The Frame Problem in Cognitive Modeling

Patrick Anselme

University of Liège, Psychology Department

In this target paper, I examine the factors that engender a frame problem in machines and why this issue is so hard to solve artificially. I begin by focusing on the frame problem within both deductivist and non-deductivist traditions of Artificial Intelligence (AI) in order to highlight its complexity and generality from symbol-friendly perspectives. Then the status of the frame problem in connectionist and dynamical systems is considered. It is shown that network architectures also have severe limitations in providing a real solution. Finally, I envisage the frame problem in living organisms. It is argued that organisms sometimes exhibit a frame problem when they encounter unusual situations insofar as they are no longer able to correctly anticipate the result of their actions in accord with their background knowledge. However, in general, organisms have no frame problem, either because they were designed to operate in very stable niches or because they can learn to appropriately deal with unpredictable changes. Affective states (including motivations, emotions, moods, etc.) are presented as an efficient natural means against the appearance of the frame problem in the animal kingdom.

Keywords: frame, representation, computation, action, logic, change, unpredictability, affect

Assuming a machine’s database, the frame problem consists in allowing this machine to avoid checking the representations that must not be affected after producing an action in a changing world. The machine should indeed ignore what is not supposed to be modified in its background knowledge if there is no indication to the contrary. For instance, imagine that a mobile robot’s world consists in two blocks A and B on the ground. Block A is red and block B is green. Then, for any reason, block A is put on block B by the robot. The question is: what about the color of the two blocks? Is block A still red and block B green after moving? Of course, this question seems totally

---

1 The author thanks Andy Clark for his helpful comments and many discussions on this topic.
2 Address for correspondence: Patrick Anselme, University of Liège, Psychology department (B 32), Quantitative psychology and cognitive science 5, Boulevard du Rectorat, B 4000 Liège, Belgium. Email: panselme@elgacom.net


Supplied by The British Library - "The world's knowledge"
absurd to human beings because we know that an object’s color does not change because it moves, except maybe in very special cases (e.g., block A has been painted very recently, so that some red paint runs onto block B and thereby changes its color). And a robot can be given a priori appropriate representations to deal with this simple situation. In reality, however, the “special cases” are somewhat frequent and become quickly unpredictable. They are the source of the frame problem in machines, as well as in organisms.

The problem of the two blocks above is obvious to us because we are able to evaluate the causal relations among many events in many situations. Indeed, we have learned a lot of things about the world since birth in order to protect us against the surprising character of most changes we encounter. On the other hand, machines do not have such an opportunity. Their background knowledge does not develop so much because machines learn only what we ask them to. If they do not usually suffer from the frame problem, it is because they are designed to operate in very stable situations, by means of very low dynamical interactions, more or less easily identifiable in advance by their designer. Machines are provided with the right set of constraints that allows them to correctly do their job. It is only when the mode of interaction of the machine with its environment prevents its designer from taking all precautions necessary that machines exhibit a frame problem. It is a bit as if one were in a situation that one does not know very well: it is then difficult for us to evaluate the effects and non-effects of our actions on our representations of the situation. Therefore, in special cases, we can also be paralyzed by the frame problem.

Most authors belonging to the deductivist tradition of AI consider the frame problem as a representational problem that should contain an acceptable technical solution in a logical formalism (Section 1). In contrast, the authors belonging to the non-deductivist tradition of AI envisage the frame problem in a more computational perspective, without giving up, in general, its representational dimension (Section 2). Here, the frame problem is not seen as a purely technical issue in logic, but rather as a psychologically relevant issue that has been highlighted by the methods of AI. After presenting both deductivist and non-deductivist (symbol-friendly) approaches, I will focus on the too rarely discussed status of the frame problem in connectionist and dynamical architectures (Section 3). Can they help its solution? For that, I think that several difficulties must first be solved concerning the ability of networks to store information and thus the ability to frame itself. It is shown that the frame problem has some psychological resonance for living organisms, including humans but also much more cognitively limited species like rats (Section 4). Finally, although this article does not provide any concrete solution to the frame problem, I highlight the determining role played by affectivity in the acquisition and activation of agent’s representations of the world. Without understanding these two key notions (acquisi-
tion and activation of representations), the solution to the frame problem would necessarily remain out of reach (Section 5).

1. The deductivist approach

I will begin by focusing on the deductivist approach, although coming after the non-deductivist one in the history of AI, because it is within that tradition that the frame problem initially appeared and was attempted to be solved. Here, the frame problem is viewed as a technical issue in logic. Its link with the mind of psychological agents is almost non-existent. Let us consider the case of monotonic logic and then a few words will be said about non-monotonic logics which have tried, without great success, to go beyond the limits of monotonicity.

1.1. Monotonic logic

According to McCarthy and Hayes (1969) intelligence is composed of an epistemological part and a heuristic part. The epistemological part concerns the representational model of the world in such a way that all subsequent representational states follow from the data expressed in that representational model; the heuristic part is the mechanism by which it becomes possible to logically infer a valid conclusion on the basis of available data. These authors assert, however, that several problems encountered in AI, including the frame problem, arise in constructing the epistemological part of intelligence. Thus, the frame problem would be caused by the fact of having to define a representational level in machines. The difficulty is indeed sizeable since human designers must determine in advance what axioms their machines need in order to reach a pre-specified goal in a changing environment. If we declare a set of axioms to be sufficient for reaching a particular goal in a given situation, it is perfectly possible that they prove rapidly insufficient because of a sudden change that modifies situation parameters, so preventing the goal from being reached. Other axioms then become necessary for countering that change. But the problem is that we do not know a priori what changes the machine will have to face, because the world is exhaustively unpredictable and thus exhaustively unknowable. Therefore, we do not know a priori what axioms we need to declare in order that the machine does its task. Furthermore, the rise in the number of machine’s axioms will raise the number of axioms that are necessarily irrelevant in relation to its goal, so that the machine must not squander its computational resources by checking all its representations to find the right ones.

