions:

a. *Why was, or is, the Frame Problem such a pivotal problem in the field of logical AI?*

b. *Have current solutions refuted the claims made by philosophers that the Frame Problem was a problem related to cognitive faculty, and would thusly not be solvable within logic.*

With these two questions in mind, the rest of this paper tries to sketch the history of this highly interesting problem, through which the attempt is made to reach conclusive answers to both of them. The rest of this paper has been divided into sections pertaining to:

- The early history of the frame problem in section 2, written by T. Schmits
- The philosophical take on it in section 3, written by M. Kamermans (3.2 in collaboration)
- Early solutions to it in section 4, written by M. Kamermans
- Modern logical solutions to it in section 5, written by T. Schmits
- A final conclusion in section 6, written in collaboration.

2 – The origins and early development of the Frame Problem

2.1 The Logical approach to AI since 1958

2.1.1 The objectives of AI and the role of representation

Although there is no undisputed description of AI’s objectives, there is a common consensus to be found among many researchers, adequately characterised by Shanahan:

> *“Objective One: To engineer artefacts capable of performing tasks which, if performed by a human, would be said to demand intelligence;*
> *Objective Two: To arrive at a scientific understanding of the principles underlying intelligent behaviour, as manifest in humans, in animal, and especially in the products of AI’s first objective.”* [Shanahan, 31]

‘Tasks performed by animals’ could be added to objective one, because many a researcher has used animal behaviour as a guideline in his or her research. Nonetheless, the characterisation quite hits the spot, as it does not describe what intelligence actually is, but it describes intelligence as being a prerequisite for certain behaviour and states that intelligent behaviour is subjective to perception.

Objective two, however, is more important. It characterises that AI is trying to arrive at a scientific understanding to what intelligent behaviour actually is based upon. Because, indeed, of what use is an intelligent artefact if we have no understanding of its working? It is vital for the field of AI to produce theories, adequate analysis and well understood principles in relation to intelligent behaviour, because otherwise it will contribute very little to science. Therefore, in relation to objective one, some scientific assumptions must be made about what the prerequisites are to intelligent behaviour, and these scientific assumptions are, of course, highly controversial.

Many researchers though held on to the idea of symbolic representation. They state that intelligent behaviour can be described best as being the result of correct reasoning and correct representation, and that these two should be formalised in such a way that rigorous and scientific analysis is possible. Logicians are among these researchers and their account on representation has played a large role since the beginning of AI. They are also responsible for discovering the Frame Problem, as we will see later on.

2.1.1 The introduction of formal logics as medium for knowledge representation

Ever since McCarthy proposed formal logic as a medium for knowledge representation in 1958, there were many discussions about its importance in relation to the subject of AI. The next two quotations of Hayes describe a twofold argument claiming that formal logic IS important (the first quotation being very much related to AI's second objective described in the previous section):

> *"One of the first tasks which face a theory of representation is to give some account of what a representation or representational language means. Without such an account, comparisons between representations or languages can only be very superficial. Logical model theory provides such an analysis"* [Hayes, 12]
> *"… virtually all known representational schemes are equivalent to first-order logic …"* [Hayes, 14]

The first quotation does not describe the importance of formal logics because, although formal logics provide scientific analysis, there is no reason to believe that no other formalism should provide the same. But it is the second argument in combination that closes the argument. And although the second argument is a little more farfetched, it is not unjustified. The justifications for the claim that all possible representation schemes are reducible to first-order logic are more philosophical by nature than would be expected, but Hayes argues [12]:

- In so far as the world is comprehensible at all, it must be composed of objects standing in various relations to one another, and these fundamental components (objects and relations) correspond to the fundamental components of language of formal logic;
- Facts concerning the relations in which individuals stand to each other can only be combined in a fixed number of ways, corresponding to the connectives of logic.

Hayes himself pointed out in the same paper that the only argument against these claims is that the world might not be like that at all, and that it does not consist of individuals with relations between them. But to Logicians this argument is beside the point, as there have not been sound arguments why the world should not be treated as such. Kirsch describes the next two important assumptions made by logicians not previously mentioned in this paper [Kirsch, 16]:

- Cognition can be studied separately from perception and motor output;

- The kinematics of cognition and the domain knowledge needed for cognition can be studied separately from learning, psychological development and evolution.

Again, these are disputable in principle, but because of the primacy of the linguistic character of knowledge logicians claim their approach to knowledge representation is the best approach.

All these arguments have not claimed formal logics to be the one and only prerequisite for intelligent behaviour, but they do claim formal logics to be the most scientific way to obtain the goals of Artificial Intelligence by using it as a medium for knowledge representation.

2.1.2 A brief description of the situation calculus

The best-known formalism for knowledge description in a changing world is the situation calculus, developed by McCarthy and Hayes and published in 1969 in a paper called "Some philosophical problems from the standpoint of artificial intelligence" (although McCarthy described it earlier in a unpublished memo in 1963). From there on it evolved into a more complex formalism, though essentially staying the same. I will describe its basics briefly as it is essential for understanding the next section in this chapter.

The ontology of the situation calculus consists of situations, fluents and actions.

- Situations are best described as a 'snapshot' of the state of the universe;
- Fluents are things that change across situations;
- Actions generate new situations from based on old ones, usually by changing fluents.

Secondly, it includes predicate symbol Holds, which relates fluents to situations for which these fluents are true. For example, let us take a look at how to describe a tower of blocks on a table (the most used example ever to explain situation calculus):

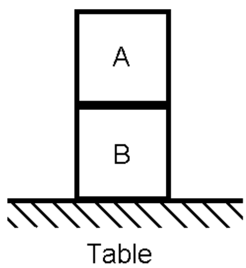


figure 2.1

Figure 2.1 shows a certain situation that is properly described as follows:

Holds(S0, Clear(BlockA))
Holds(S0, On(BlockA, BlockB))
Holds(S0, On(BlockB, Table))
Holds(S0, Clear(Table))

S0 is a denotation of the situation that is described by picture 2.1, and as you can see, fluent *On(BlockA, BlockB)* is related to S0 by *Holds*. It is a way of describing what fluents a certain situation consists of (what fluents a certain situation 'holds'). Third, the situation calculus includes *Results* which is a mapping of actions and situations onto new situations. For example, let us consider the action of putting Block A on the table:

PutOn(BlockA, Table)

With *Results* we can denote the situation that arises if the previous action is performed in situation *S0*:

Results(PutOn(BlockA, Table), S0)

Now we can describe what the previous situation looks like, by relating fluents to the previous situation using *Holds*. To state that Block B on the table after the previous action, we say:

Holds(Results(PutOn(BlockA, Table), S0), On(BlockB, Table))

We can now also write certain formulae, which describe the effects of certain actions, appropriately called effect axioms. These will have to take into account the preconditions of that action, that is, the fluents that need to hold for certain actions to succeed:

$$Holds(s, Clear(x)) \wedge Holds(s, Clear(y)) \wedge x \neq Table \wedge x \neq y \rightarrow$$
$$Holds(Results(PutOn(x, y), s), On(x,y))$$

This, in a nutshell, is the situation calculus.

2.1.3 The Original Frame Problem

McCarthy and Hayes soon discovered that problems would arise if questions would be asked about fluents that did not change due to the actions performed. For example, let's just say we introduced this next fact to situation *S0*:

Holds(S0, Red(BlockA))

Will it also be true that *Holds(Results(PutOn(BlockA, Table), S0), Red(BlockA))*? Clearly, any human would answer 'no' to this question as there is no reason to believe the colour of Block A has changed just by moving it, but in this system there is no *proof* of that fact yet. We would have to introduce the hypothesis that things stay red as they are moved:

Holds(s, Red(x)) $\rightarrow$ *Holds(Results(PutOn(x, y), s), Red(x))*

This axiom is called a frame axiom, describing the persistency of a fluent that is unaffected by an action. But, of course, many fluents stay the same as blocks are moved. There will be enough space on the table to place another block upon, whatever the situation. Assuming Block B and the table have other colours, these colours will also remain the same after a *PutOn* action, and so on.

The amount of frame axioms necessary to describe situations adequately is considerably large in comparison to the domain of n fluents and m actions, because in general most fluents are unaffected by most actions. Generally speaking, every time we add a new action we are going to have to add roughly as much frame axioms as there are fluents, and every time we add a new fluent we are going to have to add roughly as much frame axioms as there are actions. Thus, the total amount of frame axioms required for a domain of n fluents and m actions will be $\mathrm{m} \times \mathrm{n}$! For example, a domain of 500 fluents and 100 actions would require 50,000 frame axioms. This amount $m \times n$ is dependant on how these naïve frame axioms are defined. Researchers have explained naïve frame axioms to be of forms that will result in a total number of $2 \times m \times n$ frame axioms. Nonetheless, this mount is of the same order and, as $m \times n$ is a too large amount of frame axioms, $2 \times m \times n$ will definitely be as well! For sake of argument $m \times n$ will be used from here on.

As we would like intelligent agents to act in far bigger domains, we would have to find a way of minimising these frame axioms, otherwise the time an agent needs to predict situations is enormous. This, in a nutshell, is the original logical Frame Problem.

2.2 The impact the Frame Problem had on the field of AI

The sheer amount of papers that describe the Frame Problem and possible solutions to the Frame Problem is enormous. Surprisingly, not only logicians have been involved into solving the Frame Problem, there have been many anti-logicians and philosophers writing about the Frame Problem as well, most of them reformulating the Frame Problem and giving it a new definition, based on different perspectives. Some claimed the logical Frame Problem to be an instance of *"the general problem of stating 'laws of motion' which adequately describe the world"* [Ford & Hayes, 11], or Hume's Problem of Induction [Fetzer, 8]. Fodor said that *"the Frame Problem is just [the problem when to stop thinking] from an engineer's perspective"* [Fodor, 10]. Then again, Pat Hayes claimed that *"Fodor doesn't know the Frame Problem from a bunch of bananas"* [Hayes, 14]. Last, but not least, ideological foes of AI such as Dreyfus and Searle cited the Frame Problem as being the cause of death of the field of AI. As is clear to see, the Frame Problem proved to be a firm battleground where many disciplines were pitted against each other and it even played a large role in the rather shocking conversion of a stalwart logician, McDermott, to a confirmed anti-logician [McDermott, 21].

2.2.1 The introduction of the philosophical aspect to the Frame Problem

The earliest reference to the Frame Problem from a philosophical point of view was made by Dennett in 1978, though his most memorable essay concerning the Frame Problem was published in 1987, called *"Cognitive Wheels: The Frame Problem of AI"*. In this essay he describes the Frame Problem to be *"… a new, deep epistemological problem – accessible in principle, but unnoticed by generations of philosophers – brought to the light by the novel methods of AI, but still far from being solved"* [Dennett, 5]. After this publication, an enormous amount of essays and papers have been published in reaction to Dennett's claim.

Now one might ask, why did the Frame Problem become the subject of heated philosophical discussion? Did Dennett actually discover a new, deep epistemological problem? We will try to answer the first question in twofold, after which the second question will have to be reformulated.

The first reason why Dennett was responsible for a wave of philosophical discussion is not based on scientific hypotheses, but it is solely speculation. Dennett first claimed the Frame Problem to be important, because *"a certain trick we have just performed is flat impossible"* (referring to what humans are capable of) [Dennett, 5]. Then he kicked a lot of people in the shins with his tone, claiming that many people had failed to acknowledge or explain 'the bigger Frame Problem'. To prove our point we will quote Dennett [5]:

"One does not have to hope for a robot-filled future to be worried by the Frame Problem. It apparently arises from some very widely held and innocuous-seeming assumptions about the nature of intelligence, the truth of the most undoctrinaire brand of physicalism, and the conviction it must be possible to explain how we think."

