



The Frame Problem

First published Mon Feb 23, 2004; substantive revision Mon Feb 8, 2016

To most AI researchers, the frame problem is the challenge of representing the effects of action in logic without having to represent explicitly a large number of intuitively obvious non-effects. But to many philosophers, the AI researchers' frame problem is suggestive of wider epistemological issues. Is it possible, in principle, to limit the scope of the reasoning required to derive the consequences of an action? And, more generally, how do we account for our apparent ability to make decisions on the basis only of what is relevant to an ongoing situation without having explicitly to consider all that is not relevant?

- [1. Introduction](#)
- [2. The Frame Problem in Logic](#)
- [3. The Epistemological Frame Problem](#)
- [4. The Metaphysics of Common Sense Inertia](#)
- [5. The Frame Problem Today](#)
- [Bibliography](#)
- [Academic Tools](#)
- [Other Internet Resources](#)
- [Related Entries](#)

1. Introduction

The frame problem originated as a narrowly defined technical problem in [logic-based artificial intelligence](#) (AI). But it was taken up in an embellished and modified form by philosophers of mind, and given a wider interpretation. The tension between its origin in the laboratories of AI researchers and its treatment at the hands of philosophers engendered an interesting and sometimes heated debate in the 1980s and 1990s. But since the narrow, technical problem is largely solved, recent discussion has tended to focus less on matters of interpretation and more on the implications of the wider frame problem for [cognitive science](#). To gain an understanding of the issues, this article will begin with a look at the frame problem in its technical guise. Some of the ways in which philosophers have re-interpreted the problem will then be examined. The article will conclude with an assessment of the significance of the frame problem today.

2. The Frame Problem in Logic

Put succinctly, the frame problem in its narrow, technical form is this (McCarthy & Hayes 1969). Using mathematical logic, how is it possible to write formulae that describe the effects of actions without having to write a large number of accompanying formulae that describe the mundane, obvious non-effects of those actions? Let's take a look at an example. The difficulty can be illustrated without the full apparatus of formal logic, but it should be borne in mind that the devil is in the mathematical details. Suppose we write two formulae, one describing the effects of painting an object and the other describing the effects of moving an object.

1. $Colour(x, c)$ holds after $Paint(x, c)$
2. $Position(x, p)$ holds after $Move(x, p)$

Now, suppose we have an initial situation in which $Colour(A, Red)$ and $Position(A, House)$ hold. According to the machinery of deductive logic, what then holds after the action $Paint(A, Blue)$ followed by the action $Move(A, Garden)$? Intuitively, we would expect $Colour(A, Blue)$ and $Position(A, Garden)$ to hold. Unfortunately, this is not the case. If written out more formally in classical predicate logic, using a suitable formalism for representing time and action such as the situation calculus (McCarthy & Hayes 1969), the two formulae above only license the conclusion that $Position(A, Garden)$ holds. This is because they don't rule out the possibility that the colour of A gets changed by the $Move$ action.

The most obvious way to augment such a formalisation so that the right common sense conclusions fall out is to add a number of formulae that explicitly describe the non-effects of each action. These formulae are called *frame axioms*. For the example at hand, we need a pair of frame axioms.

3. $Colour(x, c)$ holds after $Move(x, p)$ if $Colour(x, c)$ held beforehand
4. $Position(x, p)$ holds after $Paint(x, c)$ if $Position(x, p)$ held beforehand

In other words, painting an object will not affect its position, and moving an object will not affect its colour. With the addition of these two formulae (written more formally in predicate logic), all the desired conclusions can be drawn. However, this is not at all a satisfactory solution. Since *most* actions do not affect *most* properties of a situation, in a domain comprising M actions and N properties we will, in general, have to write out almost MN frame axioms. Whether these formulae are destined to be stored explicitly in a computer's memory, or are merely part of the designer's specification, this is an unwelcome burden.

The challenge, then, is to find a way to capture the non-effects of actions more succinctly in formal logic. What we need, it seems, is some way of declaring the general rule-of-thumb that an action can be assumed not to change a given property of a situation *unless* there is evidence to the contrary. This default assumption is known as the *common sense law of inertia*. The (technical) frame problem can be viewed as the task of formalising this law.