Deductivists generally consider the frame problem as a problem of representations, not computations. Of course, they agree that certain computational difficulties, such as data updating, are related to the frame problem, but they assert that these difficulties are not part of the issue: the frame problem "is concerned with what axioms to input to the machine, not with what
the machine should do with them once it has them"; so the framing "is not a problem for the robot, but for us, its designers" (Hayes, 1987). For Hayes, the frame problem only stems from the requirement to describe a world in which states are in constant movement due to actions and events, that is, a world in which the representations are potentially alterable from one moment to the next. It is not, contrary to a widespread notion, the problem of reasoning in change as a whole, or what Hayes ironically calls the "General Problem of Artificial Intelligence" (GPAI), but only its representational part.

The so-called epistemological part of intelligence was initially modeled by McCarthy through the concept of "situation calculus." A situation is a set of facts describing the state of affairs at some moment in time as completely as possible. It is not anything static: a situation can change because of an agent's actions or events, so that its complete description at time $t$ becomes another complete description at time $t+1$. The role of an agent's actions remains, however, limited to the description of environmental changes. Indeed, these environmental changes are specified in terms of a transition from one state to another, rather than in terms of actions modifying certain attributes and not others. Therefore, every attribute of which the status has not been defined remains non-specified. McCarthy's situation calculus then proves to be very constraining since it forces us to declare every value of the attributes in every possible state of the environment at any moment in time. (There is a one-to-one correspondence between a given moment in time and an environmental state insofar as time is linear and composed of discreet moments.) Once everything has been declared in the machine's memory at time $t$, each change that occurs at times $t+1$, $t+2$, etc. forces us to declare the set of things making the situation again, while maintaining the previous descriptions in memory. This is a property of monotonic logic: a fact is added to the database as soon as it is proved to be true, so that the database grows incessantly.

In situation calculus, things that must change when an action takes place are specified by action axioms, whereas things that cannot change after an action are specified by frame axioms. The existence of these frame axioms is at the origin of the frame problem in machines (Dennett, 1987). Indeed, they are often difficult to formulate (Ginsberg, 1987), in addition to using up most of the system's time processing (Haselager, 1995). Their specification depends on the number of agent's actions and on the number of fluents, which are supposed to give partial information about situations by indicating which facts are true in what situations. "A number of fluents are declared as attached to the frame and the effect of an action is described by telling which fluents are changed, all others being presumed unchanged" (McCarthy & Hayes, 1969). But, when an action is performed in the world, driven by action axioms, we are never sure that what is supposed to remain unchanged must be systematically ignored by the machine because of the dynamic nature of the world. For example, if a robot's representation of a lamp on the
table had to remain unchanged after its action and, in the meantime, someone has moved the lamp, this makes the frame axiom ON(lamp, table)[s] wrong. Robot designers should then declare another frame axiom according to which ON(lamp, table)[s] is true, unless someone moves the lamp; unless it falls on the floor because of the cat; unless... Thus declaration of fluents is endless.

Not only is the number of frame axioms to be initially considered gigantic, but also many exceptions exist for each frame axiom already declared. We have no guarantee that what is specified will have importance in an agent's exchanges with the world. Conversely, we have no guarantee that what is not specified will have no importance either. The principles of McCarthy's situation calculus are simple and then seem natural, yet the constraints it involves are the main technical cause of the frame problem.

1.2. Non-monotonic logics

Other theoretical alternatives were then suggested in order to make up for the limits of monotonic logic, which are grouped under the label of non-monotonic logics (e.g., McCarthy, 1980; Reiter, 1980; Moore, 1985; Shoham, 1986; Lifschitz, 1987). The idea behind all these formalisms is that a fact persists in the database in the absence of information to the contrary (Ginsberg & Smith, 1988). Non-monotonic logics thereby allow the avoidance of redundancies stemming from the use of frame axioms (Baker, 1991) as well as inconsistencies in the axioms (Hanks & McDermott, 1986).

For instance, if someone believes that "all birds fly" and then he or she learns that "certain birds, like penguins and ostriches, do not fly," his or her initial belief must change to "a majority of birds fly" rather than being maintained in the database near this second belief. Default reasoning and circumscription are the two main non-monotonic methods of AI. Default reasoning (Reiter, 1980) consists of a set of sentences in first-order logic that are coupled with a set of default rules, so that these rules are used when information is lacking. Circumscription (McCarthy, 1980) is a technique allowing a person or program to jump to certain conclusions: if a system can show, from a set of facts, that objects sharing a certain property exist, then the system circumscribes the objects in question as satisfying that property. For example, we are able to deduce from a series of things we know about navigation that a boat is usable, unless something prevents it (e.g., its hull is broken). Here again, unfortunately, we have to face the "unless" operator and its endless list of exceptions as in the case of monotonic logic: the correctness of a conclusion "depends on our having 'taken into account' all relevant facts when we made the circumscription" (McCarthy, 1980).