"… Hume, like virtually all other philosophers and 'mentalistic' psychologists, was unable to see the Frame Problem because he operated at what I call a purely semantic level, or a phenomenological level. …"

"It is rather as if philosophers were to proclaim themselves expert explainers of the methods of a stage magician, and then, when we ask them to explain how the magician does the sawing-the-lady-in-half trick, they explain that it really is quite obvious: The magician doesn't really saw her in half; he simply makes it appear he does. 'But how does he do that?', we ask. 'Not our department', say the philosophers – and some of them add, sonorously: 'Explanation has to stop somewhere'."

The second reason is the many problems the Frame Problem has been taken for. It has been said that: *"a definition of the 'Frame Problem' is harder to come by than the Holy Grail"*, and indeed, many papers start with 'a brief description of what the Frame Problem is', all of them being different in some way or the other. Eventually, many researchers offered a solution to a 'Frame Problem' they themselves came up with! But as long as other people didn't acknowledge their definition of the Frame Problem, they would definitely deny their solution as being 'the one that cracked the Frame Problem'. Heated discussions took place, debating definitions and solutions alike. One of the most memorable discussions about the Frame Problem, in our opinion, is the one between Fodor and Hayes (of which the earlier mentioned quote of Hayes involving a bunch of banana's is a part of). In *"The Robot's Dilemma: the Frame Problem in Artificial Intelligence"* published in 1987 (ed. Pylyshyn) an excellent example of 'a heated exchange' is found.

Hayes in reaction to Fodor's "Modules, Frames, Fridgeons, Sleeping Dogs, & the Music of the Spheres", 1987 *: "Fodor misuses the term Frame Problem in at least three different ways. … perhaps Fodor's discussion is not meant to be taken [as precisely as Hayes does]. But then all he is saying is something like: Gee, AI is really hard. We don't need to work with philosophers to know that, thanks. …"*

Fodor in reaction to Hayes' "What the Frame Problem Is and Isn't", 1987: *"I have to agree with Hayes that, if you don't call [what Fodor in [Fodor, 10] described as being the Frame*

Problem] the Frame Problem, that makes the Frame Problem easier. What I don't understand is why Hayes finds this banal observation comforting. If you don't call it cancer, then what you do call cancer won't be what you die of. But you end up dead either way."

2.3 Reformulations and Fringe Problems

The next section describes problems that have often been related to the Frame Problem, though researchers have brought up many more. The ones mentioned here, though, are important enough for many researchers to have dealt with them, so it might be useful to know what they mean. They are described here in chronological order, using the times when they were first referenced to.

2.3.1 The Qualification Problem

The qualification problem was defined by McCarthy in 1986: roughly speaking, the problem of specifying what conditions must be true in the world for a given action to have its effect. Secondly, how can we be sure that all the preconditions that are built in effect axioms are all the preconditions there are?

The classic example is the case of turning the ignition key: For the car to work when the ignition key is turned, a lot of preconditions must be true. For example, the battery must be alive, there must be gas in the tank, an engine should be present, etc. As with the original Frame Problem, this is a representational problem with a computational aspect, but it does differ. Even if the Qualification Problem would be overcome, the Frame Problem would still be present and finally, the qualification problem can be overcome with default reasoning like circumscription [Shanahan, 31].

2.3.2 The Ramification Problem

Finger pointed out in 1988 that there is a problem accounting for derived and delayed effects that are tedious to infer. Let's consider the next example: If I put a newspaper on top of an air vent, then my hand will be empty, and the newspaper is lying on top of the vent. But in addition, the room may become stuffy if it is the only air vent, though it is not the placing of the newspaper on top of the vent that directly causes the room to be stuffy. These ramifications have complicated the logical Frame Problem, and not many solutions have been able to cope with this difficulty in a satisfactory way.

2.3.3 The Temporal Projection Problem

Morgenstern and Stein pointed out in 1988 a collection of problems, called the temporal projection problem. It is generally described as a combination of the persistence problem, the problem of reasoning about change, the ramification problem and the backward temporal projection problem.

The persistence problem is the difficulty of determining what stays the same in a changing world, the problem of reasoning about change being its counterpart. The combination of these problems is often referred to as the forward temporal projection problem (or the prediction problem). The backward temporal projection problem is the difficulty of reasoning about what happened at an earlier time point if we are told what is true at a later time point. The temporal projection problem has been quite a subject of study, and solutions to the frame problem have often been confronted with it.

2.3.4 The Problem of Induction

The problem of induction is not a problem many researchers have confronted their solutions with, but many have responded to Fetzer's statement. Fetzer deviated quite far from the original Frame Problem by McCarthy and Hayes by stating in 1987 that the Frame Problem is an instance of the problem of induction. Much later he stated more accurately:

"The problem of Induction as Hume viewed it was one of justifying some inferences about the future as opposed to others. The Frame Problem, likewise, is one of justifying some inferences about the future as opposed to others. The second problem is an instance of the first." [Fetzer, 8]

The induction problem proposed by Hume is this: 'What justifies our moving from a finite number of particular observations to conclusions that cover most cases we have no observed?' And indeed, how can we justify our belief that next July will be warmer than next February? Note that it is not the question how we obtained this belief, but how we can justify it. The problem of induction is of course a more philosophical problem than the previous ones, but Fetzer's statement does appeal to intuition.

It has been widely accepted, though, that the Frame Problem is not the Problem of Induction in disguise. As Dennett put it: *"Having access to absolutely justified probabilistic knowledge is of little use to a robot who has to make a plan to save its spare battery if it does not know how to apply this knowledge at hand"* [Dennett, 5], or as Hayes put it: *"Even if we had the laws of science given to us by an angel on tablets of bronze, the frame problem would still be there."* [Hayes, 14]

3 – Philosophy and the Frame Problem

The Frame Problem, and the discussion it caused, was not just something that logicians and anti-logicians were interested in. Philosophers saw in it a deeper problem, carrying much more meaning than being just a problem encountered when trying to model the world in terms of logics. Dennett and Fodor were two of the first to argue that we were in fact dealing with a rather fundamental epistemological problem that exists in every system that needs to intelligently reason about matters concerning the world. Reasoning from a philosophical standpoint, Dennett, Fodor and others (such as Haugeland and Janlert) have given their view on the frame problem, related problems, and the repercussions they have on much more than what the problem initially applied to.

3.1 Philosophical takes on the Frame Problem

3.1.1 Dennett's take on the matter – An epistemological view

Dennett's 1984 article "*Cognitive Wheels: The Frame Problem of AI*", republished in the 1987 defining work "The robot's dilemma", was one of the first of many articles to spark the discussion about the philosophical nature of the Frame Problem. To illustrate the problem, Dennet sketched various situations in which a series of robots all failed at a seemingly simple task, retrieving a battery on a dolly in a room. The first robot fails because it doesn't notice there's a time bomb on the dolly. The second robot fails because while it sees the bomb, fails to realise that moving the dolly also moves the bomb. The third robot fails because it has been programmed to make sure it takes into consideration all the right things, and should ignore all other possible bits of information and derivable information, thus deadlocking it in an endless process of thinking of information and disregarding it as not the right thing.

Using this as demonstration, Dennett explained that the Frame Problem as originally outlined by McCarthy and Hayes was not really the whole story, but was in fact a manifestation of a bigger "frame problem": The problem of determining which information is relevant to reasoning in a situation, and which information is to be ignored.

One interpretation that might have offered a solution to that problem was Hume's idea that humans perform this relevancy pruning task through associatism, with certain transition paths between ideas and actions being reinforced, but Dennett could not accept this idea, using the "why" game as demonstrator that this theory breaks down there where AI needs it most. He proposes the situation where two children both take cookies from the jar, where one is spanked for this, and the other is not. The child that gets spanked eventually stops, and Dennett starts to ask why. The why game runs through "because the child gets spanked" – "because spanking causes pain"- "because the child doesn't want to experience pain" and another "why?" shows the problem to Hume's solution: it's not easy to answer, if it can be answered at all.

According to Dennett, AI forces to the surface the problems that are so easily overlooked in trying to find out what the relevant information for tasks is, because AI needs to start from scratch, not having the luxury of a system that possesses any form of common sense at the start of the design process. He argued that this bigger frame problem was not a new problem at all: humans too suffer from the problem that relevancy is not always properly determined, and when we make a mistake, we are surprised, as we are confronted with our own inability to have properly determined the nature of the situation.

In Dennett's own words, "*This suggests a graphic way of characterizing the minimal goal that can spawn the frame problem: we want a […] robot to be surprised […]. To be surprised you have to have expected something else, and in order to expect the right something else, you have to have and use a lot of information about the things in the world*." [Dennett, 5]

This sounds a lot like the induction problem, and Dennett noted this too, but pointed out that even if the induction problem was solved, the frame problem would persist, thus demonstrating that the two are not entirely the same problem: suppose a reasoning agent would, through some miracle of science, be able to draw conclusions about the world which were never wrong, based on all the knowledge that it has. Thusly free of the induction problem, it would actually still suffer from the frame problem in that it still needs to somehow represent and process all this knowledge it has in some way. The actual representation and processing required to reach a conclusion, however, is not related to the truth-value, perceived probability or even subjectivity of the conclusions themselves.

According to Dennett, solutions to the frame problem as had been offered so far had all been inspired by different fields and different minds, but they could still all be grouped under a single label stressing an important aspect: they were all what Dennett called cognitive wheels, in analogy to the concept of wheels when looking at the natural world. The wheel was an excellent engineering solution to the problem of efficient motion, but it was nowhere to be found in nature itself. Cognitive wheels then, were solutions to the problem of modelling cognitive faculties, which had absolutely no relevance when looking at them from the real world, contrasting them to for instance how cognitive faculties worked in humans – which was the role of hard AI; the modelling of cognitive function in artificial systems. In Dennett's view, AI should have been (and presumably should still be) wary of the readiness to use such cognitive wheels as solution to problems encountered, as the use of a counter-intuitive and highly artificial solution may cause AI to ignore what it set out to do in the first place.

Dennett himself has no solution to the frame problem. As he saw it, the deep epistemological issue of the frame problem was not one reserved to AI, but applied to all living, reasoning systems, and as such may not even have a solution at all. This view demonstrates an interesting difference between philosophy and AI as science: where philosophy is not hampered by the existence of this conundrum, AI as science concerned with modelling cognitive functions may not be able to progress in this goal until it has learned to deal with this problem. Dennett concludes that AI may not have the answer now, but that at least having asked the question, is a good start.

3.1.2 Fodor's take on the matter – Epistemology and the scientific view

Fodor, also in "The robot's dilemma", related the Frame Problem to being the problem of formalising our intuition about inductive relevance. Looking at cognition in humans, Fodor argued that there were two key issues at play [Fodor, 10]:

> *"The frame problem goes very deep; it goes as deep as the analysis of rationality"*

> *"Outbreaks of the frame problem are symptoms of rational processing; if you're looking at a system that has the frame problem, you can assume that it's rational at least to the extent of being unencapsulated."*

According to Fodor, only modular cognitive faculties – exemplified by such things as audio parsing – are free of the frame problem, as they operate on a purely perceptual and informationally encapsulated principle. Audio parsing (though not the interpretation) for instance, only relies on an incoming audio signal that can be chopped up according to some "grammar" that tells the faculty how to chop up the signal so that all the signal chunks correspond to the right sounds. Fodor believed that the only cognitive tasks we had been able to understand and recreate in AI, were those tasks that did not demonstrate the frame problem: only the modular tasks. Non-modular tasks, he argued, suffered incurably from the frame problem by the fact that they were informationally unencapsulated.

The concept of the "sleeping dogs strategy", a form of the commonsense law of inertia, was rejected as solution by Fodor on the basis of a very strong epistemological stance: depending on how the world is conceptualised, no proverbial dogs would be left sleeping. The example Fodor gave to illustrate this, was the concept of a fridgeon. A well-defined concept that applied to everything, as long as Fodor's fridge was turned on, and didn't apply to anything at all, when it was turned off. Fodor argued that by flicking the switch on his fridge, he was changing a property of every single particle in the universe. He argued that regardless of how 'not real' the idea of a fridgeon was, the fact that it was a perfectly valid way to look at the world – everything being a fridgeon or not, depending on whether the fridge was on or not – meant that the sleeping dogs approach simply did not always work.