The main obstacle to doing this is the *monotonicity* of classical logic. In classical logic, the set of conclusions that can be drawn from a set of formulae always *increases* with the addition of further formulae. This makes it impossible to express a rule that has an open-ended set of exceptions, and the common sense law of inertia is just such a rule. For example, in due course we might want to add a formula that captures the exception to Axiom 3 that arises when we move an object into a pot of paint. But our not having thought of this exception before should not prevent us from applying the common sense law of inertia and drawing a wide enough set of (defeasible) conclusions to get off the ground.

Accordingly, researchers in logic-based AI have put a lot of effort into developing a variety of *non-monotonic* reasoning formalisms, such as *circumscription* (McCarthy 1986), and investigating their application to the frame problem. None of this has turned out to be at all straightforward. One of the most troublesome barriers to progress was highlighted in the so-called *Yale shooting problem* (Hanks & McDermott 1987), a simple scenario that gives rise to counter-intuitive conclusions if naively represented with a non-monotonic formalism. To make matters worse, a full solution needs to work in the presence of concurrent actions, actions with non-deterministic effects, continuous change, and actions with indirect ramifications. In spite of these subtleties, a number of solutions to the technical frame problem now exist that are adequate for logic-based AI research. Although improvements and extensions continue to be found, it is fair to say that the dust has settled, and that the frame problem, in its technical guise, is more-or-less solved (Shanahan 1997; Lifschitz 2015).

3. The Epistemological Frame Problem

Let's move on now to the frame problem as it has been re-interpreted by various philosophers. The first significant mention of the frame problem in the philosophical literature was made by Dennett (1978, 125). The puzzle, according to Dennett, is how "a cognitive creature ... with many beliefs about the world" can update those beliefs when it performs an act so that they remain "roughly faithful to the world"? In *The Modularity of Mind*, Fodor steps into a roboticist's shoes and, with the frame problem in mind, asks much the

same question: “How ... does the machine's program determine which beliefs the robot ought to re-evaluate given that it has embarked upon some or other course of action?” (Fodor 1983, 114).

At first sight, this question is only impressionistically related to the logical problem exercising the AI researchers. In contrast to the AI researcher's problem, the philosopher's question isn't expressed in the context of formal logic, and doesn't specifically concern the non-effects of actions. In a later essay, Dennett acknowledges the appropriation of the AI researchers' term (1987). Yet he goes on to reaffirm his conviction that, in the frame problem, AI has discovered “a new, deep epistemological problem—accessible in principle but unnoticed by generations of philosophers”.

The best way to gain an understanding of the issue is to imagine being the designer of a robot that has to carry out an everyday task, such as making a cup of tea. Moreover, for the frame problem to be neatly highlighted, we must confine our thought experiment to a certain class of robot designs, namely those using explicitly stored, sentence-like representations of the world, reflecting the methodological tenets of *classical AI*. The AI researchers who tackled the original frame problem in its narrow, technical guise were working under this constraint, since logic-based AI is a variety of classical AI. Philosophers sympathetic to the [computational theory of mind](#)—who suppose that mental states comprise sets of propositional attitudes and mental processes are forms of inference over the propositions in question—also tend to feel at home with this prescription.

Now, suppose the robot has to take a tea-cup from the cupboard. The present location of the cup is represented as a sentence in its database of facts alongside those representing innumerable other features of the ongoing situation, such as the ambient temperature, the configuration of its arms, the current date, the colour of the tea-pot, and so on. Having grasped the cup and withdrawn it from the cupboard, the robot needs to update this database. The location of the cup has clearly changed, so that's one fact that demands revision. But which other sentences require modification? The ambient temperature is unaffected. The location of the tea-pot is unaffected. But if it so happens that a spoon was resting in the cup, then the spoon's new location, inherited from its container, must also be updated.

The epistemological difficulty now discerned by philosophers is this. How could the robot limit the scope of the propositions it must reconsider in the light of its actions? In a sufficiently simple robot, this doesn't seem like much of a problem. Surely the robot can simply examine its entire database of propositions one-by-one and work out which require modification. But if we imagine that our robot has near human-level intelligence, and is therefore burdened with an enormous database of facts to examine every time it so much as spins a motor, such a strategy starts to look computationally intractable.