Despite the important work on non-monotonic logics and McCarthy's (1986) enthusiasm for its suitableness for explaining the mind, many authors remain somewhat skeptical concerning its psychological relevance. First of all, non-monotonic formalisms are incomplete and doubts exist as to
whether they can be achieved (Davis, 1980; McDermott, 1982). Furthermore, their contribution to understanding mind is, often enough, unrealistic. Baker states thereby that the general principle that a fact persists in the absence of contrary information does not work because no basic model corresponding to our intuition exists. "In a sequence of events, often one can eliminate an expected abnormality at one time by introducing a totally gratuitous at another time" (Baker, 1991). Finally, non-monotonic logics are essentially viewed as an efficient formalism for common sense. However, if common sense is widely non-monotonic, this does not mean that non-monotonic logics capture all its important dimensions. Common sense seems to a large extent to be implicit (unconscious/procedural) rather than explicit (conscious/declarative), but such a distinction is basically lacking in all logical formalisms.

1.3. The sleeping dog strategy

Admittedly, the real world is locally changing and uncertain because of consequences of certain actions or events, but it is globally inert, so that most representations an agent has should not be affected in a particular situation. As already noticed, the agent should not check these unaffected representations.

According to McDermott (1987), the frame problem is the inertia problem and could find an acceptable "solution" with what Haugeland calls the "sleeping dog" strategy: the system will totally ignore facts, without even glancing at them, if there is no indication that they have been affected (Haugeland, 1987). In sum, McDermott defends the idea that there is no possible or necessary solution to the frame problem because "no working AI program has ever been bothered at all by the frame problem" (McDermott, 1987, italics). This idea is akin to Hayes's (1987) which explains that many AI systems do not think about change, and when they do, it is always in very predictable ways, so that the frame problem does not arise in their reasoning. In addition, McDermott argues that since we humans are unable to predict all consequences of our actions, when they are numerous or unusual, it would be absurd to ask a machine to do it: the frame problem would then be just an unrealistic symptom of the fact of accepting a level of complexity that we are unable to manage. Hence, the sleeping dog strategy proves efficient enough to face the majority of situations a machine may encounter.

Fodor (1987) suggested an argument according to which the sleeping dog strategy is not a solution but rather a relocation of the frame problem. Indeed, the sleeping dog strategy is based on the idea that most facts do not change. However, we have no certitude that facts assumed to be fixed cannot sometimes change. That depends, Fodor says, on how facts have been isolated. Fodor gives the following example: imagine that every physical particle of the universe becomes a "fridgeon" when Fodor's fridge is turned on. Hence, when Fodor turns his fridge on, this changes the state of every
physical particle of the universe (they become a "fridgeon"), such that a
great number of new facts are created by Fodor's action. "The sleeping dog
strategy proposes to keep the computational load down by considering as
candidates for updating only representations of such facts as an event
changes. But now there are billions of facts that change when I plug in the
fridge; one fact for each particle, more or less" (Fodor, 1987). Thus the sleeping
dog strategy is not a right or wrong solution because it is not a solution
at all. As long as we do not determine what counts as a fact, the sleeping dog
strategy is an empty technique. In sum, McDermott (1987) is right only due
to an artifact: "the programs run because the counterexamples are not al-
lowed to arise. The programmer decides what kinds of properties get speci-
fied in the database, but the decision is unsystematic and unprincipled"
(Fodor, 1987).

There is another reason for rejecting the sleeping dog strategy: it has an a
posteriori character for agents that deal with a changing environment. The
strategy implies that agents check a number of their representations, only
those suspected of having changed, but necessarily after acting. On the con-
trary, with living organisms, and particularly humans, anticipation plays a
crucial role in driving action. It is often before acting that agents modify their
representations because they expect, even implicitly, that their action will
have consequences that will affect some of their representations. Planning
and decision-making are rational strategies avoiding having to rush head-
long and then observing perhaps dramatic consequences. The sleeping dog
strategy as an a posteriori tool would not only be fatal for organisms in many
cases, but also means, oddly, that organisms understand nothing at all about
the environment in which they operate. Such a non-understanding is unfor-
unately a fact for machines, but techniques like the sleeping dog strategy, in
which the representational reassessment comes systematically after action,
cannot help them to improve their mastery of the situations they encounter.

Failure of the sleeping dog strategy shows us that the only means of
really defining the representations that should not be affected in an agent's
mind after an action involves determining what will change in this agent's
mind and in the world as result of this action. The set of unaffected repre-
sentations depends on context, just like the set of affected representations,
since when a piece of knowledge is no longer unaffected it becomes auto-
matically affected, and conversely. A reformulation of the frame problem is
then required: it is not the inertia problem, as believed by deductivists, but
rather the problem of managing the dynamics between affected and unaffected rep-
resentations when a change occurs after an agent's action. I will come back to this
definition in Sections 4 and 5 in order to show, among other things, that the
frame problem also arises in the animal kingdom. At the moment let us only
consider its direct implication for cognitive modeling; the "frame" is not the
set of unaffected representations because these do not bring about any prob-
lem in themselves. The "frame" is rather the set of representations affected
by the consequences of actions and serving as a guide for sensory-motor interaction of agents with their worlds.