The problem of choosing the way in which to represent the world and actions within it, was according to Fodor the reason why one could never implement the sleeping dog's solution as absolute solution for the Frame Problem. As there would always be an infinite amount of "kooky" concepts (like the fridgeon) conceivable, the problem of which conceptualisation to use for the world was essentially the problem of formalising one's intuition about inductive relevance: which concepts would be cookies, and which wouldn't?

Fodor concluded that because of this, the Frame Problem was really the same as the problem of formalising our intuition about inductive relevance, which was a particular hard and deeply philosophical problem, directly related to human cognition and all that entails, and shouldn't therefore not be left in the hands of AI, but rather should be treated by the philosophers and cognitive scientists. As he so aptly puts it in his response to McDermott's criticism on the philosophical view of the problem: "*The frame problem – to say it one last time – is just the problem of nondemonstrative inference; and the problem of nondemonstrative inference is – to all intents and purposes – the problem of how the cognitive mind works. I am sorry that McDermott is out of temper with philosophers; but, frankly, the frame problem is too important to leave it to the hackers.*"[Fodor, 10]

3.2 The philosophy behind the logic

Logic finds its origin in the philosophy of Aristotle. From the start, logic had been a way to model reasoning, being the only sound system that allowed proper inference of conclusions based on premises. Since the conception of cognitive science, the Philosophy of cognition too has been of great influence on how we perceive the cognitive processes to run in our own heads, but only recently has the step been made to attempt to model cognition in artificial systems. For the past fifty year there has been nothing short of a quest to model cognition and intelligence in artificial systems, using a wide variety of approaches, but of these, only the logical one can be said to be scientifically justifiable.

People such as Hayes [1977], Kirsch [1991] and Shanahan [1997] claim that there has to be a single architecture that fully models the cognitive process, and that it is justifiable to believe that this architecture can entirely be modelled through logic. And to paraphrase Shanahan, while perhaps it isn't appropriate for the purposes of designing representations to talk about the real world as if it consists of individuals with relations between them., no conclusive argument to this effect has been put forward.

In opposition of this, people such as Dennett and Fodor do not deny that logic should be capable of properly modelling cognition, but instead argue that because this "single architecture" assumption may be false, the approach taken cannot lead to any realistic result. Searle's interpretation [1980] goes beyond this and is much stronger in that he does not believe that cognition is in any way reducible, thus claiming that there can be no underlying singular architecture to cognition itself. Taking it one step further still, Dreyfus [6] argues that there can in fact be no cognition without a worldly body that allows cognition to manifest itself and interact with the world, adding the necessary context to the abstract concepts formed by the mind.

Given these contrasting views, with on the one hand Hayes, Kirsch, Shanahan and others, and on the other hand people like Dennett, Fodor, Searle, Dreyfus, it is clear there are both arguments in logic's defence as well as its against it. This places the use of logic as *the* way to reach a proper model of cognition at an interesting place: on the one hand, it may arguable be the wrong approach. On the other hand, it's the only known scientifically justifiable approach we know of. In fact the only way to end this discussion, is either by the logical approach leading to a working model of cognition, or by irrefutable evidence that logic *cannot* capture cognition. Since there has not been either such a model, nor such evidence proving such a model cannot exist, there remains an impasse. From a philosophical view though, logic is still the only candidate there is with which reasoning about cognition can be done in such a way that philosophy itself at least cannot deny the validity of the method.

4 – (Partial) solutions to the frame problem, and the problems these solutions themselves brought with them.

In this chapter we shall present an outline of some of the solutions that were posed in relation to the Frame Problem, and in which respect these solutions do and do not solve it. The chapter divides approaches in solutions into two sections: the logical approaches, outlined in 5.1, and the approaches taken from outside the logical field, outlined in section 5.2.

As many before us have already noted, before one can perform an objective evaluation of the solutions proposed in relation to the Frame Problem, it is essential to have specific criteria that solutions need to meet before claims can be made on degrees of solving the problem. Morgenstern [22] in her 1996 paper suggested seven criteria to be used in evaluating the Frame Problem, of which we found five to be essential criteria.

The most obvious criterion that Morgenstern stated was that a solution should solve the Frame Problem, not a derivative of the Frame Problem, or a redefinition of it. This may sound like an obvious criterion, but in the past 25 years it has been more rule than exception to redefine the Frame Problem and solve this redefinition instead. we have omitted Morgenstern's second criterion, which states that a solution should be intuitive. This is not to say that it is a bad criterion to work with, but sometimes the best solution to a problem may not be the most intuitive one – becoming intuitive only once one has become familiar with the solution.[1]

Morgenstern's third criterion, however, is a strong one: solutions need to be truthful. That is, no assumptions or approximations may be left in the theory, as these can invariably lead to a simply incorrect theory (*for an example of this, see section 4.1.2, regarding the modal language Z*). A fourth and important point outlined by Morgenstern is that the solution should be able to apply to a concrete example of the problem: many solutions have been offered that involving solving a highly constructed "toy problem". If a solution is claimed to be a genuine solution to the Frame Problem, it should not only hold for these toy problems, but also for concrete real-world problems.

A fifth criterion Morgenstern suggests, is conciseness. That is to say, it should not fall into the trap of the Frame Problem: giving an endless list of patch-rules is no better than an endless list of frame axioms. Finally, the last criterion Morgenstern suggests is that a solution should be theoretically founded. A procedural approach, for instance, may not explicitly be a logical theory, but it should still be possible to derive or demonstrate the logical principles behind the theory as being sound and correct.

---

1) A classical example of this is research into the ability of people who had never been introduced to formal logic in their life about the validity of modus ponens. A problem such as "spears are long sticks. Abrim has a spear. Does Abrim have a long stick?" would intuitively be answered with 'yes' by anyone already familiar with modus ponens, but this turned out to by no means be a natural intuition in people who had never been explained the reasoning that made modus ponens valid.

Morgenstern's last criterion of simplicity, applying Occam's razor to multiple theories of different complexity but equal effectiveness, we believe to already be covered by the need to be concise, and the requirement that the theory needs to solve the right problem. If there are competing theories of different complexity that both solve the problem, then it is not a matter of selecting the simplest one, but finding out why they both solve the problem and whether they are perhaps forms of each other. This then leaves us with the following criteria:

1. Solving the right problem
2. Being truthful
3. Applying to concrete examples
4. Conciseness
5. Being theoretically founded

Deserving mention before we continue, is that a genuine solution to the Frame Problem should not just hold for single-actor worlds, but should also allow both concurrent single-agent actions, as well as concurrent multi-agent actions in a world to be handled. As an example of concurrent multi-agent actions, consider that while a robot that finds its way through an office building perfectly is certainly be a major achievement, having the restriction of only being able to do so when the environment remains static, where no one does something the robot may not be familiar with, is still far from a working AI agent. As single-agent concurrent action we can imagine a system that's supposed to co-ordinate a (in approximation) continuous path. This can be done through for instance adjusting the speed, and making turns, but as long as the system cannot adjust the speed while making a turn, either a large segment of possible paths is ignored, or the number of actions taken becomes a problem, as a continuous path is an in approximation infinite number of finite-duration actions.

This criterion of concurrency is implied in the need for the solution to apply to concrete examples, and for this reason has not been listed as a separate criterion. Armed with the five criteria for what constitutes a good solution, what follows are the different approaches taken to solving the Frame Problem, and an evaluation of the solutions they resulted in.

4.1 Logical approaches

In this section we will look at some of the solutions that have been offered for the Frame Problem from inside the field of logic, where it was first discovered by McCarthy and Hayes in 1969. The solutions offered during the earlier discussion of the problem, in the period between the mid 80's to the early 90's can be categorised in three sections, being the procedural, the monotonic and the non-monotonic approaches to solving the Frame Problem. In the following three sections we shall look at these approaches and the solutions they have led to, as well as evaluate these solutions in terms of the criteria stated earlier in this chapter.

4.1.1 Procedural approaches to solving the Frame Problem

In 1971, Fikes and Nilsson developed the procedural language STRIPS [Fikes & Nilsson, 9], as a new approach to theorem proving, and problem solving, using first order predicate logic. In STRIPS, the state of the world is represented as a conjunction of only positive FOP literals, which are ground and function free, and no free variables inside a world state are allowed. Goals in the STRIPS language are simply states, and a state *s* in STRIPS satisfies any goal *g* when *s* contains all literals in *g*. It is important to note that STRIPS operates on the closed world assumption. Thus, any state not explicitly in a STRIPS model of a world is assumed to simple not exist, and any condition not explicitly present is assumed to be false, operating on the Negation As Failure principle.

Actions in STRIPS are represented as action schema: an action name which may contain free variables, a set of preconditions that need to be true before the action can take place (covering all the variables used in the action) and the effect that this action has on the state of the world that this action is performed in. An example of this is:

ACTION(relocate(*x*, *from*, *to*),
PRECOND: At(*x*, *from*) ^ Location (*from*) ^ Location(*to*),
EFFECT: ¬At(*x*, *from*) ^ At(*x*, *to*))

The effect of an action is often represented as a 'delete-list' and an 'add-list', denoting the two sets of negative effect literals, and positive effect literals. STRIPS works by taking a world state and allowing any action whose preconditions can be unified with this world state. This leads to certain literals of the world state being modified through the action's effect, leading to a result state. This process can be repeated until a particular goal has either been satisfied or no actions can be taken anymore in the world state. Literals are never added to a world state from the effect add-list if they are already present in the world state, and literals that are negated are not added to the world state but ignored under the closed world assumption. If a literal in the world state is found as negation in the delete list, then it is removed from the world state. The argument goes that through the use of this procedural approach, the frame problem is solved by a form of default logic: the closed world assumption guarantees that we don't need frame axioms to say which objects are not affected by which action, because unless it is explicitly stated that an action modifies an object, it simply doesn't.

There are, however, a few problems that result from the STRIPS approach. Firstly, STRIPS will not work in a concurrent action world. If we have a world where there are two actors, and two blocks – let's call them A and B – and one of the actors moves block A, while the other moves block B, then there is no way to represent the fact that while Moving A does not causally move block B, block B will not be at its original location once A has been moved. Likewise, it means that there is no sure way to guarantee that block B has not moved in a concurrent situation.

Secondly, and perhaps more importantly, STRIPS does not allow for temporal reasoning. The use of instantaneous cause-and-effect resolution means that the possibility that one action takes longer to perform than does another is ignored, preventing STRIPS from providing an accurately model of what happens in the world, be this the real world, or a virtual world.

An important recent expansion of STRIPS is the development of ADL, or Action Description Language. Unlike STRIPS, ADL works with an open world assumption, and allows negative literals as part of its world state too, with literals that are not described in the world state being assumed 'unknown', rather than false. It allows for quantifiable variables in goals, rather than forcing ground literals, and supports variable type declaration such as relocate(p: Persons, from: Location, to: Location) rather than explicitly requiring set name predicates.

However, even in this expansion the Frame Problem still persists in that concurrent actions will still lead to problems that can only effectively be solved by adding in ADL-equivalents of frame axioms to cope with the potential concurrent actions. Also, like STRIPS, ADL lacks a temporal dimension along which actions take place, making ADL applicable only on relatively toy worlds too. While this does not prevent STRIPS and ADL from being useful tools in planning that does not require a time constraint in its actions, it does mean that as solutions to the Frame Problem, they do not cover all the ground that needs to be covered for them to work as such.

4.1.2 Monotonic approaches to solving the Frame Problem

Logic is said to be monotonic when it adheres to the simple rule that if a set Γ of formulas implies a consequence C, then a larger set $\Gamma \cap A$ will also imply this consequence C. Monotonic approaches to solving the Frame Problem are all based on this fundamental concept in logic, subsumed by classical logic and modern branches such as modal logic, and predicate logic. Within the classical framework of logic, various solutions to the Frame Problem have been suggested, The most interesting of which is Reiter's solution through the use of situation calculus [Reiter, 24]; the very language that exposed the Frame Problem in the first place. Before we look at Reiter's solution however, we examine the approach through modal logic, using the modal language Z, by Brown and Park [2, 3], and the modal language ZK, a Kripke type structure, as proposed by Schwind [29].