Thus, a related issue in AI has been dubbed the *computational aspect* of the frame problem (McDermott 1987). This is the question of how to compute the consequences of an action without the computation having to range over the action's non-effects. The solution to the computational aspect of the frame problem adopted in most symbolic AI programs is some variant of what McDermott calls the “sleeping dog” strategy (McDermott 1987). The idea here is that not every part of the data structure representing an ongoing situation needs to be examined when it is updated to reflect a change in the world. Rather, those parts that represent facets of the world that have changed are modified, and the rest is simply left as it is (following the dictum “let sleeping dogs lie”). In our example of the robot and the tea-cup, we might apply the sleeping dog strategy by having the robot update its beliefs about the location of the cup and the contents of the cupboard. But the robot would not worry about some possible spoon that may or may not be on or in the cup, since the robot's goal did not directly involve any spoon.

However, the philosophical problem is not exhausted by this computational issue. The outstanding philosophical question is how the robot could ever determine that it had successfully revised all its beliefs to match the consequences of its actions. Only then would it be in a position safely to apply the “common sense law of inertia” and assume the rest of the world is untouched. Fodor suggestively likens this to “Hamlet's problem: when to stop thinking” (Fodor 1987, 140). The frame problem, he claims, is “Hamlet's problem viewed from an engineer's perspective”. So construed, the obvious way to try to avoid the frame problem is by appealing to the notion of *relevance*. Only certain properties of a situation are relevant in the context of any given action, so the counter-argument goes, and consideration of the action's consequences can be conveniently confined to those.

However, the appeal to relevance is unhelpful. For the difficulty now is to determine what is and what isn't relevant, and this is dependent on context. Consider again the action of removing a tea-cup from the cupboard. If the robot's job is to make tea, it is relevant that this facilitates filling the cup from a tea-pot. But if the robot's task is to clean the cupboard, a more relevant consequence is the exposure of the surface the cup was resting on. An AI researcher in the classical mould could rise to this challenge by attempting to specify what propositions are relevant to what context. But philosophers such as Wheeler (2005; 2008), taking their cue from Dreyfus (1992), perceive the threat of infinite regress here. As Dreyfus puts it, “if each context can be recognized only in terms of features selected as relevant and interpreted in a broader context, the AI worker is faced with a regress of contexts” (Dreyfus 1992, 289).

One way to mitigate the threat of infinite regress is by appeal to the fact that, while humans are more clever than today's robots, they still make mistakes (McDermott 1987). People often fail to foresee every consequence of their actions *even though* they lack none of the information required to derive those consequences, as any novice chess player can testify. Fodor asserts that “the frame problem goes very deep; it goes as deep as the analysis of rationality” (Fodor 1987). But the analysis of rationality can accommodate the boundedness of the computational resources available to derive relevant conclusions (Simon 1957; Russell & Wefald 1991; Sperber & Wilson 1996). Because it sometimes jumps to premature conclusions, bounded rationality is logically flawed, but no more so than human thinking. However, as Fodor points out, appealing to human limitations to justify the imposition of a heuristic boundary on the kind of information available to an inferential process does not in itself solve the epistemological frame problem (Fodor 2000, Ch.2; Fodor 2008, Ch.4; see also Chow 2013). This is because it neglects the issue of how the heuristic boundary is to be drawn, which is to say it fails to address the original question of how to specify what is and isn't relevant to the inferential process.

Nevertheless, the classical AI researcher, convinced that the regress of contexts will bottom out eventually, may still elect to pursue the research agenda of building systems based on rules for determining relevance, drawing inspiration from the past successes of classical AI. Whereupon the dissenting philosopher might point out that AI's past successes have always been confined to narrow domains, such as playing chess, or reasoning in limited microworlds where the set of potentially relevant propositions is fixed and known in advance. By contrast, human intelligence can cope with an open-ended, ever-changing set of contexts (Dreyfus 1992; Dreyfus 2008; Wheeler 2005; Wheeler 2008; Rietveld 2012). Furthermore, the classical AI researcher is vulnerable to an argument from holism. A key claim in Fodor's work is that when it comes to circumscribing the consequences of an action, just as in the business of theory confirmation in science, anything could be relevant (Fodor 1983, 105). There are no *a priori* limits to the properties of the ongoing situation that might come into play. Accordingly, in his modularity thesis, Fodor uses the frame problem to bolster the view that the mind's central processes—those that are involved in fixing belief—are “informationally unencapsulated”, meaning that they can draw on information from any source (Fodor 1983; Fodor 2000; Fodor 2008; Dreyfus 1991, 115–121; Dreyfus 1992, 258). For Fodor, this is a fundamental barrier to the provision of a computational account of these processes.