2. The non-deductivist approach

First, non-deductivists consider the frame problem as a psychologically relevant issue, not just a technical/logical one (Janler, 1987). According to Dennett (1987), for example, the frame problem is not just an engineering problem in AI, it is a profound epistemological problem brought to light by the recent methods of AI. (Non-monotonic logics also present the frame problem in psychological terms but, as monotonic logic, they focus only on a technological solution.) Second, non-deductivists consider, often enough, the frame problem in terms of Hayes's GPAL: how can a machine sensibly reason about a changing world? Maybe this attitude is erroneous because, as Hayes notes, the framing is just a part of the GPAL. The updating problem, the ramification problem, the qualification problem, etc. constitute its other parts. However, the more general character of the non-deductivist approach invites us to look more easily at both representational and computational dimensions of the frame problem. It is no longer merely the technical issue of the right representations to declare in order to face environmental changes, but rather the psychological issue that consists of facing environmental changes both representationally and computationally. This does not mean that the frame problem is equivalent or even similar to the GPAL, but that the frame problem, as a part of the GPAL, has representational and computational implications.

There are non-deductivists who see the frame problem in strictly computational terms. So Haugeland asserts that the frame problem is similar to the knowledge access problem, so that "the challenge is not how to decide for each fact whether it matters, but rather how to avoid that decision for almost every bit of knowledge" (Haugeland, 1987). According to Fodor, the frame problem is "the problem of putting a 'frame' around the set of beliefs that may need to be revised in the light of specified newly available information" (Fodor, 1983, p. 112-113). In his view, "this isn't just "a bookkeeping problem; it is the general problem of inductive confirmation" (Fodor, 1983, p. 115). Fetzer also insists on the role of induction: "the frame problem cannot be solved without a solution to the problem of change, which is the dimension of the frame problem that is a 'special case' of the problem of induction" (Fetzer, 1993; see also Fetzer, 1990).

Bayes' theorem is often presented as an interesting computational solution to the frame problem in that it allows us to modify the probability of an event occurring from evidence. Bayes' theorem specifies how to combine the prior probability of a hypothesis with the probability of an evidence in order to determine the posterior probability of the hypothesis.
According to Korb (1998), "causal learning is necessary and sufficient for solving the frame problem" and such causal learning can be carried out by Bayesian networks. These networks are indeed able to reason under uncertainty and take into account the causal characteristics of the world as closely as possible within isomorphic micro-models. In addition, they learn and adapt autonomously to change by dealing with data that designers have not pre-codified. In sum, "their behavior mimics that of isomorphic micro-models, to the extent that their deviations from isomorphism are small or unimportant to the variables of practical interest" (Korb, 1998). Bayesian networks can then "discover more exact relationships between the variables" and "make more precise predictions." Korb is certainly right to say that the frame problem is strongly, even essentially, related to the question of causality. The ability to detect and learn causal relations in the world ensures the system's architectural constraints to best fit those of its environment. On the other hand, he is probably wrong to consider that the most relevant variables (those of "practical interest") should be input to the system in advance. As already put, in complex situations designers do not know what these variables are. Korb seems to pay little attention to the role of latent variables. Yet what determines the relevance of one particular variable in a given situation is often linked to context, that is, a set of variables, sometimes hidden, that constrains perception of the world. If the context changes, an initially relevant variable can suddenly become irrelevant, and vice versa because of the subtle, underlying determinism of latent variables. This criticism has nothing to do with the fact of not accepting imperfections in the model, contrary to what Korb suggests, because latent variables are numerous and can really prevent Bayesian networks from achieving their goals most of the time.

3. What about network architectures?

3.1. Connectionist systems

Connectionist architectures account for a wide range of psychological phenomena, such as form recognition, generalization and lexical priming, via constraints closer to biological structures than those of symbolic devices (e.g., Grossberg, 1980; Rumelhart & McClelland, 1986; Smolensky, 1988). Classically, neuron-like nodes are arranged in layers, which comprise an input layer (receiving data), a hidden layer (where a representation emerges as an activation vector) and an output layer (which is generally compared with a target pattern to reach). Each layer is usually fully connected to the following via its neuron-like nodes. As information spreads through the network, the connection weights modify themselves, thus allowing learning to take place gradually, trial after trial. Learning will be either supervised
(reaching a predefined target pattern) or unsupervised (self-organizing) according to the kind of algorithms the network uses.

According to Churchland (1989), connectionism offers an interesting approach to informational encoding and retrieval allowing us to bring a solution to the frame problem viewed as a search problem. Contrary to symbolic methods, connectionism does not need serial and local search strategies, which are computationally costly for machines. A mechanism of spreading activation across a large number of nodes modifies the connection weights in the network in accord with their excitatory or inhibitory character, driving the input signal to its destination. “Since information is stored not in a long list that must somehow be searched, but rather in the myriad connection weights that configure the network, relevant aspects of the creature’s total information are automatically accessed by the coded stimuli themselves” (Churchland, 1989, p. 178).