The modal language Z

The modal language Z, as proposed by Brown [2], was designed in order to allow the development of a simplified, intuitive, semantics. Reasoning from McCarthy's 1977 claim that "*For AI's purposes, we would need all [the above] modal operators in the same system. This would make the semantic discussion of the resulting modal logic extremely complex*", Brown argued that if the modal language Z contains all other intentional concepts, then the complexity of the entire system could not be bigger than the complexity of the language Z itself, and as such made Z a solid starting point to solving the frame problem. In [2], Brown explained the inner working of the modal language Z, explaining how it's similar to an *S5* modal logic, but stronger because it contains axioms that allow individual facts to be proved to be logically possible with respect to a body of knowledge, after which (in co-operation with Park) he set out to solve the frame problem [3].

Using Z, an attempt was made to solve the frame problem by redefining the frame problem as "the problem of determining what is logically possible as defined by the modal logic described with Z." This immediately raises questions on whether or not this means the frame problem is actually solved by this approach, as solving a redefinition is not the same as solving the original problem. What Brown and Park end up doing is giving a proof of the solvability of the Yale Shooting Problem (*see section 4.1.3 for a detailed explanation*), thus showing it can handle the problem of persistence, which they follow with an attempt to solve a more classical toy problem of two blocks, A and B, atop a surface, with block B resting on block A, and block A resting on the surface.

The reasoning system Z uses a PHYSICAL-LAWS construction in order to reject possible resultant worlds in which (real world) physical impossibilities such as AT(A,location1) after a MOVE(A,location1,location2) action are deduced. When applying this in their proof for solving the block-world frame problem however, when one of the resultant worlds shows the possibility of block A (the lower block) having been moved to a new location and block B (the upper block) having remained where it originally was, even though it is still on top of block A, this possible world is rejected as potential outcome, using as argument that the PHYSICAL-LAWS construction would not allow this [3, pp.168].

This is quite curious. Rather than having the physical laws dictate what is possible, let us analyse the world that it rejects and see if it should actually be impossible for this to occur in the real world: we look at the following situation that demonstrates problems at the fringes of the physically possible for this modal language. In a block world a block is resting on another block, which itself is resting on a surface. The upper block is exactly twice the size of the lower block, and is poised precariously on top of the lower block. According to physical laws, this situation is wholly possible, and given a force being applied as indicated, with the right momentum to momentously overcome friction, situation 2 may follow from situation 1 without any laws of physics being violated.

Situation 1 resultant Situation 2

After relocation of the lower block in this fashion, it is quite obvious that no part of the lower block is anywhere at its original location, whereas the upper block has not moved and is still poised precariously on top of the lower block. It becomes obvious that the proof given by Brown and Park operates on the silent assumption that the blocks in their proof were of such a nature that moving the lower block implicitly already excluded the possibility of the upper block remaining at its location in the world.

This is, at the very least, careless. If the "physical laws" construction is supposed to allow us to tell what is possible and what is not, it should not be able to disallow possibilities in the world. A generalisation can be made here that using a *physical laws* concept is essentially making the Frame Problem inherent to the solution: an explicit list of all physical laws, required in order to have an agent rather than a human solve a problem using the language, could potentially be many times longer than a list of all possible frame axioms that would be required in a model. Also, in order to correctly reason with real physical laws in any language, any object represented in the language would require a list of properties being described in a way so that the physical laws can be properly applied in comparison to what would really happen in the world (be this the real world or a virtual world).

Relying on physical laws in a model to handle the selection process of possible worlds, brings with it the need to actually fully describe the world in terms of objects, actions, and how these actions affect objects through the physical laws of the system being modelled. While this may solve the original frame problem, the explicit solution for any complex world would be many times longer than any inclusive list of frame axioms would be for the same world.

The modal language ZK

Schwind takes an alternative approach, also using a modal language, termed ZK (which is not an extension of the previously mentioned modal language Z). The modal language ZK is an extension on classical predicate logic, where all structures are Kripke structures, using two state relations, + and –, to indicate directly following and preceding states within logical expressions (with + being equal to Wright's *Mn* construction), thus embedding in the system a temporal aspect, as well as the operator L (“is logically true”, equal to *S5*) and the common modal operator ◊ [Schwind, 29]. The problem with Schwind's approach using ZK, however, is stated by herself at the very start of the paper as proposing “*A solution to the frame problem […] by reducing it to the logical problem of consistency proving.*”

While it might be of great value to logic if Schwind succeeded in solving the problem of consistency proving, under the first criterion mentioned at the start of this chapter, this approach is not a proper solution for the Frame Problem itself. While beyond the scope of this paper, those interested in the validity of Schwind’s approach to solving the problem of consistency proving can find the theorem prover for ZK in detail in [Schwind, 28]

Reiter’s solution (sometimes)

In 1990, Davis came up with the idea of framing primitive fluents by events, which essentially fixes what is changed by what (in the logic used) [Davis, 4]. This was done by designating certain states and fluents as ‘primitive’ fluents, and everything else as derived. The idea behind this was that the frame axioms would then only alter specified primitive fluents and states, as the rest remained unaltered. This approach meant that only one frame axiom was required for each action type, plus an axiom for each primitive state, giving rise to a number of frame axioms well under the $n*m$ number McCarthy and Hayes stated in their 1969 publication [20].

However, Davis realised that this solution would be unable to handle concurrent actions, and the distinction between primitives and derived concepts was a highly artificial one. A second suggestion by Davis, independently suggested by Schubert in [27], was that of explanation closure, in which a fluent can only be changed by a particular event. If the event does not occur, then the fluent does not change. This led to a total number of $2*F$ explanation closure axioms, as a pair of axioms for each fluent of the form:

$$\begin{Bmatrix} R(x,s) \wedge \neg R(x,do(a,s)) \supset \alpha_R(x,a,s) \\ \neg R(x,s) \wedge R(x,do(a,s)) \supset \beta_R(x,a,s) \end{Bmatrix}$$

Where R is a fluent, x an object, a an action and s a situation. The α and β in these closure axioms are the generic labels for any and all actions that may respectively facilitate a negative or positive change in this model. This is quite a contrast to the otherwise required $2*A*F$ number of frame axioms, where F stands for the number of fluents in the model and A for the number of actions. Schubert noted, however, that this only worked under the Explanation Closure Assumption:

"*The only way fluent R can change truth value under action* a, *is if* α_R *were true. This means, in particular, that* α_R *completely characterises all those actions* a *that can lead to this change. The same holds for* β_R."

This concept was expanded on by Reiter in 1991, in his article "*The frame problem in the situation calculus: a simple solution (sometimes) and a completeness result for goal regression*", which as its title implies offers only a partial solution, but does not pretend to offer a full one like most other approaches did. The "sometimes" the article title applies to the fact that Reiter's solution only works under what Reiter calls the Completeness Assumption for Fluent Preconditions, using a slightly different notation for effect axioms:

$$\left\{\begin{array}{l} \pi_a(x,s) \wedge \varepsilon_R^+(x,y,s)) \supset R(y,do(a(x),s)) \\ \pi_a(x,s) \wedge \varepsilon_R^-(x,y,s)) \supset \neg R(y,do(a(x),s)) \end{array}\right\}$$

Where πa (x,s) denotes the action preconditions, εR $(x, y, s))$ the fluent preconditions, and $R(y, do(a(x),s))$ is the fluent state that results from performing action a on x in situation s. The completeness assumption for fluent preconditions states that: "The fluent precondition ε_R^+ (x, y, s)) specifies all the conditions under which action a, if performed, will lead to the truth of R for y in a's successor state. Similarly, ε_R^- (x, y, s)) specifies all the conditions under which action a, if performed, will lead to the falsity of R for y in a's successor state."

Reiter proposes successor state axioms that take the form:

$$\text{Possible}(a,s) \supset [R(\text{do}(a,s)) \equiv \gamma_R^+(a,s) \vee R(s) \wedge \neg\, \gamma_R^-(a,s)]$$

, and for each action A, an action precondition axiom of the form $\pi_A(s) \supset \text{Possible}(A,s)$. Furthermore, unique names axioms for actions and states were required, but in doing so this solution managed to reduce the number of required frame axioms from the original $n*m$ with n actions and m fluents to the much smaller $n+m$ instead.

Clearly, this is not so much a true solution to the frame problem as it is a way to make the problem less pressing, which was quite a feat. While it means that within the monotonic logic the frame problem persists, it can be reduced to a minor problem inherent to the fact that monotonic logic is used, rather than an insurmountable obstacle.

4.1.3 Non-monotonic approaches to solving the Frame Problem

Lastly, as alternative to monotonic logic, is the non-monotonic logic. Unlike monotonic logic, the concept that if a set Γ of formulas implies a consequence C, then a larger set $\Gamma \cap A$ will also imply this consequence C is not true under a non-monotonic approach. Axioms added after derivations based on the set Γ have been performed, may in fact invalidate any or all of the derivations made prior to the introduction of new axioms. This approach allowed for solutions to for instance the "Tweety Bird" problem[2], and it was thought that perhaps a solution to the Frame Problem, too, could be found through the use of the non-monotonic principle.

Circumscription

The dominant solution from the non-monotonic corner is the idea of circumscription, introduced by McCarthy in 1980 [18], which was reworked in 1986 [19], where the idea was to restrict extensions to a predicate or set of predicates as much as possible. This would for instance be able to describe a sentence such as " Birds usually fly" by rephrasing it to "Birds fly, unless they are abnormal birds": $\forall x\ bird(x) \wedge \neg abnormal(x) \Rightarrow flies(x)$. The class of abnormal birds would then be restricted as much as possible. That means that for instance an ostrich would be unable to fly, but until it can be proven that a bird is abnormal in some way it would be valid to conclude that it can in fact fly (which hits very close to how common-sense reasoning is performed).

The formalisation of Circumscription as stated in [McCarthy, 18] is as follows. Let S(P) be a second-order logical sentence with P being a tuple of some of the free predicates in S(P), and let E(P,x) be a well-formed formula in which P and a tuple of individual variables, x, occur free. The circumscription of E(P,x) relative to S(P) is $S'(P)$ defined by:

$$S(P) \wedge \forall P'[S(P') \wedge [\forall x\ E(P',x) \supset E(P,x)] \supset [\forall x\ E(P',x) \equiv E(P,x)]$$

This formalisation states that the collection of P and x that can satisfy E(P,x), is also necessarily the only collection of P and x that does so. McCarthy, in [18], says himself that "*Circumscription is not a "non-monotonic logic''. It is a form of non-monotonic reasoning augmenting ordinary first order logic*". While for the 1985 redefinition this is still true except that it applies to

2) The Tweety Bird problem: "all birds fly", "Tweety is a bird", "Does Tweet fly?" The conventional answer would be 'yes', but if you later learned that Tweety had a broken wing, then the answer would have to be 'no'. This invalidation of prior conclusions by new axioms added to the model is not expressible in monotonic logic.

second order logic rather than first order logic, the advantage of this approach over normal monotonic approaches should be clear: instead of restricting scenarios by ruling out all sorts of highly improbably possibilities within it, thus creating a possibly false world model (such as when using the modal language Z), this approach lets one get away with only stating the known world model, and assuming there are no abnormalities until these are explicitly added to the model, for whatever reason.

McCarthy proposed extending this system to allow for temporal reasoning, and in doing so, seemingly solved the Frame Problem [McCarthy, 19], by defining a rule for saying which fluents stay the same (or 'hold') over time, formalised as $\forall$ *f, e, s*: Holds(*f,s*) $\wedge$ $\neg$ abnormal(*f,e,s*) $\Rightarrow$ Holds(*f, result(e,s)*), where *f* is a fluent, *e* an event and *s* is a situation. With this addition, it seemed that McCarthy had solved the frame problem, by making inherent to the solution the fact that things that shouldn't be modified through an action, are in fact not modified through an action – without the use of frame axioms.

