Self-* information systems: why not?

Geoffrey S. Canright
Telenor R&D, 1331 Fornebu, Norway

Note: This is a position paper. That means that it is full of opinions and devoid of results. It is intended only to stimulate thought and discussion. It is written in an informal style, and I use the first person freely. Finally, note that this is directed to attendants to the self-* conference; hence I make frequent reference to self-* concepts, without defining them. My principal aim here is to generate interesting questions.

I start with the first, and perhaps broadest, question in the self-* invitation:

Q1: Is there a valid scientific basis for self-* computing, or is it mainly hype?

A1: Yes, there is a valid scientific basis.

To support this answer I point to biology, ie, the fact that life exists. Whatever kind of * we think of, life exhibits self-*. Therefore the existence of life is a proof by example that self-* is possible.

This very short answer seems to be cheating. That is, can the answer be as simple as that?

Here is a possible objection, phrased as a new question:

Q2: OK, so biological systems are self-*. But might there be some properties or capabilities of living systems that cannot be mimicked or reproduced by technological information systems?

A2: Now the answer is not so easy. But I would say that there is no important property of living systems that cannot also appear in, or be built into, future information and communication systems.

In other words, Q2 is a rephrasing of Q1, which focuses on the difficult part: what is it about living systems that makes them able to be self-*? And whatever that "it" is, can future information systems also have "it"?

Well, what is "it"? We have long ago abandoned "vitalism": the idea that there was something *essentially* different about living matter, as compared to nonliving matter [1]. In fact, modern science has come almost completely to the opposite view: living stuff is the same as nonliving stuff. It is just organized differently, it is more "lively"; it is self-*.

And this opposite view is being confirmed daily, in a growing stream of research results which take as their starting point that there is no essential difference, and then demonstrate that by, say, coupling silicon circuits to living neuronal circuits. Plenty of other examples can be cited!

So life is life because it is organized in interesting ways. Organization requires information flow. That is, parts of an organized whole must communicate with one another—perhaps only during morphogenesis, but more likely an ongoing flow of information is needed to maintain the pattern of organization.

So, rather crudely, life is pattern which maintains itself by appropriate flow of information. We can expand the list of verbs here: life is patterns that establish, maintain, repair, reproduce, and evolve themselves. And the maintenance of these patterns requires a certain flow of information—within the cell, within the organisms, within the ecosystem, etc.

Given this picture of what life is, there is then no essential difference between what living systems can do and what artificial systems can do. Therefore, if life can have self-* properties, so can nonliving systems.

In particular, systems which deal in patterns built from information rather than hard matter, and which allow for detailed and microscopic control over the flow of information, seem like *promising* candidates for life-like properties. That is, it seems in a sense easier to establish lifelike properties in a distributed information system than in, for example, mechanical systems such as robots. The patterns in information systems are essentially pure information—something that is easily copied and transported. And underlying physical technology—an outstanding example being the Internet—allows for good information flow between various parts of the patterns. In contrast, the robots' parts are less easily copied or transported; and good communication among a flock of robots (especially if they are mobile) is challenging.

Thus we have an answer to Q2, and an argument for that answer.

If we accept A1 and A2, then we are ready to try to build artificial systems with lifelike properties. But then we must ask, How to do it? The obvious followup question is, how did life do it? That is, how did life become life, and in so doing acquire lifelike, self-* properties?
This question gets no number, because it comes purely from biology, and we think we know the answer: evolution.

Stuart Kauffman [2] likes to argue that natural selection is not the only mechanism which can give rise to organized patterns. He points out that in many interesting cases, organization comes "for free". And of course he is right. Yet this point is largely irrelevant to our biological question, whose answer is still: life became life, and then became what it is now, through evolution. And natural selection played a crucial role in determining the course of that evolution. The point here is that the question of "organization", fascinating as it is, is not really relevant. If life were to be a mass of *disorganized* patterns which still managed to maintain and reproduce themselves, *better than any other patterns can,* then nothing changes in our argument. That is, life is not only patterns that maintain themselves: today's life is those patterns that are *most* successful—compared to extinct life—at maintaining themselves (including of course via reproduction). Perhaps these most successful patterns gain some *advantage* from being "organized" (precise definition lacking); but here I only focus on the *persistence* (in the sense of survival and reproduction) of patterns. Hence the question of the role of "organization" (which is both highly interesting and highly nontrivial) seems not to be relevant to my argument.