Churchland is right to say that connectionism offers an alternative to the symbolic tradition concerning the ability of machines to quickly access their knowledge due to the mechanism of spreading activation. However, it is unlikely that this can solve the frame problem. Not only is Churchland’s view an oversimplification of what really happens, but connectionism also suffers from several limitations that prevent the network architectures from framing anything. These limitations are:

**Catastrophic forgetting.** In general, connectionist systems cannot develop representations corresponding to very different input-output pairs in a same hidden layer, so that the output has no occasion to be framed by multiple representations. A connectionist network is capable of learning several highly correlated representations in order to construct a category. For example, some sort of conceptual category for birds can be formed by learning different exemplars of birds (sparrow, pelican, crow, etc.). Any type of similarity among a series of representational properties in the network brings about their associations by overlapping, according to the dimensions considered by the designer (here: feathers, beak, ability to fly, etc.). However, categorical learning remains insufficient to solve the frame problem: framing an action is only possible provided that several categories can be maintained in a same network. Unfortunately, this operation leads to catastrophic forgetting of representations by interference in connectionist systems (Ratcliff, 1990; McCloskey & Cohen, 1989). When some information is learned, it completely and suddenly destroys the information previously learned if it is too different. A series of techniques attempt to attenuate the risk of catastrophic forgetting, but this is often done to the detriment of the network’s ability to generalize (e.g., Kortge, 1990; Levandowsky & Goebel, 1991; Murre, 1992; McRae & Hetherington, 1993; for more realistic models, see French, 1992; McClelland, McNaughton & O’Reilly, 1995).

**Interactions among representations.** Providing a solution to catastrophic forgetting is the way to go, but that is not enough insofar as the range of
interactions between representations in connectionist networks is very limited. For example, the loss of information seems unavoidable when several distributed representations are combined in a more complex distributed representation (Haselager & Van Rappart, 1998). Another worry concerns the fact that all representations cannot interact by overlapping. In the brain, one representation can activate another without sharing any common traits. There are computer techniques for synchronizing neuronal activation, so that radically different representations can be used at the same time, but these techniques involve transforming the simple associationist idea that usually underlies the connectionist paradigm (e.g., Singer, 1995; Roelfsema Engel, König & Singer, 1996). Finally, flexibility of connectionist representations is maybe not as great as it seems a priori. According to Clark (1991), distributed representations exhibit some flexibility owing to their emergent properties and the fact that they constitute example-driven knowledge. They can be consequently altered and reorganized up to a certain point when the input space remains unchanged. But these representations make the system inflexible and thereby non-adaptable when the input space changes over time. Then the network needs to be re-trained to face the change that happened (see also Clark & Karmiloff-Smith, 1993).

Regulation of the informational flow. Connectionism seems to offer a computational solution to the frame problem because, as Churchland claims, neural networks have supple trajectories of activation allowing automatic and quick access to the relevant information. However, a connectionist system is like a “reflex”: once the right input is there, the corresponding output, with which an association was previously learned, follows immediately by propagating activation across every node of the network. In contrast, Clark (in preparation) notices that when information comes into the brain, it is rapidly routed and filtered to its destination without encountering most other cognitive resources. The flow of information is modulated into the brain, but this is hardly the case in connectionist devices. All nodes are not, of course, equally stimulated in an artificial network, but they are affected anyway. What then makes such a modulation possible in the brain? Van Essen, Anderson & Olshausen (1994) suggested that among the large populations of neurons serving as encoders of knowledge, certain neurons can be specialized in trafficking the flow of information within cortical areas. In their model, the “shifter circuits” allowing the relevant knowledge to be routed to its right destination are presumed to be innate. However, as Clark (in preparation) points out, such a presupposition is plausible in the limits of their simple model, which focuses only on some aspects of visual attention, recognition and motor control, but it seems much less credible for higher-level cognition.
3.2. Dynamical systems

Purely dynamical systems, of which connectionist devices are a subcategory (van Gelder, 1995), designate a set of physically describable state-dependent complex systems. They are capable of self-organizing different endogenous and exogenous variables over time (e.g., Skarda & Freeman, 1987; Thelen & Smith, 1994; Giunti, 1997). Dynamicists think that a system's behavior is not planned by the system before acting, but rather takes place while the system's action is specifying the trajectory of its achievement. These systems must gradually shape and adapt their sensory-motor coupling with the world on the basis of their experiences in order to improve their behavior in a particular environment (e.g., following along a wall, pursuing a moving target, avoiding an obstacle, etc.). Behavior is, in general, very simple and can be only exhibited when the source of stimulation is present. Thus, if a robot is pursuing some moving target visually and this target is temporarily hidden because going behind an obstacle, the robot will directly turn itself off owing to an interruption in the dynamical process linking the robot with its environment. Contrary to symbolic and connectionist systems, which are representational systems, any information here cannot be kept in memory out of immediateness.

No need to say more. Dynamical systems are incapable of solving the frame problem since they store no knowledge and therefore develop no cognitive frame for action. The frame problem is in no way eliminated, but merely evaded because dynamical systems are too elementary to be bothered by the frame problem (see French & Anselme, 1999). Their physical states, determined by numerical variables, change as a result of environmental changes: But (a) these physical states are not causally efficient outside of the system's immediate stimulation, and (b) the kind of environmental changes capable of inducing a frame problem in machines is absolutely not that with which dynamical systems are able to manage. It has been shown that some preliminary knowledge of the environment in which an agent operates must be available and accessible to this agent in order to interpret its changes. And the frame problem is to know how that is possible. However, dynamical systems are only concerned by changes whose tractability is dependent on the immediate status of the variables involved. Their task is to adjust their sensors and effectors to the immediate scene in an a-historical relationship, not to deal with changes from preliminary background knowledge of the scene in question.

