

# HOW TO SAVE THE WORLD

*IN FIVE SIMPLE STEPS*

By Sabine Hossenfelder

**Abstract:** If you knew how humanity should steer the future, what difference would it make? The major challenge that humanity faces today is not that we lack ideas for what to do, as I am sure this essay contest will document. No, the major challenge, the mother of all problems, is to convert these ideas into courses of action.

We fail to act in the face of global problems because we do not have an intuitive grasp on the consequences of collective human behavior, are prone to cognitive biases, and easily overwhelmed by data. We are also lazy and if intuition fails us, inertia takes over. How many people will read these brilliant essays?

For the individual, evaluating possible courses of action to address interrelated problems in highly connected social, economic and ecological networks is presently too costly. The necessary information may exist, even be accessible, but it is too expensive in terms of time and energy. To steer the future, information about our dynamical and multi-layered networks has to become cheap and almost effortless to use. Only then, when we can make informed decisions by feeling rather than thinking, will we be able to act and respond to the challenges we face.

## THE PROBLEM.

The most remarkable fact about humans is the utter uselessness of our infants. Humans, in contrast to all other species, must learn almost everything necessary for survival. It takes us a long time to reach maturity, time in which parents have to prevent their offspring from eating sand, chopping off fingers, or accidentally wiping out the human race by growing super-resistant bugs behind their ears.

But our ability to learn, combined with technics to communicate information, is also what enables us to adapt to changing environments faster than gene selection could possibly achieve. We are awed by inborn knowledge – butterflies that recall routes of their ancestors – but we outpaced our competitors by changing the rules of what it means “to adapt” itself. We do not wait for physiological changes to result in better Darwinian fitness. Instead, we modify our environment, ourselves, and our interaction with the environment to get there faster, to “fit” better than any other species.

The root of the problems that humanity faces today is that our adaptation as a species has fallen behind the changes we have induced ourselves. As human interactions get more complex, as networks spread globally and become tightly coupled, we need systems that are able to learn and in return help let us learn about the system. But we don't have them.

The political, economic, and social systems that govern our lives are presently adaptive by trial and error. But much like gene selection is too slow to have yet adapted humans to a mostly sedentary city life and goat memes, the adaptation of our systems by trial and error is too slow to solve the problems we presently face. May that be climate change, the global water crisis, the big garbage swirl, or the fragility of our financial systems – our inability to process these problems means that we, as the actors in the system, do not respond, indeed cannot respond, to the information we have in a timely manner. And so, problems persist and build up.

The necessary information for individuals to learn and react to systemic trends may be available, even accessible, but it is too expensive. Information is presently costly, not necessarily financially, but in the amount of effort required to obtain and understand it. Relevant information is too difficult to find or comprehend and doing so requires too much time and energy. Blaming people for being politically disinterested, scientifically illiterate, or plainly unintellectual doesn't do anything to address the costliness of information and thus doesn't do justice to the origin of the problem. The individual investment necessary to process information about trends and relations in our systems is currently too high and personal benefits do not outweigh the disadvantages.

There is no shortage of well-intentioned institutions and organizations that aim at one or several of humanity's problems. The biggest, most existential, problems have been collected by the Future of Humanity Institute [1]. These are the problems that can lead to extinction, near-extinction or progress stagnation of the human species, including but not limited to nuclear accidents, asteroid impacts, artificial intelligence and biotechnology. In case you're not a worrier, check their webpage.

The Future of Humanity Institute is also a point in case because it has next to no influence on the future of humanity. The same can be said about all other initiatives that collect data about our global networks and use buzzwords like complexity and interdisciplinarity. For example the United Nations Global Pulse [2], the FutureICT [3] or the PSIR Model [4]. All such initiatives fail to address the main problem, which isn't to collect information, but feeding this information back into the system, back to the many humans who are the initiators of change.

Some believe quantum computing will solve our problems. It won't. Computation alone cannot solve our problems for the same reason political utopias, however beautiful and ingenious, never solved any problems: Because humans don't care what somebody or some thing thinks they should be doing. They'll do whatever they please. The only way to change their ways is to please them. Please them differently than before, and change will follow.

No, the biggest challenge mankind faces today is not the development of some breakthrough technology. The biggest challenge is to create a society whose institutions integrate the knowledge that must precede any breakthrough technology: The knowledge about the systems themselves that is necessary for the realization, adaptation and use of technologies. All of our big problems today speak of our failure, not to envision solutions, but to turn our ideas and knowledge into reality. We have a social problem, not a technological one.

We reached this gridlock because the human brain did not evolve to understand the consequences of individual actions in networks of billions of people. We are bad in making good long-term

decisions and do not care much what happens in other parts of the planet to people we have not and will most likely never meet. We have no intuitive grasp on the collective behavior of large groups and their impact on our environment, and what little grasp we have is prone to cognitive biases and statistical errors, many of which are now subject of new scientific areas like game theory, behavioral economics and decision science.