It is tempting to see Fodor's concerns as resting on a fallacious argument to the effect that a process must be informationally encapsulated to be computationally tractable. We only need to consider the effectiveness of Internet search engines to see that, thanks to clever indexing techniques, this is not the case. Submit any pair of seemingly unrelated keywords (such as “banana” and “mandolin”) to a Web search engine, and in a fraction of a second it will identify every web page, in a database of several billion, that mentions those two keywords (now including this page, no doubt). But this is not the issue at hand. The real issue, to reiterate the point, is one of relevance. A process might indeed be able to index into everything the system knows about, say, bananas and mandolins, but the purported mystery is how it could ever work out that, of all things, bananas and mandolins were relevant to its reasoning task in the first place.

To summarize, it is possible to discern an epistemological frame problem, and to distinguish it from a computational counterpart. The epistemological problem is this: How is it possible for holistic, open-ended, context-sensitive relevance to be captured by a set of propositional, language-like representations of the sort used in classical AI? The *computational* counterpart to the epistemological problem is this. How could an inference process tractably be confined to just what is relevant, given that relevance is holistic, open-ended, and context-sensitive?

4. The Metaphysics of Common Sense Inertia

An additional dimension to the frame problem is uncovered in (Fodor 1987), where the metaphysical justification for the common sense law of inertia is challenged. Although Fodor himself doesn't clearly distinguish this issue from other aspects of the wider frame problem, it appears on examination to be a separate philosophical conundrum. Here is the argument. As stated above, solutions to the logical frame problem developed by AI researchers typically appeal to some version of the common sense law of inertia, according to which properties of a situation are assumed by default not to change as the result of an action. This assumption is supposedly justified by the very observation that gave rise to the logical frame problem in the first place, namely that most things don't change when an action is performed or an event occurs.

According to Fodor, this metaphysical justification is unwarranted. To begin with, some actions change many, many things. Those who affirm that painting an object has little or no effect on most properties of most of the objects in the room are likely to concede that detonating a bomb actually does affect most of those properties. But a deeper difficulty presents itself when we ask what is meant by “most properties”. What predicates should be included in our ontology for any of these claims about “most properties” to fall out? To sharpen the point, Fodor introduces the concept of a “fridgeon”. Any particle is defined as a *fridgeon* at a given time if and only if Fodor's fridge is switched on at that time. Now, it seems, the simple act of turning Fodor's fridge on or off brings about an astronomical number of incidental changes. In a universe that can include fridgeons, can it really be the case that most actions leave most things unchanged?

The point here is not a logical one. The effect on fridgeons of switching Fodor's fridge on and off can concisely be represented without any difficulty (Shanahan 1997, 25). Rather, the point is metaphysical. The common sense law of inertia is only justified in the context of the right ontology, the right choice of objects and predicates. But what is the right ontology to make the common sense law of inertia work? Clearly, fridgeons and the like are to be excluded. But what metaphysical principle underpins such a decision?

These questions and the argument leading to them are very reminiscent of Goodman's treatment of induction (Goodman 1954). Goodman's “[new riddle of induction](#)”, commonly called the *grue paradox*, invites us to consider the predicate *grue*, which is true before time t only of objects that are green and after time t only of objects that are blue. The puzzle is that every instance of a green emerald examined before time t is also an instance of a *grue* emerald. So, the inductive inference that all emeralds are *grue* seems to be no less legitimate than the inductive inference that all emeralds are green. The problem, of course, is the choice of predicates. Goodman showed that inductive inference only works in the context of the right set of predicates, and Fodor demonstrates much the same point for the common sense law of inertia.

An intimate relationship of a different kind between the frame problem and the problem of induction is proposed by Fetzer (1991), who writes that “The problem of induction [is] 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.” This view of the frame problem is highly controversial, however (Hayes 1991).

5. The Frame Problem Today

The narrow, technical frame problem generated a great deal of work in logic-based artificial intelligence in the late 1980s and early 1990s, and its wider philosophical implications came to the fore at around the same time. But the importance each thinker accords to the frame problem today will typically depend on their stance on other matters.