4. The frame problem: From robots to animals

Contrary to a widespread notion, the presence of changes is not basically the source of the frame problem. Provided that the designer knows the changes to be dealt with in advance, specific programs can be constructed to allow a
system to face these changes, however complex they may be. For example, expert systems are not usually bothered by the frame problem although they work in broad, complex environments modifying themselves over time (e.g., think of video games). Actually, expert systems operate in closed worlds, in which everything is totally constrained by the job they have to do. Both start and end states of their processing work are predefined. Thanks to that, every option of change is carefully considered in advance by the designer.

In Section 1.3, it has been suggested that the frame problem, for an agent, consists in managing the dynamics between its affected and unaffected representations when a change happens in the environment. But change alone is not its cause, as seen. What causes the difficulty for an agent to manage its representations is the conflict between some environmental changes and the agent’s cognitive dispositions. These dispositions can be a designer’s hypothesis in the case of mobile robots, but also, after all, the set of phylogenetic and ontogenetic constraints of living organisms. Animals, including humans, should suffer from the frame problem insofar as they live in the real world, which is changing and uncertain, and are constrained by their history, defined in a broad sense, which does not provide for them against all types of change. Historical constraints are only there to equip animals for dealing with the changes with which they are (as a species or as individuals) usually confronted. Thus a frame problem is to be expected as soon as an agent, robot or animal, has to act in (or just represent) an unusual situation.

Let us consider examples of the inability of humans and of non-humans to correctly manage their representations because of a conflict between a change in the environment and their (history-based) cognitive dispositions. So when politicians are invited to a country whose cultural traditions are very different from those of their own country, they risk a frame problem: ignoring the native rituals and acting inappropriately (without being conscious of it) can have unexpected consequences for politicians and even lead to diplomatic incidents. Similarly, when you are introduced to someone you have never seen before, it is rather unpolished to accost the person straightaway about controversial ideas. This way of interacting does not work because the person’s reaction would be totally unpredictable in the absence of further information. You would have a frame problem.

The frame problem, as defined above, also exists in non-humans when the evaluation of the consequences of their conditioned actions is prevented. In my view, the blocking effect, well known by animal psychologists, is an example of its occurrence in animals. The blocking effect arises when an animal learns a causal association between two stimuli A and B (e.g., if light, then shock), and then is asked to learn another causal association between a new stimulus C and stimulus B (e.g., if tone, then shock). Pre-exposure to an association A-B decreases (blocks) the learning capacities of animals for another association of the type C-B (Kamin, 1969; Mackintosh, 1973). This is a frame problem because, after learning A-B as a historical constraint defining
some cognitive disposition, animals are unable to face the change from A to C. Optimally, in an environment where C-B is relevant, the association A-B should be inhibited to allow C-B to take place. The learning of A-B as an ontogenetic constraint here is not, properly speaking, the source of the frame problem. The frame problem is rather caused by the phylogenetic limitation to inhibit A-B in order to learn C-B.

Let us take another example. A mouse can escape a cat by hiding in a place it has previously located. The presence of the cat will automatically activate the right thing to do by the mouse, rather than bringing about a review of all possible actions with their probably irrelevant consequences in the situation. However, a frame problem is not excluded in that the mouse might fail to succeed in managing its representations owing to the suddenness of the event. If the appearance of the cat were totally unexpected, the mouse could try, in its panic, to escape it by running in a direction while believing wrongly that there is a hideout close by. Here, the mouse has a frame problem since some irrelevant knowledge was automatically accessed and processed.

There is no objection to consider that the frame problem exists in many animal species insofar as it is insensitive to the explicit (conscious) or implicit (unconscious) character of the representations of the world involved. The frame problem is also insensitive to the format (symbolic or distributed) in which these representations are encoded. If the frame problem is more easily highlighted in humans than in non-human species, that is only due to the fact that humans manipulate a greater number of representations in order to face highly complex situations.

5. Why is the frame problem so occasional in the animal kingdom?

Not only robots but also animals can succumb to the frame problem. However, it is a matter of fact that animals have much fewer difficulties than robots facing hostile environmental conditions. In general, they have no frame problem. Why? What do animals have that robots do not? I do not claim to bring the solution to the frame problem in organisms but, at least, suggest an explanation of the difference observed between robots and animals with respect to the frequency of its occurrence.

The frame problem is about the computational management of the representations of the world. It is both a representational problem (what to represent) and a computational problem (how to compute). In addition, the frame problem is a temporal problem (when to compute), which is strongly associated with the computational one insofar as information in the brain has to be processed as quickly as possible in order to be retrieved at the right time. Any solution to the frame problem should necessarily take into account these three parts of the global issue.
Anselme: The Frame Problem in Cognitive Modeling

I argue that the affects, from simple needs to moods and feelings, can account for these three parts: Biological motivations (hunger, thirst, sexuality, etc.), hedonic states (pain and pleasure), emotions (fear, anger, guilt, etc.), moods (sadness, happiness, etc.), and feelings (suffering, well-being, etc.) affect the organism over its interactions with the world. Affects not only affect the body, but also the brain and its processes. For instance, many studies show that the emotions are involved in perception, memory, attention, learning, decision making, among other psychological faculties (e.g., Johnson-Laird & Oatley, 1992; Damasio, 1994). Moods distinctly influence evaluative judgment (Clere, 1992), creativity (Jamison, 1995), and memory retrieval (Laird et al., 1982; Lazarus, 1991). Therefore, could not affects as a whole be significant determinants of the solution to the frame problem in organisms?