However, upon closer inspection, it didn't: There were two problems with this approach. One problem was that concurrent actions were still not taken into account, with for instance someone moving a block while someone else moves another block not being describable without additional rules that allow for the possibility of moving one block non-causally resulting in another block having moved. The second, and quite interesting one because of its anomalous nature, was that it suffered from the "Yale Shooting Problem", a special case of the multiple extensions problem, which was discovered independently by Hanks and McDermott, and Lifschitz in 1986. This problem was of such a nature that the circumscriptive approach led to completely counter-intuitive situations to be equally likely within the logical framework, while being disregarded as viable situations in the real world.

The Yale Shooting Problem and its solutions

The Yale shooting problem emerged as an anomaly when looking at an attempt to circumscribe the following situation: We have a gun, which gets loaded at time S0, after which we wait an indeterminate amount of time, ranging at the very least over S1, and at S2 we fire the gun at our best friend. Under normal circumstances this would be tantamount to murder, but in a circumscriptive approach it may actually be the case that our best friend might very well still be alive, because of two assumptions made in circumscribing this situation: the first assumption is that humans are generally alive, and the second assumption is that guns don't spontaneously unload themselves.

The expected situation, where we kill our best friend, violates the first assumption through the *shoot* action, but it is also possible that the second assumption is violated instead: while not explicitly stated anywhere, the circumscriptive approach allows for the *possibility* that the gun becomes unloaded some time during a *wait* action. Even though the first situation would be intuitive and the second would not be, both are equally likely when circumscribing this scenario, giving rise to a reasonably serious problem.

To formalise this problem, we can describe the world as allowing three actions: *load*, *wait* and *shoot*, and allowing two fluents, *loaded* and *alive*, which apply to the gun and our best friend respectively. The action *load* has no preconditions, and effects *loaded* becoming true. The action *shoot* has *loaded* as its only precondition and has as effect the negation of *alive*, with finally *wait* having no preconditions and no effects. Lastly, we add the circumscription assumptions that “People are alive, unless they are abnormal” and “Guns don’t spontaneously unload, unless they’re abnormal”. We formalise the actions *Load* and *Shoot* as follows, leaving *Wait* out, since it has no preconditions or effects:

Load	$\forall s\ \text{Holds}(\text{loaded}, \text{RESULT}(\text{load}, s))$
Shoot 1	$\forall s\ [\text{Holds}(loaded, s) \rightarrow \text{Holds}(\neg alive, \text{RESULT}(\text{shoot}, s))]$
Shoot 2	$\forall s\ [\text{Holds}(loaded, s) \rightarrow \text{Holds}(\neg loaded, \text{RESULT}(\text{shoot}, s))]$

Also, we start out in situation S0: Holds(alive,S0) and ¬Holds(loaded,S0)

The “problem” in the Yale Shooting Problem as outline earlier is then the one where these axioms lead to the situation where the gun is unloaded, and our best friend is still quite alive. The intuitive extension is the following scenario:

S0 →	load: S1 →	wait: S2 →	shoot: S3
alive	alive	alive	¬alive (*default violation*)
¬loaded	loaded	loaded	¬loaded

However, using circumscription the following extension is also possible:

S0 →	load: S1 →	wait: S2 →	shoot: S3
alive	alive	alive	alive
¬loaded	loaded	¬loaded (*default violation*)	¬loaded

The interesting part of this problem is that both extensions violate the same number of default assumptions: the first extension violates the default that people in general are alive, which includes our best friend, where the second violates the default that a loaded gun does not spontaneously unload. If approached numerically, both extensions are (logically) equally “not right”… How then, is this problem solved?

The chronological approach to a solution

The first approach to solving this problem was done by enforcing a forward-in-time order of reasoning [Kautz, 15; Shoham, 32]. In the expected model, reasoning occurs from the earliest to the latest time moment, and unexpected models (where changes take place before the latest possible moment) are disqualified. This approach leads to the conclusion that our best friend is quite dead after we shoot the gun, since the unloading does not occur until the shoot action is performed, but while it solves the Yale Shooting Problem it's not enough to solve the frame problem: the inability to reason back in time meant that for instance in the case that one could park ones car somewhere in the morning, come back in the afternoon to find it gone, and would have to conclude it must have disappeared the moment before the discovery was made that the car was gone. Additional model rules that allowed for earlier 'removal' of the car would have to be invented for the extension to be less artificial.

The causal-based approach to a solution

Having the problem that chronological approaches lacked the ability to reason backwards in time, causal-based solutions were suggested instead, by people such as Lifschitz (1987), Haugh (1987) and Baker (1991). These differed from chronological approaches in that they offered a valid reasoning both forward and backward reasoning over time, and that they offered a solution to the Yale Shooting Problem that was strongly intuitive.

Of these, Lifschitz' suggestion gained the most ground as possible solution, and was based on the concept of an affects(a,f,s) predicate that states that a fluent f is affected by an action a in situation s when the action changes the truth value of the fluent, expressible as [$\text{affects}(a,f,s) \Leftrightarrow \text{success}(a,s) \wedge \exists v\ \text{causes}(a,f,v)$], where v is the truth value for the fluent. An action a is then successful if all of the preconditions for it hold, according to [$\text{success}(a,s) \Leftrightarrow \forall f\ (\text{preconditions}(f,a) \Rightarrow \text{holds}(f,s))$]. The actual axiom that describes how fluents change is formalised as:

$$\text{success}(a,s) \wedge \text{causes}(a,f,v) \Rightarrow (\text{holds}(f,\text{result}(a,s)) \Leftrightarrow v = \text{true})$$

According to Lifschitz, subsequently minimising the preconditions results in a solution for the qualification problem, where minimising causes would solve the frame problem. This solution handled the Yale Shooting Problem, through the fact that there is no axiom that allows a wait action to unload the gun we have in our hands, because of the minimisation on causes that this system prescribes.

It turned out that Lifschitz´ original solution, however, had some quirks left in it, which warranted that in some scenarios if the outcome was not as predicted, a “miracle” had happened. This was perceived as something that did require a less mystical explanation, and over the course of several years, Lifschitz’ theory was expanded on by various people, fixing the small problems that persisted in it, but bringing with them small problems of their own again.

It seem that, after the whole Yale Shooting Problem, solutions to it, holes in the solutions, plugs for the holes, and more holes in the plugs themselves, reasonably clear that coming up with solutions to only the Yale Shooting Problem was not the way to go about solving the real problem, namely getting around the Frame Problem. In the chapter on the modern solutions to the frame problem, we will see how the problem was eventually tackled through the basis laid out by the work of all the people that had brought forth potential solutions to the frame problem, and we will see just how pivotal the solutions from the heydays of the Frame Problem were to the modern solutions, or partial solutions, to the problem that has sparked the most debate concerning logic in AI.

4.2 Alternative approaches

With many people coming to the conclusion that the frame problem was one of form, rather than content – that is to say, a problem innate to the use of logic, rather than being a problem related to exactly which approach within the logical framework was taken – there were multiple publications on approaches that resulted in a system that performed the role of systems for which formal logic had been previously used, but did not run into the frame problem due to the non-logical approaches used.

Concepts such as statistical or probabilistic approaches towards time-based reasoning, using the world as its own representation for situated robotics, and the use of connectionist or neural networks to operate in the world were suggested as alternatives to the logical approach. While it was clear that none of these solutions effectively “solved” the frame problem – circumventing it by taking a non-logical approach to particular tasks that had been performed using formal logic – they nevertheless received some attention as ‘solutions’ to the Frame Problem in the broader sense, as they were quite capable of performing some tasks formal logic had yet to facilitate.

4.2.1 Statistical and probabilistic approaches

One approach to the whole modelling of the world problem, which sparked the Frame Problem in the first place, is to deal with it probabilistically. The field of statistics offers us things such as Bayesian networks that allow us to reason about cause and effect, and by adding temporal probability to the mix it is quite possible to reason about events and causation over a timeline. This approach is taken by Pearl [23], who for instance solves the Yale Shooting problem as follows:

Assuming a stance similar to default logic, there are 5 rules, of which two (LD1 and ¬AL2) need to be derived in some way from the other three:

LD0 – the gun is loaded at time t0

LD1 – the gun is loaded at time t1

SH1 – the gun is fired at time t1

AL1 – our best friend is alive at time t1

AL2 – our best friend is alive at time t2

Probabilities are assigned so that P(LD1|LD0) = high (= 1 - ε, where ε is an insignificantly small number), P(AL1|AL2) = high, P(AL2| AL1, SH1, LD1) = low (= ε) and P(AL2 | AL1, SH1, ¬LD1) = high. Using this, Pearl demonstrates that solving this problem using 1-ε probabilities, and the principle that causal structure is assumed on the model, the problem can be successfully solved, but this assumption comes at a price: Pearl remarks that "probability does not offer a complete solution to the frame problem because it does not provide rules for recomputing the summaries when unanticipated refinements are warranted" [23, pp.23]. Pearl does however remark that while probability alone is not the answer, probability does offer that which is missing in logic, suggesting that a synthesis of the two – where logic handles the visible part of the world, and probability handles the invisible parts – might be the answer to the frame problem.

4.2.2 Physical grounding

The other end of spectrum has approaches taken by people such as Brooks, using modular hardware components to achieve a form of interaction between agents and the world, without an explicit internalisation of the world. The frame problem doesn't pop up in this approach, because nothing at all is formalised, relying on reflex behaviour to achieve behaviour that seems to be intelligent. Brooks rejects the idea of symbol grounding – the internal representation of the world in terms of symbols, and cognitive processes operating on these symbolic representations – on the basis of psychological evidence (as stated in [Brooks, 1]) that perception is an active, task dependent operation, and as such the symbolic representation would be different depending on the task at hand. According to Brooks, symbol systems attempt to model an absolute knowable and objective truth about the world, while the attempts to facilitate this through logic creates more and more biologically implausible systems for reasoning.

Rather than using any formalisation through abstraction, Brooks' approach uses the world as such as the model. Through this, the need for a symbolic representation stops being a requirement for the system to operate in the world. Brooks thusly avoids falling into the trap he believes many fall in when trying to approach modelling and AI as a whole: "*One of the most popular ideas is generality. This quickly leads to a disease I call puzzlitis. The way to show generality is to pick the most obscure case within the domain and demonstrate that your system can handle or solve it.*" [Brooks, 1] but in doing so, also believes a logical approach to be essentially an exercise in futility, without any real proof of this perceived fact.

4.2.3 Connectionism

Another "blind" approach in the sense that there is no explicit world representation, can be found in the so-called connectionist or neural networks models. While it had gone out of fashion for a while as a serious technique in the field of AI, the use of neural networks had been rekindled with the publication of "*Connectionist models and their properties*" by Feldman and Ballard in 1982 [7], and the fact that these systems were able to successfully classify data without prior knowledge put into them. We will not actually go into the details of any particular approach to solving the frame problem, but rather explain why the frame problem doesn't exist when using a connectionist model to implement intelligence.

Similar to how Brooks' approach never runs into the Frame Problem, so too is the neural networks approach free of the Frame Problem. Rather than explicitly representing the world in any logical language, the neural network approach allows the system to come up with its own model, and corroborate this model back to the world, to determine whether it's effectively internalised those parts of the world it should work with or not. Neural networks can do this, as they are made up of many elements that all process part of an incoming signal without knowing what the complete signal looks like. The end result of the processing is compared to what the end result should be, and if there is a significant discrepancy, the network can update (through various cascade functions) the way the signals are processed in each element, or node, in the network. This updating is performed entirely by the system itself, and the end result is (ideally) a network that has, at least in some form or other, an internalisation of that part of the world that is relevant to its processing.