Within classical AI, a variety of workable solutions to the logical frame problem have been developed, and it is no longer considered a serious obstacle even for those working in a strictly logic-based paradigm (Shanahan 1997; Reiter 2001; Shanahan 2003; Lifschitz 2015). It's worth noting that logically-minded AI researchers can consistently retain their methodology and yet, to the extent that they view their products purely as engineering, can reject the traditional cognitive scientist's belief in the importance of computation over representations for understanding the mind. Moreover, insofar as the goal of classical AI is not computers with human-level intelligence, but is simply the design of better and more useful computer programs, it is immune to the philosophical objections of Fodor, Dreyfus, and the like. Significantly though,

for AI researchers working outside the paradigm of symbolic representation altogether—those working in situated robotics, for example—the logical frame problem simply doesn't feature in day-to-day investigations.

Although it can be argued that it arises even in a [connectionist](#) setting (Haselager & Van Rappard 1998; Samuels 2010), the frame problem inherits much of its philosophical significance from the classical assumption of the explanatory value of computation over representations, an assumption that has been under vigorous attack for some time (Clark 1997; Wheeler 2005). Despite this, many philosophers of mind, in the company of Fodor and Pylyshyn, still subscribe to the view that human mental processes consist chiefly of inferences over a set of propositions, and that those inferences are carried out by some form of computation. To such philosophers, the epistemological frame problem and its computational counterpart remain a genuine threat.

For Wheeler and others, classical AI and cognitive science rest on Cartesian assumptions that need to be overthrown in favour of a more Heideggerian stance before the frame problem can be overcome (Dreyfus 2008; Wheeler 2005; 2008; Rietveld 2012). According to Wheeler (2005; 2008), the situated robotics movement in AI that originated with the work of Brooks (1991) exemplifies the right way to go. Dreyfus is in partial agreement, but contends that the early products of situated robotics “finesse rather than solve the frame problem” because “Brooks's robots respond only to fixed isolable features of the environment, not to context or changing significance” (Dreyfus 2008, 335). Dreyfus regards the neurodynamics work of Freeman (2000) as a better foundation for the sort of Heideggerian approach to AI in which the frame problem might be dissolved (see also Shanahan 2010, Ch.5; Rietveld 2012; Bruineberg & Rietveld 2014). Dreyfus is impressed by Freeman's approach because the neurodynamical record of significance is neither a representation nor an association, but (in dynamical systems terms) “a repertoire of attractors” that classify possible responses, “the attractors themselves being the product of past experience” (Dreyfus 2008, 354).

One philosophical legacy of the frame problem is that it has drawn attention to a cluster of issues relating to holism, or so-called informational unencapsulation. Recall that a process is informationally unencapsulated (Fodor sometimes uses the term “isotropic”) if there is no *a priori* boundary to what information is relevant to it. In recent writing, Fodor uses the term “frame problem” in the context of all informationally unencapsulated processes, and not just those to do with inferring the consequences of change (Fodor 2000, Ch.2; Fodor 2006, Ch.4). It's clear that idealised rationality is informationally unencapsulated, in this sense. It has also been suggested that isotropy is damaging to the so-called [theory theory](#) of folk psychology (Heal 1996). (For Heal, this lends support to the rival [simulation theory](#), but Wilkerson (2001) argues that informational unencapsulation is a problem for both accounts of folk psychology.) Analogical reasoning, as Fodor says, is an example of “isotropy in the purest form: a process which depends precisely upon the transfer of information among cognitive domains previously assumed to be irrelevant” (Fodor 1983, 105). Arguably, a capacity for analogical and metaphorical thinking—a talent for creatively transcending the boundaries between different domains of understanding—is the source of human cognitive prowess (Lakoff & Johnson 1980; Mithen 1996). So the informational unencapsulation of analogical reasoning is potentially very troublesome, and especially so for modular theories of mind in which modules are viewed as (context-insensitive) specialists (Carruthers 2003; 2006).