I am not interested in the ability of organisms to recognize or express affective states. After all, an agent can recognize or express attitudes that seem to indicate the presence of affects in this agent when its behavior merely results from a set of rules (Picard, 1997). For instance, the robot Kismet (e.g., Breazeal & Scassellati, 1999) takes into account the physiological and behavioral consequences of several emotions, like intonation, in order to recognize an emotion expressed by a human and then react in the same way (e.g., anger causes Kismet's anger). But the deep causes of Kismet's emotions are ignored: there is no integration of cognitive factors allowing the robot to really recognize and express emotions in empathy with the person with which it interacts. The robot has no emotions properly speaking, and therefore cannot develop any emotional intelligence. The fact of having real emotions as well as an emotional intelligence, through which the emotions are both regulated and used, characterizes animals as opposed to robots. I do not want to relate the abundant literature on this question, but rather to envisage how the presence of affects in animals has a positive impact on the solution to their frame problems, viewed in terms of representational problems, on the one hand, and of computational/temporal problems, on the other hand.

1/ Affects, frames, and representation. What to input to the frame (in a given situation)? Insofar as the designers cannot specify everything in advance, any solution to the frame problem will be only acceptable if “the system itself is able in a reasonable way to expand and modify its knowledge structure” (Haselager, 1995, p. 94). So the real question is: what must the machine learn to input to the frame (in a given situation)? I suggest that the presence of affects in animals, especially motivations and emotions, helps them to learn the right things and then reduces the representational part of the frame problem. Let us see why.

All affects have specific conditions of satisfaction. For example, hunger is satisfied by eating food, not glass or plastic; happiness is satisfied by experiencing pleasant events; fear can be satisfied by avoiding something danger-
ous; etc. Broadly, the satisfaction conditions of positive affects, such as happiness, is what allows them to be generated; and the satisfaction conditions of negative affects, such as pain, is what allows them to be suppressed.

I argue that the satisfaction conditions of affects differ from the affective contents properly speaking, and that these satisfaction conditions are not, in general, known by agents a priori (Anselme, submitted). They must be learned to some extent according to circumstances. For example, when a thirsty dog has to search for something to drink, what is able to satisfy the dog’s motivation (i.e., water or maybe milk) is not what defines the motivational content: the dog just has a dry mouth. The dog is then in a position to learn thirst-quenching stimuli. Similarly, there is evidence that organisms can be conditioned to exhibit fear about anything. Fear can have a multitude of objects, which are a priori non-specified to agents. Animals are then forced to learn in an open-ended way the stimuli that satisfy their affects in the situations they are used to dealing with. Insofar as animals are essentially interested in managing their interactions with the world, their affective states are determining; they potentially allow agents to learn what to input to the frame in a given situation.

2/ Affects, frames, and processing time. How to access the right frame (in a given situation)? And how much time is needed? The affective properties mentioned above remain silent about how agents can access the right bits of knowledge once they have them. In nature, it is one thing to learn a series of relevant data in an open-ended way, it is another to access them at the right time in order to stay alive. However, affects also seem to play an important role in driving information. In the brain, the emotions are coded by the limbic system, an archaic structure that existed in paleo-mammals, like horses, and around which the neo-cortex developed in more recent mammals, like primates (MacLean, 1970). In addition, “it is known that there are many more connections carrying signals from the limbic system to the cortex than vice-versa” (Picard, 1997, p. 41). These findings seem to show that the emotions are neuro-biologically involved in the activation of high-level cognitive processes.

Damasio (1994) observed a positive correlation between the capacity of frontal patients to repress emotions and their capacity for making decisions, in spite of a perfect conservation of fundamental processes such as attention, memory, language and general intelligence. According to Damasio, the reasoning processes cannot be wholly explained by the fact of computing some information logically: logic can only help us to make a final decision after other processes have dramatically reduced the number of possibilities. As such, logical evaluation is too slow, and even impossible, when many options must be simultaneously considered. For example, imagine the number of factors you would have to take into account in order to choose between two equally attractive jobs or the person whom you want to marry. Damasio suggests that this inability of logic to account for decision-making is due to
the limited capacity of working memory: all cost/benefit ratios to be considered cannot be simultaneously kept in memory, even when the problems are simple, in order to establish systematic comparisons.

How can affects drive the retrieval of appropriate representations? Let us take a concrete case: fear. A human being is frightened hundreds of times in relation to very different events and learns to react to them. How can the simple emotion of fear select the right representations, in order to frame appropriate actions, when a great number of representations are associated with it? Does this not come back to the frame problem? Definitely not... if we admit that one fear is not another. For instance, the following expressions designate fear: to be afraid of heights, of drowning, of an aggressive dog, of someone’s opinion, of a plane breaking the sound barrier. In these examples, fear is the common denominator, although its perception is different in each case. Each type of fear is attached to a very specific cognitive content. It is even probable that context refines more the emotional perception and the related cognitive contents. So being afraid of someone’s opinion will be perceived differently depending on its consequences, which depend, in turn, on vast situational background.