As neural networks are a dynamic system in that they can continuously prime themselves for the task they are intended to perform (though not with the guarantee of ever succeeding – see [Feldman & Ballard, 7]) there are no real axioms or internalisation added by the designers at the start, but rather the internalisation of the world "forms" over the course of this priming, to correspond more and more with what seems to actually be the case, with the actual representation being nothing more than weight distribution values that say which part of the incoming signal is sent on with which adjusted strength to which other nodes in the network.

Again, but differing from the physical grounding approach in that even if it’s not explicit, an internalisation of the world is performed, this is not a true solution to the Frame Problem but a circumvention. While neural networks are a very interesting part of the AI field, this approach means that it is very hard to prove anything about the abilities of a system that employs it, due to the next to impossible level of determinism in these systems once they have been allowed to modify themselves. As such, while a potential approach to AI in general, it is one that is essentially highly non-scientific at the internal level. While the math that drives the system has been worked out to the meticulous detail, the actual working systems are not easy to understand after more than a few iterative priming runs. This makes it very hard to understand why an internalisation that works, works, and as such it would be unlikely that a theoretical basis could be found for a particular internal state a neural network comes up with that seemingly works for a given situation in the world.

4.3 In conclusion

The frame problem was, and stayed, an open problem. Many interesting solutions to it had been proposed, some of which offered – and still offer today – a functional way to handle logical world modelling. However, none of the proposed solutions worked for concurrent actions, which is probably the pivotal problem in dealing with the Frame Problem. While it goes without saying that a true solution to the single-action Frame Problem would be a major step forward in logic, the ultimate goal is of course to use the solution in agents that interact freely with the world, be this world physical or virtual, where concurrent actions by both single as well as multiple agents are only to be expected.

In the next chapter, we will focus on the modern approaches to solving the frame problem, building on the foundations that had been laid out during the turbulent years of the late 80’s and early 90’s, to see whether current approaches fare better at tackling what is arguably still the more tentative problem in the logical approach to AI and knowledge representation.

5 – Solutions to the Frame Problem in Modern Logics

This chapter will describe two modern logic approaches to the logical frame problem. They will be presented in chronological order, starting with Shanahan's Circumscriptive Event Calculus (1997), and following with Thielsher's Fluent Calculus (2001). These approaches have been widely accepted as being solutions to the Frame Problem and are therefore worth describing. They will also be compared to Morgenstern's requirements [22], as they provide a solid guideline for evaluating solutions to the Frame Problem.

5.1 Shanahan's Circumscriptive Event Calculus

Circumscriptive Event Calculus is described by Murray Shanahan in *"Solving the Frame Problem: A Mathematical Investigation of the Common Sense Law of Inertia"*, published in 1997. Shanahan introduced a way to use Circumscription in combination with Event Calculus, resulting in axioms that produce satisfactory conclusions when dealing with the Yale Shooting Problem, ramifications and concurrent events, without formulating the amount of Frame Axioms.

5.1.1 Shanahan's Criteria to a Solution

Shanahan himself points out three criteria a satisfactory solution should meet, being quite comparable to the requirements we've selected from the ones dictated by Morgenstern in [22]. They are

- Representational Parsimony
- Expressive Flexibility
- Elaboration tolerance

He considers the representational parsimony to be the essence of the Frame Problem. It insists that representation of the effects of actions should be in proportional to the complexity of the domain, dictated by the sum of the total number of actions, A, plus the total number of fluents, F. This criterion could be compared to our fourth requirement, 'Conciseness'. Shanahan claims that his criterion can be met by formalising 'the Common Sense Law of Inertia', which dictates *"the assumption that besides known effects, nothing else changes"* [31].

Expressive flexibility insists that the solution will carry over to more complicated domains. Not more complicated in the sense of 'more actions and/or fluents', but domains which contain ramifications, concurrency, non-determinism and continuous change. This criterion is harder to compare to our requirements, but our first requirement ‘Solving the right problem’ does require that a solution “*should be compatible with a general theory of temporal reasoning*” [Morgenstern, 22].

Finally, elaboration tolerance insists that when a situation is added with a new action that influences n fluents, roughly n sentences would have to be added to the theory, without having to reconstruct the original theory. Ideally, this would mean that appending new sentences to the old theory would yield the new theory. This is easily compared our fourth requirement ‘Conciseness’.

5.1.2 The Principles of Shanahan’s Solution

Before an accurate description of Shanahan’s solution can be made, two principles must be explained on which Shanahan’s solution is based upon. One of these have already been explained, Circumscription, but Shanahan has reinterpreted the definition of the method to fit his solution, of course not without proving that this reinterpretation is essentially the same as the original principle. The workings of the other principle, event calculus, will be described next.

Event Calculus

The event calculus was originally introduced by Kowalski and Sergot [17]. They proposed a calculus that was more versatile in its applicability than situation calculus. Situation calculus had mainly two limitations, the first being that situations are instantaneous points in time, making it hard to describe change occurring continuously over time, such as for example the gradual growth of a plant. Second, situation calculus has a hard time describing concurrency as situations are mapped from one to the other by a single action. Of course, multiple actions could be described as compound actions (i.e. actions Hit and Run could be described as Hit_and_Run), but those compound actions would need new frame axioms. Event calculus has a way of describing the previous occurrences without compound actions and has as much expressive power as the situation calculus.

Event calculus introduces a notion of linear time through events. An event could be best described as a piece of this universe with both temporal and spatial extent, for example, the event of a tap being turned on and, after a while, being turned off.

The symbol *HoldsAt* essentially has the same function as Holds in situation calculus, although it relates fluents to moments in time, instead of to situations. A description of water flowing from the tap at time t1 would be:

HoldsAt(Water_Flows_From_Tap, t1)

Now Happens is introduced, a predicate relating actions to moments in time. To describe the tap being turned on at moment *t0*:

Happens(Turn_Tap_On, t0)

A description of the fact that turning the tap on makes the water flow is constructed using Initiates:

Initiates(Turn_Tap_On, Water_Flows_From_Tap, tN)

Notice that time is included in this description. This is necessary if a certain fluent Happens at an earlier time that might prevent the flowing of water from happening, for example, the mechanism of the tap being sabotaged. Terminates, the opposite of Initiates, could be used to describe the fact that turning the tap off will make the water stop flowing.

Finally, *Clipped* is included. *Clipped* is used to describe that something happened between two moments in time, causing a certain fluent to be false. *Clipped* is a tool to reason about periods of time, for instance, to describe the fact that the flow of water has been stopped somewhere between *t0* and *t*:

Clipped(t0, Water_Flows_From_Tap, t1)

Reasoning about the gradual growth of a plant is made possible by the introduction of time, and through this, reasoning about differences between points in time is also made possible. This formalism can be used to reason about concurrent events, using multiple *Happens* axioms with the same time argument *t*. *Clipped* offers reasoning about time periods were there is incomplete information about the course of events as it just states *something* happened between two points in time. This, in a nutshell, is event calculus.

Circumscription by Shanahan

Circumscription as described earlier in this paper, is very important to Shanahan's solution. The basic idea is to limit of the set objects for which a predicate is true, also known as the minimisation of a predicate. Shanahan defines circumscription as follows [Shanahan, 31]:

Definition 7.1 Let $\phi(\rho)$ be a formula that mentions the predicate symbol ρ. The circumscription of $\phi(\rho)$ minimising ρ, written CIRC[$\phi(\rho)$; ρ], is the second-order formula,

$$\phi(\rho) \wedge \forall q\ [[\phi(q) \wedge q \leq \rho] \rightarrow \rho \leq q]$$

- $q \leq \rho$ meaning that $\forall x\ [q(x) \rightarrow \rho(x)]$ □

An example will clarify this definition. Let $\phi(\rho)$ denote the following formula:

$$Bird(Tweety) \wedge Bird(Roadrunner) \wedge Black(Sylvester)$$

Can it be shown that $CIRC[\phi, Bird] \not\models \neg Bird(Sylvester)$? The answer is yes, if one considers $Q(x) \equiv_{def} [x = Tweety \vee x = Roadrunner]$. If we instantiate q to Q in definition 7.1 than the left part of the implication, $[\phi(Q) \wedge Q \leq Bird]$, would be true. Thus, the right part of the implication in definition 7.1, $Bird \leq Q$, is necessarily true. Put in to Conjunction with $\phi(\rho)$, the following is necessarily true: $\forall x\ [[x = Tweety \vee x = Roadrunner] \leftrightarrow Bird(x)]$. Thus the set of instances for which $Bird(x)$ is true is minimised by circumscription.

A problem arises if circumscription is applied to event calculus when it is confronted with the Yale Shooting problem, as was the case with situation calculus. If the Yale Shooting problem is defined in terms of the event calculus resulting in a domain description Σ, then there are conceivable models of Σ in which *Terminates* is already minimised according to the previous description that implicate that *Terminates(Wait, Loaded, t)* is true. Again, satisfactory conclusions cannot be made using circumscription as the gun is magically unloaded by waiting, saving our best friend once again. Shanahan, however, formalised an event calculus fit for circumscription.

5.1.2 Circumscriptive Event Calculus

With the circumscriptive event calculus Shanahan reintroduces situations into event calculus, which he defines as states. He introduces them to gain the benefits of the previous described circumscription for event calculus, without the formalism losing its expressive power. Shanahan claims it's safe to index Initiates and Terminates on states because he defines a state as a set of fluents, therefore being timeless. For example, a proper description of a set s would be $s = \{F1, F2, F3\}$. This is very important as this set defines frame fluents, as Shanahan calls them, on which his 'common sense law of inertia' applies. The next axioms are being presented as the first part of Shanahan's formalism:

$$s1 = s2 \leftrightarrow \forall f\ [In(f,s1) \leftrightarrow In(f,s2)] \qquad (S1)$$

$$\forall s1,f1\ \exists s2\ \forall f2\ [In(f2,s2) \leftrightarrow [In(f2,s1) \vee f2 = f1]] \qquad (S2)$$

$$\exists s\ \forall f\ [\neg In(f,s)] \qquad (S3)$$

- *In(f,s)* represents that fluent f is a member of the set s.

To put these in words:

'Two sets are equal if and only if they contain the same fluents.' (S1)

'Any fluent can be added to any set to give another set' (S2)

'An empty set exists' (S3)

It is very important to point out that Shanahan also defines the set state to be able to contain negations of fluents. For example, *s* could be described as *s = {F1, Not(F2)}*. Thus, it is now possible to tell which fluents holds in a certain state and which fluents do not hold:

$$HoldsIn(f,s) \leftrightarrow In(f,s) \quad (E1)$$

$$\neg HoldsIn(f,s) \leftrightarrow In(Not(f),s) \quad (E2)$$

So there are now there are two ways of discerning whether a fluent holds in a situation:

- If state *s = {F1, Not(F2)}*, then *HoldsIn(F1,s)* is true and *HoldsIn(F2,s)* is false.
- If state *s = {F1}* and $HoldsIn(F1) \rightarrow \neg HoldsIn(F2)$ is added to the domain description, then again *HoldsIn(F1,s)* is true and *HoldsIn(F2,s)* is false.

This is also a way of discerning which fluents are primitive and which are derived. The fluents that are part of state descriptions are primitive, dictated by the situation. The truth-value of fluents dictated by domain constraints, but which are not part of the set that define the state are called derived fluents. Next, Shanahan makes a connection between states and time, using the axiom:

$$State(t,s) \leftrightarrow \forall f1\ [in(f1,s) \leftrightarrow [Initiated(f1,t) \vee \exists f2\ [f1=Not(f2) \wedge Terminated(f2,t)]]] \quad (E3)$$

Now the whole idea is to let the 'common sense law of inertia' apply to the primitive fluents, thus only letting those primitive fluents change that have been directly affected by an event. This will make frame axioms obsolete as the only way a fluent can become negated in the set which defines the state is when it is explicitly *Terminated*, efficiently solving the Frame Problem! The original objection concerning the impossibility of concurrent events made by Davis (who coined the terms primitive fluent and derived fluent in 1990) is not applicable as the event calculus has a way of dealing with concurrent actions by letting multiple events happen at the same time *t*.