Dreyfus claims that this “extreme version of the frame problem” is no less a consequence of the Cartesian assumptions of classical AI and cognitive science than its less demanding relatives (Dreyfus 2008, 361). He advances the view that a suitably Heideggerian account of mind is the basis for dissolving the frame problem here too, and that our “background familiarity with how things in the world behave” is sufficient, in such cases, to allow us to “step back and figure out what is relevant and how”. Dreyfus doesn't explain how, given the holistic, open-ended, context-sensitive character of relevance, this figuring-out is achieved. But Wheeler, from a similarly Heideggerian position, claims that the way to address the “inter-context” frame problem, as he calls it, is with a dynamical system in which “the causal contribution of each systemic component partially determines, and is partially determined by, the causal contributions of large numbers of other systemic components” (Wheeler 2008, 341). A related proposal is put forward by Shanahan and Baars (2005; see also Shanahan 2010, Ch.6), based on global workspace theory (Baars 1988), according to which the brain incorporates a solution to the problem of informational unencapsulation by instantiating an architecture in which a) the responsibility for determining relevance is not centralised but is distributed among parallel specialist processes, and b) a serially unfolding global workspace state integrates relevant contributions from multiple domains.

Bibliography

- Baars, B. (1988), *A Cognitive Theory of Consciousness*, Cambridge University Press.
- Brooks, R.A. (1991), "Intelligence without Reason", in *Proc. 12th International Conference on Artificial Intelligence*, pp. 569–595.
- Bruineberg, J. & Rietveld, E. (2014), "Self-organization, Free Energy Minimization, and Optimal Grip on a Field of Affordances", *Frontiers in Human Neuroscience*, 8: 559, doi:10.3389/fnhum.2014.00599
- Carruthers, P. (2003), "On Fodor's Problem", *Mind and Language*, 18(5): 502–523.
- (2006), *The Architecture of the Mind*, Oxford University Press
- Chow, S.J. (2013), "What's the Problem with the Frame Problem?", *Review of Philosophy and Psychology*, 4: 309–331.
- Clark, A. (1997), *Being There: Putting Brain, Body, and World Together Again*, MIT Press.
- Dennett, D. (1978), *Brainstorms*, MIT Press.
- (1987), "Cognitive Wheels: The Frame Problem in Artificial Intelligence", in Pylyshyn (1987).
- Dreyfus, H.L. (1991), *Being-in-the-World: A Commentary on Heidegger's Being and Time, Division I*, MIT Press.
- (1992), *What Computers Still Can't Do*, MIT Press.
- (2008), "Why Heideggerian AI Failed and How Fixing It Would Require Making It More Heideggerian", in *The Mechanical Mind in History*, P. Husbands, O. Holland & M. Wheeler (eds.), MIT Press, pp. 331–371.
- Fetzer, J.H. (1991), "The Frame Problem: Artificial Intelligence Meets David Hume", in Ford & Hayes (1991).
- Fodor, J.A. (1983), *The Modularity of Mind*, MIT Press.
- (1987), "Modules, Frames, Fridgeons, Sleeping Dogs, and the Music of the Spheres", in Pylyshyn (1987).
- (2000), *The Mind Doesn't Work That Way*, MIT Press.
- (2008), *LOT 2: The Language of Thought Revisited*, Oxford University Press.
- Ford, K.M. & Hayes, P.J. (eds.) (1991), *Reasoning Agents in a Dynamic World: The Frame Problem*, JAI Press.
- Ford, K.M. & Pylyshyn, Z.W. (eds.) (1996), *The Robot's Dilemma Revisited: The Frame Problem in Artificial Intelligence*, Ablex.
- Freeman, W.J. (2000), *How Brains Make Up Their Minds*, Phoenix.
- Goodman, N. (1954), *Fact, Fiction, and Forecast*, Harvard University Press.
- Hanks, S. & McDermott, D. (1987), "Nonmonotonic Logic and Temporal Projection", *Artificial Intelligence*, 33(3): 379–412.
- Haselager, W.F.G. & Van Rappard, J.F.H. (1998), "Connectionism, Systematicity, and the Frame Problem", *Minds and Machines*, 8(2): 161–179.
- Hayes, P.J. (1991), "Artificial Intelligence Meets David Hume: A Reply to Fetzer", in Ford & Hayes (1991).
- Heal, J. (1996), "Simulation, Theory, and Content", in *Theories of Theories of Mind*, eds. P.Carruthers & P.Smith, Cambridge University Press, pp. 75–89.
- Lakoff, G. & Johnson, M. (1980), *Metaphors We Live By*, University of Chicago Press.
- Lifschitz, V. (2015), "The Dramatic True Story of the Frame Default", *Journal of Philosophical Logic*, 44: 163–196.
- McCarthy, J. (1986), "Applications of Circumscription to Formalizing Common Sense Knowledge", *Artificial Intelligence*, 26(3): 89–116.
- McCarthy, J. & Hayes, P.J. (1969), "Some Philosophical Problems from the Standpoint of Artificial Intelligence", in *Machine Intelligence 4*, ed. D.Michie and B.Meltzer, Edinburgh University Press, pp. 463–502.
- McDermott, D. (1987), "We've Been Framed: Or Why AI Is Innocent of the Frame Problem", in Pylyshyn (1987).
- Mithen, S. (1987), *The Prehistory of the Mind*, Thames & Hudson.
- Pylyshyn, Z.W. (ed.) (1987), *The Robot's Dilemma: The Frame Problem in Artificial Intelligence*, Ablex.
- Rietveld, R. (2012), "Context-switching and Responsiveness to Real Relevance", in J.Kiverstein & M.Wheeler (eds.), *Heidegger and Cognitive Science: New Directions in Cognitive Science and Philosophy*, Palgrave Macmillan, pp. 105–135.
- Reiter, R. (2001), *Knowledge in Action: Logical Foundations for Specifying and Implementing Dynamical Systems*, MIT Press.