In short: life has used evolution and selection to become what it is. Now we want to build artificial systems that are "good", by some criteria which we human engineers impose. Some of those criteria are that they should be "lifelike", ie, good at building, maintaining, and repairing themselves. And it is just these criteria which are selected by natural selection.

Then I come to my next numbered question.

Q3: What role can evolution and natural selection play in the development of artificial systems with self-* properties?

I find more than one answer to Q3.

A3.0. It seems foolish *not* to use variation and selection. The alternative is to try to create new, lifelike systems purely by *design*. This strikes me as an extremely difficult way to go—and my experience with the BISON project just strengthens that view. I believe we need both creative design, and some kind of variation and selection process.

A3.1. Evolution via natural selection, with variation fuelled by random mutations, is "trial and error". Trial and error is acceptable engineering practice when it is done "offline": in a laboratory, wind tunnel, simulator, etc. Trial and error is *not* an acceptable approach for online, running systems. Therefore, do it, but do it offline.

This answer seems simple enough. There is, probably, a catch. That is, offline tests, for lifelike, self-organizing systems, may not be reliable enough.

We only have to look to biology to see this. Firm Z (or Professor P) invents a new, genetically modified organism. This is the "trial" part of "trial and error". But now we want to avoid the worst part of "error": for instance, we don't want to let loose an organism that will (to state the case in the extreme) lead to the extinction of the human race. But that's too easy: even if we can rule that out (with very high probability), there are a large number of other undesirable outcomes that we also wish to rule out. To do so, we must be able to *predict* the effects of introducing a new organism into the whole, worldwide ecosystem. This kind of prediction is notoriously hard. Most current societies are living—and not by choice—with unwanted and unforeseen ecological consequences of choices made earlier. Here I think of invasive species, desertification, extinctions … various kinds of ecological instabilities which—being instabilities—can be very hard to predict.

I am tempted to digress, to give an example from physics [3]. It is not very relevant, but it is colourful! During the second world war, scientists at Los Alamos, working to invent fission bombs (and thinking about fusion bombs), had to face and answer the question: could the detonating of any of these bombs lead to a runaway reaction in which the Earth's atmosphere, or the oceans, might be ignited and burned up? These fine scientists found that such a result was nearly impossible. Then the experiment was done.

Back to the point: complex systems are hard to predict. More specifically, the prediction of the consequences of new organisms/mechanisms, in a system as complex as an ecosystem, is hard. And yet one needs some ability to predict, in order to extrapolate from offline simulations to online experiments.

In short: evolution, variation and selection are most safely done offline. The problem is then to be able to *rely* on the offline results, as predictors of the online results. There is the danger of drawing erroneous conclusions from the offline results; and then the further danger of not being able to reverse, or even halt, the undesirable outcome of the subsequent online experiments.

This brings us to a third answer to Q3.

A3.2. We have no choice: evolution and selection are already operating on the Internet—*online*—and will continue to do so.

Proof by example: computer viruses. They are close to perfect analogs of biological viruses. The principal difference that I see, is that their variation is mostly or entirely still human-driven, ie random mutations seem to play little or no role. But there is clearly variation, selection, and evolution going on.

That is, hackers have already begun to experiment with digital, self-reproducing systems. So far, we use non-adaptive and mostly non-self-reproducing systems to fight computer viruses—someone (human) must develop the antivirus software, and someone must take the responsibility for downloading the antivirus software. Thus we are fighting evolving, self-reproducing systems with evolving, non-self-reproducing systems. In each case, the evolution is essentially human-driven. But the thing we are fighting seems clearly more lifelike than the things that are fighting it.

Is it then a good idea to "fight fire with fire"? This brings us to Q4.

Q4. Suppose we can build lifelike, self-* systems in principle, and also we know how to do it in practice —at least, to some extent. The question is then: are such systems *desirable*?

One gets the impression that a Yes answer is at least implicit in the formulation of the self-* conference. However, it should also be clear from the above discussion that self-* systems can have a down side: systems that are self-reproducing, self-propagating, etc, can be very hard to get rid of—even if they turn out to have undesirable properties.

Perhaps however it is reasonable to assume that the situation need never be worse than it is now—where "most" engineers and hackers are interested in making "nice" systems. Here, in the context of this question, "nice" must mean at least two things: the old meaning, that the aim is to have effects that are judged as desirable, or at least acceptable, by most people; and a new meaning, that any self-* systems should be made in such a way so that they can be deactivated, rendered inoperable, "killed"—if they get out of line.