Finally, the formula *HoldsAt(f,t)* represents that a fluent *f* holds in situation *s* present at time *t*.

$$HoldsAt(f,t) \leftrightarrow \exists s\ [State(t,s) \wedge HoldsIn(f,s)] \qquad (E4)$$

This final axiom makes this calculus compatible with standard notation of previously proposed problems as the Yale Shooting Problem, so that it can come up with satisfactory conclusions.

The conjunction of axioms (S1) to (S3) with (E1) to (E4) provides Shanahan's event calculus (EC). This EC in combination with circumscription will produce satisfactory results when confronted with classic problems related to the Frame Problem. The next section will provide an example.

5.1.3 The Application of Circumscriptive EC

Next the application of Circumscriptive EC on the Yale Shooting scenario is described. The domain for the Yale shooting scenario consists of three types of event: *Load*, *Sneeze* and *Shoot*. As might be noticed, in the event calculus description of the Yale shooting problem *Wait* has been replaced by *Sneeze*. This is because if the original *Wait* action is used instead of *Sneeze*, then there would a difference in time between the 'loading event' and the 'waiting event'. But that period of time can already be described as waiting, thus resulting in a rather awkward description of a situation: 'First someone loads a gun, then he does nothing until he starts waiting'. *Sneeze* will prove to have a function equivalent to *Wait* as it provides an opportunity for the minimisation to go wrong.

The domain also consists of two types of fluent, which are: *Alive* and *Loaded.* All these are formalised as follows, denoted by Σ^{YSS} (Note that there is no axiom for Sneeze as it does not affect any fluent!):

Initiates(Start, Alive, s) (CE7)
Initiates(Load,Loaded,s) (CEY1)
$HoldsIn(Loaded, s) \rightarrow Terminates(Shoot, Alive)$ (CEY2)

Happens(Start, 0) (CE6)
Happens(Load, 10) (CE3)
Happens(Sneeze, 15) (CE4)
Happens(Shoot, 20) (CE5)

Shanahan's Circumscriptive EC will come to the conclusion $\neg\ HoldsAt(Alive, 25)$, which is what could be described as being satisfactory. It will come to this conclusion, because

$CIRCec[\Sigma^{YSS}] \models State(20,s) \leftrightarrow \forall f\, [In(f,s) \leftrightarrow [f = Alive] \vee [f = Loaded]].$

Which means that, since the event of loading the gun at time $t = 10$ the states until $t = 20$ will be defined by *{Alive, Loaded}*. The *Sneeze* event at time *15* did not influence any fluent in these fluent sets, because the EC axioms were used in combination with circumscription. As (CEY2) dictates that event *Shoot* at time $t = 20$ will result in the termination of event *Alive*, all states at time $t > 20$ will be denoted by the set *{Loaded}*. Thus,

$CIRCec[\Sigma] \models \neg HoldsAt(Alive, 25).$

Shanahan describes multiple applications, showing satisfactory conclusions using the previous method of proof. The Circumscriptive EC, for example, properly processes ramifications, as derived fluents follow from the axioms describing the domain. Shanahan also proves that Circumscriptive Event Calculus produces satisfactory conclusions when dealing with non-determinism and concurrent events without an addition of an intolerable amount of axioms.

5.1.4 Does this solve the Frame Problem?

It would seem that Shanahan produced something that is very close to a solution to the Frame Problem. Again, it must be pointed out that there is no universal consensus to what a solution should consist of, but one could say that Shanahan is indeed very close to a solution if the next points are taken into account.

Shanahan has indeed succeeded in obtaining his requirements. The number of axioms needed to describe a domain is proportional to the complexity of the domain, dictated by the sum of the total number of actions, *A*, plus the total number of fluents, *F*. Shanahan's first requirement is also commonly viewed as being the logical criterion for solving the frame problem, so it could be said that Shanahan has found a way to formalise 'the common sense law of inertia', resulting in satisfactory conclusions about a certain domain, without needing an endless list of frame axioms.

It would also seem that Shanahan's solution is concurrent to the requirements dictated by Morgenstern as Shanahan not only aptly describes the logical Frame Problem, he sets requirements which are commonly viewed as necessary for a solution to the logical Frame Problem and founds his solution on sound theory. Also, as Shanahan states himself, his method is quite general purpose and not limited to reasoning about action.

5.2 Thielscher's Fluent Calculus

There is another calculus which offers a solution to the Frame Problem and which is quite popular among programmers that develop Cognitive Agents, Thielscher's fluent calculus (FC). It is widely accepted as being a solution to the Frame Problem in situation calculus, and the essence of the solution is quite similar to Shanahan's way of solving the Frame Problem as we will see later on.

Thielscher's FC is based on classical situation calculus and Reiter's successor situation axioms, which have been mentioned earlier. First, a more elaborate explanation of certain aspects of Reiter's calculus is described because they founded certain aspects of Thielscher's calculus.

5.2.1 More Elaborate Description of Reiter's Solution

Combining Pednault's effect axioms and Schubert's explanation closure axioms, Reiter formalised the following axioms as described earlier in section 4.1.2 [Reiter, 24]:

1. For each fluent R, successor situation axioms,:

   $Possible(a,s) \supset [R(do(a,s)) \equiv \gamma^{+}_{R}(a,s) \vee R(s) \wedge \neg \gamma^{-}_{R}(a,s)]$

   - $\gamma^{+}_{R}(a,s)$ denotes the positive fluent preconditions of the fluent R, which must be satisfied in order that fluent R can become true in the successor state do(a,s).
   - $\gamma^{-}_{R}(a,s)$ denotes the negative fluent preconditions of the fluent R, which must be satisfied in order that fluent R will become false in the successor state do(a,s).
2. For each Action A, a single action precondition axiom of the form:

   $\pi A(s) \supset Possible(A,s)$

   - $\pi A(s)$ denotes the preconditions defining that action A is possible in state s

Reiter displayed that these axioms will effectively function as frame axioms. The action precondition axiom will define all the preconditions for being able to perform action *A*, and successor situation axioms define what fluents can change if *A* is performed. Now the only way a fluent can change truth-value is if it is explicitly described by these axioms. These axioms only work under the Completeness Assumption for Fluent Preconditions though as described earlier, but Thielscher has enveloped these concepts in fluent calculus without this assumption being necessary.

5.2.2 Thielscher's Fluent Calculus

Thielscher's described the fluent calculus in a report, aptly named 'the Fluent Calculus' and published in 2001. It has become popular as it has been designed as a specification language for robots that have to meet the requirement of being able to address the following aspects in a uniform way [Thielscher, 33]:

1. Non-determinism and Uncertainty
2. Knowledge and Sensing Actions
3. Ramifications
4. Concurrency

Aspect 2, 'Knowledge and Sensing Actions', is an aspect that dictates the problem that a robot might need to actively sense his surroundings, updating its knowledge about the world, before it can start to plan ahead. As this aspect is not a universal requirement for a solution to the Frame Problem it will not be of any importance here. Aspects 1, 3 and 4 are classics though, and the fluent calculus has a way to deal with them in a satisfactory way.

Fluents and States

It is of importance to note that the fluent calculus is a logic of sorts, which means that, among others, it reserves the sort FLUENT for fluents and it reserves the sort STATE for states. It also has a striking similarity to Shanahan's event calculus as states are defined as a collocation of fluents:

InitialState = InRoom(R401)

State1 = InRoom(R401) ∘ AtDoor(D12)) ∘ HasKey(D12)

As can be seen, a state may be defined by a single fluent, thus FLUENT is a subset of STATE. The functions "∘" and "-", which are of type STATE $\times$ STATE $\rightarrow$ STATE, describe a mapping of two states into a new state in the following way :

State1 = InitialState ∘ AtDoor(D12)) ∘ HasKey(D12)

InitialState = State1 - (AtDoor(D12) ∘ HasKey(D12))

This definition of fluents and states is crucial to Thielscher's approach to the Frame Problem as we will see next.

State Update Axioms

Thielscher defines state update axioms that are based on Reiter's work. He describes state update axioms, which define the effects of action A in the fluent calculus, as follows,

$$Poss(A(x),s) \wedge \Delta(x,State(s)) \supset State(Do(A(x),State(s)) = State(s) \circ \theta + - \theta -$$

$\Delta(x,State(s))$ defines the additional preconditions for the defined change of fluents. For example, an action *Inhale* might need two state update axioms where $\Delta(x,State(s))$ might be *AboveWater(State(s))* in the first and *BelowWater(State(s))* in the second axiom. This replaces Reiter's preconditions for a fluent to become true or false, as it actually is a precondition for the axiom to be applicable. Next, θ^+ defines the positive effects and θ^- defines the negative effects in the situation, thus updating state appropriately. So as you can see, change is modelled by specifying the difference between two states, and this explicit notion of states is how the fluent calculus approaches the Frame Problem. Under the assumption that positive and negative effects are disjoint, state update axioms describe only fluents that change, thus making frame axiom formulations obsolete.

Action preconditions are defined the same way as Reiter, so the total amount of axioms needed to describe a domain will roughly be $A + F$. For example, a situation with 500 fluents and 100 actions will be adequately described by 600 axioms in total, which is a minimal amount in comparison to the 50,000 naïve frame axioms needed in classical situation calculus!

5.2.3 Application of the FC

Thielscher's FC is the foundation of FLUX (Fluent Calculus Executor), a constraint logic programming implementation. This programming language is a practical application of the FC, which is simple in concept and capable of dealing with incomplete states, and the internet provides online demonstrations of the precise workings of FLUX on *http://www.fluxagent.org/*.

Theoretical application of the FC, however, is quite straightforward and when its approach to the Frame Problem is taken into account, it does not differ that much with Shanahan's solution. However, time is not included in Thielscher's FC, so there are some major differences in formalising ramifications and concurrent events.

Thielscher reserves the predicate *Causes*, of type FLUENT $\times$ STATE $\rightarrow$ FLUENT, to describe which event triggers what ramification. Next, Thielscher reserves the predicate *Ramify* to relate situations by applying a series of fluents that cause other fluents to be become true or false. It might seem that this will cause an addition of an intolerable amount of frame axioms, but this is not true. The simple FC can be expanded by adding *Ramify* to state update axioms, and by extending the domain with axioms that define what a fluent *Causes*, if it causes anything. This will prevent the domain description from becoming larger than the domain complexity.

Concurrent actions are described as compound actions and again, this does not cause the FC to fail meeting the requirement of representing complexity. Successor state axioms must be defined for concurrent axioms, where needed. Only state update axioms are defined where concurrency changes the results in comparison to sequential treatment of these actions. Otherwise, actions are processed sequentially.

5.2.3 Does this solve the Frame Problem?

Reiter is less modest than Shanahan by blatantly stating that the state update axioms are provably correct solution to the Frame Problem, without stating a definition of the Frame Problem [Thielscher, 33]. But comparing his fluent calculus to our criteria, his claim might be right nonetheless. His fluent calculus meets all the requirements, though differences to Shanahan must be pointed out. First, our third requirement, ‘Applying to concrete examples’, is definitely applicable to Thielscher's approach. FLUX is a well-known and often used programming language, producing concrete results. But Shanahan’s approach is more faithful to the first requirement. As Thielscher did not include time in his calculus, reasoning about gradual events is quite difficult in the fluent calculus. However, The fluent calculus meets the fourth requirement, which can be described as the essence of the Frame Problem, ‘Conciseness’. Thielscher’s fluent calculus has widely been accepted as a solution to the logical Frame Problem.

5.3 A Conclusion about State Representation

Shanahan’s and Thielscher’s notion of states are quite similar, and this might just be the essence of a solution to the Frame Problem. Both Shanahan and Thielscher define states to be a collocation of fluents that may be negated, called primitive fluents, which are true in a situation to which the state is linked, without the state being a complete description of a situation. Which fluents will be added or deleted from a state is dictated by the formalisation of domain axioms, ergo, the set of primitive fluents is defined by how domain axioms are formalised.