- Russell, S. & Wefald, E. (1991), *Do the Right Thing: Studies in Limited Rationality*, MIT Press.
- Samuels, R. (2010), “Classical Computationalism and the Many Problems of Cognitive Relevance”, *Studies in History and Philosophy of Science*, 41: 280–293.
- Shanahan, M. (1997), *Solving the Frame Problem: A Mathematical Investigation of the Common Sense Law of Inertia*, MIT Press.
- (2003), “The Frame Problem”, in *The Macmillan Encyclopedia of Cognitive Science*, L. Nadel (ed.), Macmillan, pp. 144–150.
- (2010), *Embodiment and the Inner Life: Cognition and Consciousness in the Space of Possible Minds*, Oxford University Press.
- Shanahan, M. & Baars, B.J. (2005), “Applying Global Workspace Theory to the Frame Problem”, *Cognition*, 98(2): 157–176.
- Simon, H. (1957), *Models of Man*, Wiley.
- Sperber, D. & Wilson, D. (1996), “Fodor's Frame Problem and Relevance Theory”, *Behavioral and Brain Sciences*, 19(3): 530–532.
- Wheeler, M. (2005), *Reconstructing the Cognitive World: The Next Step*, MIT Press.
- (2008), “Cognition in Context: Phenomenology, Situated Robotics, and the Frame Problem”, *International Journal of Philosophical Studies*, 16(3): 323–349.
- Wilkerson, W.S. (2001), “Simulation, Theory, and the Frame Problem”, *Philosophical Psychology*, 14(2): 141–153.

Academic Tools

-  [How to cite this entry.](#)
-  [Preview the PDF version of this entry](#) at the [Friends of the SEP Society](#).
-  [Look up this entry topic](#) at the [Indiana Philosophy Ontology Project](#) (InPhO).
-  [Enhanced bibliography for this entry](#) at [PhilPapers](#), with links to its database.

Other Internet Resources

- [Articles on the Frame Problem](#) in *Psycoloquy*, a refereed electronic journal sponsored by the American Psychological Association.

Related Entries

[artificial intelligence: logic and](#) | [cognitive science](#) | [folk psychology: as a theory](#) | [folk psychology: as mental simulation](#) | [induction: problem of](#) | [learning theory, formal](#) | [logic: action](#) | [mental representation](#) | [mind: computational theory of](#)

Copyright © 2016 by
Murray Shanahan <m.shanahan@ic.ac.uk>

This is a file in the archives of the Stanford Encyclopedia of Philosophy.
Please note that some links may no longer be functional.

Stanford | Center for the Study of
Language and Information

The Stanford Encyclopedia of Philosophy is [copyright © 2015](#) by [The Metaphysics Research Lab](#), Center for the Study of Language and Information (CSLI), Stanford University

Library of Congress Catalog Data: ISSN 1095-5054