Of course, we immediately face the problem: what if the Achilles-heel mechanism—that is, the one built in to make the self-* thing easy to kill—is the one that "gets out of line"? Well then, it seems, it's back to the drawing board. By definition, there is no easy fix for this kind of failure.

Why should such systems ever fail? Well, there are the usual reasons, coming from human fallibility. I find further reason for concern from A3.1. That is, self-* systems are likely to be harder to predict than more "inert" systems such as physical machines. And predictions from offline studies may prove to be wrong, perhaps disastrously wrong.

Note that Q4 is not a scientific question in itself. Its answer however depends on scientific answers: how well can we bound the error in predicting what new self-* systems will do? How small is the likelihood of this or that disastrous outcome?

The alert reader will point out that these questions have always been present for engineered systems: what is new here? Let us number this important question.

Q5. Is the unpredictability of lifelike information systems different from that of traditional, mechanical, engineered systems, such that the former are likely to be significantly harder to engineer with?

A very alert reader might note that I never really answered Q4. I will not really offer an answer to Q5 either. Instead I will simply offer some discussion.

We have learned that all systems are unpredictable in principle. Engineering is about *bounding* the unpredictability. We never cease to find examples in the news, of highly engineered systems whose behavior goes outside expected bounds—space shuttle catastrophes, bridge collapses etc. The element of unpredictability is always present, and we can never absolutely rule out catastrophic events.

Q5 then asks, Can self-managing, self-repairing, self-* systems exhibit a kind of unpredictability that is in some sense greater—in principle, and *in practice* (since we are focused on engineering)—than that seen in traditional engineered systems?

Again I would turn to biology in order to seek answers. The biological picture then suggests a very tentative answer, as follows. Biology has succeeded at developing patterns (cells, organisms) with a very nice degree of predictability. Just think of the reliability of morphogenesis, from one generation to the next, over thousands of generations. Or of the reliability of DNA replication. In fact, there is selection *against* too much instability in biology.

Let us then assume that we can design systems—say, algorithms which are self-*—that are, like biological organisms, quite reliable in the sense that they reproduce not only their code, but also their behavior, in a very reliable fashion. Have we not then answered the reliability/predictability question raised by Q5?

No. We need to be able to predict the behavior, not only of the "parts" (organisms), but also of the whole system of interacting organisms. Consider how obviously unstable biology is, at the level of ecosystems. Reliable organisms can face a new environment, due (eg) to the introduction of new organisms, so that they must either change or die. The organism is just as reliable as we thought; but its environment is not. The extreme example:

biology has produced humans, who reliably produce more humans, who in turn drive ecological change at an accelerating pace. Ecologies are unstable.

Similarly, my guess is that the worst unpredictability from engineered, self-* systems would arise at the level of the ecosystem. Firm Z (or Professor P) develops a fine, new, self-managing digital organism; and it does pretty much what its inventor expects. We suppose that this is one of the first such digital organisms; so perhaps its environment is relatively stable. But then Firm Z' puts its organism out on the net; Firm Z'' adds yet another; etc. The ecosystem of these organisms becomes more complex, richer—and more unpredictable. Can we test our newest digital organism offline, and predict, accurately enough, how it will interact with all the other ones out there? I do not offer an answer here. I only suggest that the unpredictability problem will be toughest at the ecosystem level.

Even this however may not be anything new for human societies. The picture I have just sketched has much in common with economic systems—which are, in a sense, a cultural form of human ecosystem, where money replaces food/energy, and bankruptcy stands in for death. Economic systems are also notoriously hard to predict; but it seems that the frequency of big catastrophes can be held to a low enough level that the whole system does not die. That is, few are tempted to renounce money in all its forms. My final, very tentative, suggestion is then that the coming digital ecosystem—already partly visible—will be unruly and hard to predict, but not much worse in practice than our existing economic systems.

References.

[1] For a good historical account see Graeme K Hunter, *Vital Forces: the discovery of the molecular basis of life*, Academic Press, 2000.

[2] Stuart A Kauffman, *The Origins of Order: self-organization and selection in evolution,* Oxford University Press, 1993.

[3] For two versions of this story see: Richard Rhodes, *The Making of the Atomic Bomb,* Simon & Schuster, 1986, pp 418—419; and Maxim D Frank-Kamenetskii, *Unraveling DNA: the most important molecule of life*, Addison-Wesley, 1997, p 124.