The ‘common sense law of inertia’ applies to the primitive fluents as a primitive fluent will not be removed from a state description as long as there is no axiom explicitly referring to the particular fluent. Information about other fluents in a situation can be derived from these primitive fluents, but these so called derived fluents do not contribute to a description of a state. Thus, there is no explicit notion of these fluents changing truth-value as actions occur and information about the truth-value of derived fluents in successor situations must be derived from the primitive fluents that the successor state consists of.

This element of both calculi makes the formulation of frame axioms obsolete, without the calculi losing their ability to formulate conclusions about complicated domains that are satisfactory by interpretation. Let the Frame Problem be solved.

6 – Conclusions based on thirty years of history

It is unlikely that McCarthy and Hayes could have foreseen the monster they had unleashed upon the world of AI and logic with their formulation of an at the time somewhat annoying problem which couldn't be solved with some more thinking and some more research. In the thirty years that it has taken the actual original logical frame problem to be solved, the vast number of additional problems that the Frame Problem exposed and the – at times – wild reactions towards its meaning and its implications for logic and AI has been staggering. In 1969 the problem was not much more than 'the problem of having to write down too many frame axioms', by the end of the 80's it had evolved into a problem that wasn't just receiving attention from nearly every conceivable sub-field of AI, but had also become a serious philosophical problem for which there might not be an answer at all.

In the beginning the original Frame Problem was not much more than a quirk in a logical system that was essentially the prototype of the situation calculus. However, it did not take long for people to come up with redefinitions of what the Frame Problem was supposed to be according to those that attempted to solve it, which lead to such problems as the induction problem, the problem of relevant information selection and the temporal projection problem, all being considered instances or even 'true forms' of the Frame Problem.

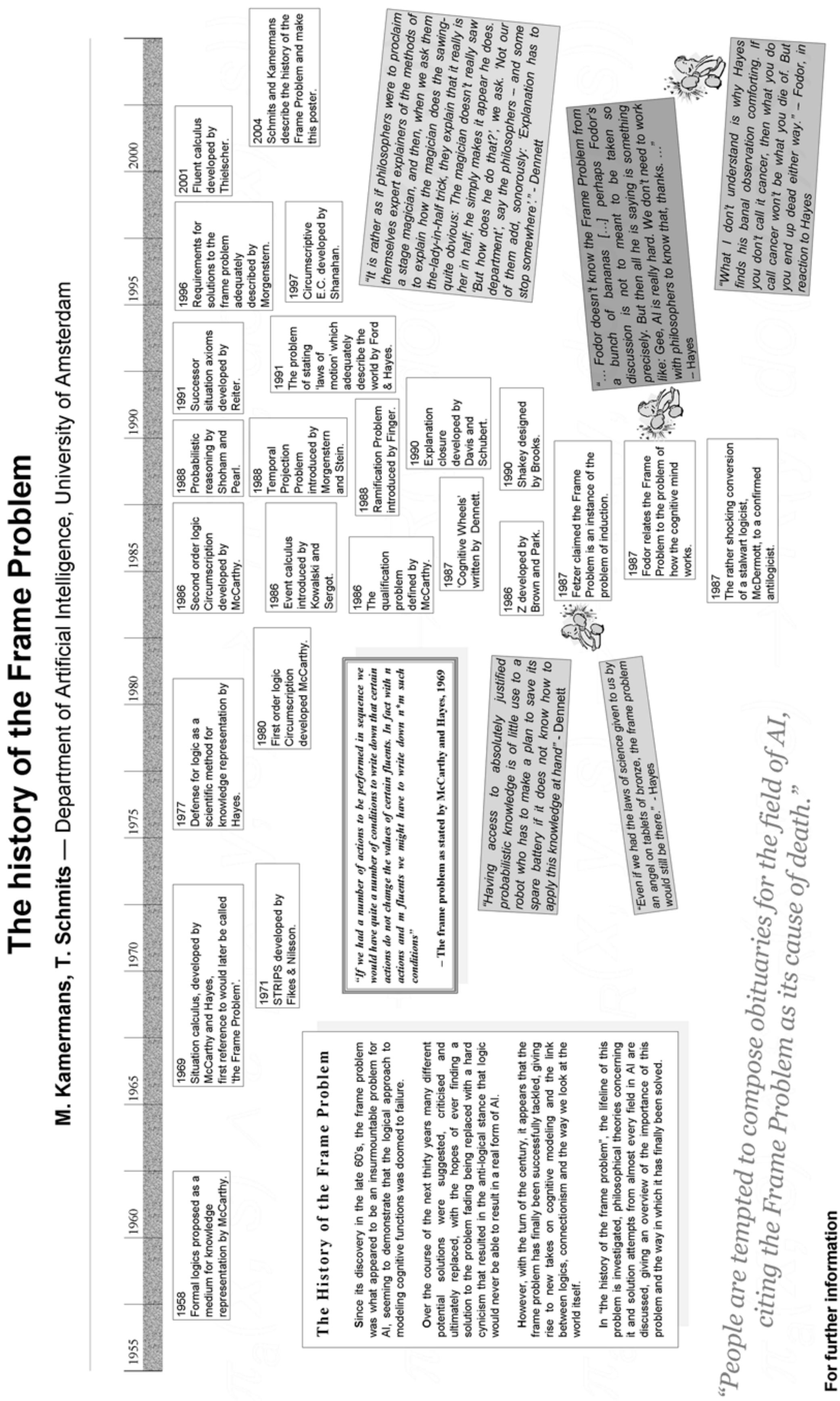
General interest in the Frame Problem was sparked even more with the publication of Dennett's article dealing with the Frame Problem in 1984, in which he treated it in a somewhat popular science form, claiming it was actually surprising no one had noticed this problem before. What followed was nothing short of a hype - the scientific community roused by Dennett's words not only jumped at the opportunity to demonstrate that the problem could in fact be solved if only the problem was described in its essence rather than the formulation initially used by McCarthy and Hayes. This lead to many publications, all of which attempted to describe this essence, followed by a solution to their description of 'what the problem really was.'

During this period of searching for proper formulation and solutions, the field of logic had come up with numerous partial solutions for the logical problem. And, with the turn of the century in sight, a solution to the logical Frame Problem was finally found, which was not only based upon proper logical theory but was also sound and correct. This solution was also remarkable because it built on various logical theories that all resulted from attempts to try to capture aspects of a 'bigger problem' that the situation calculus had missed, but had all in their own way failed to offer a true solution. But in the end, logic had solved a problem that it had itself created.

However, this leaves the question of the philosophical 'frame problem'. While logic had managed to, over the course of thirty years, solve the original frame problem and brought to a close a turbulent period in its long history, philosophy still struggles with the bigger issues that the Frame Problem brought to light.

Will the essence of the logical solution provide some insight for solving the deeper epistemological problem that the frame problem is taken for outside its logical context? To date, philosophy has yet to dare test the logical solution against the philosophical formulations of the frame problem, and so until it does, we shall be forced to wait for the first true philosophical analysis of the logical solution.

Appendix A – the timeline of the Frame Problem



Appendix B – bibliography

[1] Brooks, “Elephants don’t play chess”, *Robotics and Autonomous Systems (6), pp. 3-15*, 1990

[2] Brown, “A modal logic for the representation of knowledge”, *The Frame Problem in Artificial Intelligence – proceedings of the 1987 workshop pp.135-157*, 1987

[3] Brown & Park, “Actions, reflective possibility, and the frame problem”, *The Frame Problem in artificial intelligence – proceedings of the 1987 workshop pp.159-191*, 1987

[4] Davis, “Representations of commonsense knowledge”, *San Mateo, CA: Morgan Kaufmann*, 1990

[5] Dennett, Cognitive wheels: the frame problem of AI”, *The robot’s dilemma, pp.41-74 (1987) reprinted from “Minds, machines and evolution”*, *Ablex Publishing*, 1984

[6] Dreyfus, “What Computers Can’t Do”, *MIT Press,* 1972

[7] Feldman & Ballard, “Connectionist models and their properties”, *Cognitive Science 6 p205-254*, 182

[8] Fetzer, “The Frame Problem: Artificial Intelligence meets David Hume”, *Reasoning Agents in a Dynamic World: The Frame Problem, pp. 55-69, JAI Press*, 1991

[9] Fikes & Nilsson, “STRIPS: A New Approach to the Application of Theorem Proving to Problem Solving”, *Artificial Intelligence 2 no.3,4 - January, pp. 189-208*, 1971

[10] Fodor, “Modules, frames, fridgeons, sleeping dogs, & the music of spheres”, *The robot’s dilemma, pp.139-149*, *Ablex Publishing*, 1987

[11] Ford & Hayes, “Reasoning Agents in a Dynamic World: The Frame Problem”, JAI Press, 1991

[12] Hayes, “In Defence of Logic”, *Proceedings of IJCAI 77, pp.559-565*, 1977

[13] Hayes, “The Second Naïve Physics Manifestation”, *Formal Theories of the Commonsense World*, *Ablex pp.1-36*, 1985

[14] Hayes, “What the Frame Problem is and Isn’t”, *The robot’s dilemma, pp.123-138*, *Ablex Publishing*, 1987

[15] Kautz, “ The Logic of Persistence”, *Proceedings of AAAI-1986, pp.401-405*, 1986

[16] Kirsch, “Foundations of AI: the Big Issues”, *Artificial Intelligence 47, pp. 3-30*, 1991

[17] Kowalski & Sergot, “A Logic-Based Calculus of Events”, *New Generation Computing, vol 4, pp.67-95*, 1986

[18] McCarthy, “Circumscription – a form of non-monotonic reasoning”, *Artificial Intelligence 13, pp.86-116*, 1980

[19] McCarthy, “Applications of circumscription to formalizing common sense knowledge”, *Artificial Intelligence 26 pp.89-116*, 1986

[20] McCarthy and Hayes, “Some Philosophical Problems from the Standpoint of AI”, *Machine Intelligence vol. 4, pp. 463-502*, 1969

[21] McDermott, “We’ve been Framed: Or Why AI is Innocent of the Frame Problem”, *The Robot’s Dilemma, pp.113-122,* *Ablex Publishing*, 1987

[22] Morgenstern, “The problem with solutions to the frame problem”, *The robot’s dilemma revisited*, 1996

[23] Pearl, “Probabilistic reasoning in intelligent systems: networks of plausible inference”, *San Mateo, CA: Morgan Kaufmann*, 1988

[24] Reiter, “The frame problem in the situation calculus: a simple solution (sometimes) and a completeness result for goal regression”, *Artificial Intelligence and Mathematical Theory of Computation: Papers in Honor of John McCarthy*, 1991

[25] Rener, "Formen der Wissensrepräsentation" (providing an outline of Fodor’s “Modularity of Mind” Proseminararbeit zum Proseminar), *Institut für Psychologie der Universität Bern*, 1999

[26] Russell & Norvig, “Artificial Intelligence, a Modern Approach: Second Edition”, *Prentice Hall*, 2003

[27] Schubert, “Monotonic solution to the frame problem in the situation calculus: An efficient method for worlds with fully specified actions”, *Knowledge Representation and Defeasible Reasoning, pp. 23-67*, 1990

[28] Schwind, “Un Démonstrateur de Théorèmes pour des Logiques Modales et Temporelles en PROLOG”, *5ème congrès reconnaissance des formes et intelligence artificielle*, 1985

[29] Schwind, “Action theory and the frame problem”, *The frame problem in artificial intelligence – proceedings of the 1987 workshop pp.121-134*, 1987

[30] Searle, "Minds, Brains, and Programs", *Behavioral and Brain Sciences 3, pp. 417-424*, 1980[a]

[31] Shanahan, “Solving the Frame Problem: A Mathematical Investigation of the Common Sense Law of Inertia”, *MIT Press*, 1997

[32] Shoham, “Reasoning about change”, *MIT Press*, 1988

[33] Thielscher, “The Fluent Calculus”, *Dresden University of Technology*, 2001

Acknowledgements

We would like to thank Theo Janssen for his constructive feedback during the writing of this paper. We are certain that it would not have been finished in only 23 days if it had not been for his help.