It is therefore reasonable to imagine that each situation of life corresponds to a specific affective pattern, eventually composed of a combination of several affects (motivations, emotions, feelings, etc.). Any change of situation can lead to a frame problem when an agent acts as if there were no change, so that an affect could be generated by this agent as a result of its action’s effect in order to avoid or reproduce this action thereafter. This can be summarized by the slogan: “an effect, an affect.” The unexpected consequences of an action (or event) bring about an affect that decreases the probability of confronting the unpredictable character of what happened. For instance, if an uncontrolled movement makes me to knock over my friend’s cup of coffee on her dress, I will be surprised (initially) and embarrassed (subsequently). These emotional patterns will help to increase my vigilance in order to satisfy the emotions engendered (i.e., avoiding doing such a movement thereafter). Similarly, the blocking effect, preventing an animal from learning correlations between events, can be suppressed when the animal is surprised by a new association (Kamin, 1969). Recall that pre-exposure to an association A-B (e.g., light causes shock) during several trials decreases (blocks) the learning capacities of animals for another association of the type C-B (e.g., sound causes shock). However, without going into detail, Rescorla (1971) showed that when a rat is conditioned to learn an association light → shock, as well as an association sound → no shock, it then learns to associate the sound to the shock after being exposed to a coupling light-sound as a cause of the shock. This is due to the surprising link of the sound to the shock, initially uncorrelated.

In summary, affects help the solution to the frame problem because they are involved in the acquisition and activation of appropriate frames for ac-
tion. They reduce an agent's risk of being paralyzed by a representational problem (what to represent) and a computational/temporal problem (how/when to compute), which constitute the main parts of the frame problem. Thus affects are a condition necessary, if not sufficient, for its solution. Not only do they allow the acquisition of any kind of representations, insofar as these representations have some usefulness in relation to their satisfaction, but they also seem to have specific computational trajectories allowing these representations to be retrieved at the right time. In a sense, an affect is like a module, that is, an entity that processes information very quickly without interfering with other modular entities (Fodor, 1983). I do not have enough space to discuss the link between modularity and the frame problem, although Fodor has already stressed the role of modular structures in orienting computations in wide databases (Fodor, 1987, 2000). Affects are not modules in Fodor's sense, but they share several of their properties, such as quickness and some sort of encapsulation avoiding information being spread around, that would explain their role in the orchestration of frames for action. So a frame problem can arise in a new situation when an agent believes having recognized a familiar pattern, which arouses an affect successfully associated with that pattern in the past, and then acts in expectation of consequences that do not happen.

An outstanding question is that of determining how the situation, whose perception is supposed to activate some corresponding affect, is initially recognized. But this issue will not be considered here.

The opportunity of many animal species to develop cognitively under parental protection during several weeks, months or years, according to the species, is an effective means of learning a lot of effect-affect associations without risking a fatal frame problem. Once equipped with a good package of such associations, organisms can quit their parents and confront the real world. This does not concern the tick, an insect that pumps mammal's blood through the skin, because it is only sensitive to very specific stimuli, such as the presence of butyric acid and a temperature of 37°C (von Uexküll, 1965). The other stimuli are ignored, not because the tick judges them uninteresting, but because it does not detect them at all. Since the tick does not have to deal with these stimuli, it has no frame problem, and therefore need not develop cognitively. It is operational at birth. In contrast with more evolved species, the more complex the environment to be dealt with, the longer the period of parental protection from birth will be. So a clear difference can be observed between social and non-social species. But a difference is also observable in closer species, for instance within primates, according to the complexity of social constraints they have to manage. On average, lemurs reach the adult age after 2.5 years, monkeys after 7, apes after 11, and humans after 20. The frame problem is so fundamental that it could explain why all organisms have affective states and why these affective states become richer as agent-world interactions become more complex.
Anselme: The Frame Problem in Cognitive Modeling

How affectivity can be concretely and plausibly modeled in computers in such a way that it helps the management of representations and their learning is now another story. Admittedly, the affects are not just communication tools between individuals, but are also at the root of developing cognition. However, this idea remains largely unexplored within the AI community, although it is becoming more and more empirically supported. Modeling the affects as a starting point for cognition is currently an open, hard question but it might be the way to go to clarify the discussion on the frame problem.

6. Conclusion

(1) What is the frame problem? The difficulty of managing the dynamics between affected and unaffected representations when a change happens in the environment. (2) What is its cause? The conflict between certain environmental changes and the agent's cognitive dispositions. (3) When does it arise? The frame problem arises when the representation an agent has about the impact of an action (or event) differs from what really happens because of situational changes. The main paradigms of cognitive science have been considered as to their ability to face the frame problem. In my view, there is no hope of seeing machines going beyond this issue in the near future: the frame problem is never solved, just evaded, either due to appropriate artifacts (symbolic systems), or due to intrinsic limits of information processing (connectionist systems and dynamical systems). Then it has been shown that, sometimes, organisms suffer from it as well. However, they seem to avoid its occurrence most of the time. It was argued that the affective states of organisms play a central role in reducing its main representational, computational, and temporal implications, even though how this happens remains to a large extent somewhat mysterious.

References


Clark, A. (in preparation). Global reason and authoritative sources, or, how search engines can save cognitive science.