These cognitive shortcomings are not only obstacles to solving our problems, they *are* the problem. But these are obstacles that science can overcome.

## THE SOLUTION.

The human brain has the capacity to evaluate decisions that have long-term and large-scale consequences. However, frequently decisions which are beneficial on long time- or distance scales conflict with those on short time- or distance scales. Due to evolutionary developed reward circuits, this conflict is often resolved in favor of short times and distances<sup>1</sup>.

But we know how to solve these problems. We solve them by bringing close that what is far away. This is why people in weight-loss programs (distant) are encouraged to reward themselves (close) for holding onto their diet. This is why they pin photos (distant) on their fridge (close). This is why the World Wildlife Fund lets you adopt baby animals of endangered species (distant) and sends you a certificate (close). This is why you are shown all the photos of hungry, ill, injured or otherwise suffering children. You get the picture. It brings distant information closer and taps onto your emotional responses, which is a fast, simple, and effective reaction. It wires back into the circuits that your brain is used to work with. This wiring can be abused, all right. But used the right way, it carries the solution to our problem.

“Gamification” is a recent variant of this mechanism. Gamification is growing popular to help people balance their own priorities, typically by providing instant rewards (in terms of collecting points) for behavior users previously themselves identified as desirable (say, eating healthy). Seen from a system’s perspective, this is an external feedback loop that allows humans to use old brain circuits to adapt good (here: healthy) behavior faster than gene selection could achieve with a turnover rate of many generations. The interesting aspect about gamification is how little is necessary to make this feedback loop work. All it takes is a simple and intuitive visualization that lets users immediately grasp how well an action matches with their stated goals. The keywords here are: Simple, intuitive, and immediate. This is cheap information.

The solution to our problems is a generalization of this feedback loop: To give people access to cheap information about the consequences of collective human actions, and in return use their reaction to this information to improve the system, ie the way individual actions are coordinated.

---

<sup>1</sup> Here, with “distance” I am referring not necessarily to spatial distance, but to distance in social networks and other infrastructure networks.

The point here is not to manipulate people into changing their ways because I or you or some supercomputer thinks it would be better if we'd do more of this or more of that. The point is to help people make decisions. The way we presently make decisions, part of our priorities remain neglected because we cannot assess how well we would be working towards them. It's too complicated, too costly. But it's not like we are happy with this. Most people notice the tension, the neglect of some of their priorities, and are left with bad consciousness, the nagging voice that says you should make better decisions. If only you had the time and it wasn't so difficult.

The feedback system that we need has to give the user an intuitive feeling for how well a decision matches with recorded priorities. If such a feedback in the future can be given by a brain implant, it will be like an additional sense. How does this decision taste? How well does it match with my preferences? Does this choice look harmonious? Does it sound good? Such a feedback is the natural extension of our ability to judge the result of our actions in small groups. This is what it takes to make information cheap, really cheap, so that using it becomes almost effortless.

This feedback loop might include for example information about how well buying a product matches with the relevance one has assigned to certain health goals or its environmental impact or its contribution to the local economy. This is information which a customer doesn't normally have when making a purchase (though economic theory maintains it is taken into account). And even if they had the information, they probably wouldn't study it.

Other examples are questions like: If I dispose of that plastic bottle here, how likely is it to be recycled or to end up in the ocean? If I buy the fair trade coffee, does it work towards something I value? Do I help the homeless guy more by giving him some dollars or by donating that money somewhere? How much of the tax I pay on this item subsidizes projects I support? It's not that people do not care. It's just that in practice it takes too much effort to look into the details. And they actually do not want to know the details. All they want to know is whether, according to best present knowledge, a certain decision works towards their goals. And most of the time that is really all they need to know.

Let me use another example, a somewhat shocking one that however illustrates well distance among people. A recent study by researchers from Princeton University asked participants to judge the competence of political candidates by split-second looks at photos. It turned out that this snap judgment predicted very well who would eventually be voted [5].

How incredibly shallow we are. But forgive us. We decode human faces constantly and effortlessly and the human brain always tries to save energy. We use emotional response to somebody's look to assess how much we can trust them. That's not an optimal assessment for informed decision making. It certainly gives me to think that my opinion of political candidates probably depends on the shape of their nose. I really should go and read all these programs, comments and opinion pieces. But I have an essay to finish before the deadline, then write this overdue report and hurry to pick up the kids from daycare. Maybe I'll look up these candidates next week. Or the week after that. If only information wasn't so costly. If only it wouldn't take up so much time and energy.

But now imagine you could look at a candidate and in fact get a simple, fast, sensorial or emotional feedback how certain selected priorities and interest of this person match with yours. This would dramatically lower the cost of information. It would bring close that what is far away.

The ingredients for closing this feedback loop already exist, they just aren't combined suitably. Above I mentioned gamification to bridge long time distances. Other applications that make information less costly in terms of time and energy are sites dedicated to help you decide which party to vote based on answers to a set of questions, or dating sites that match your interests with potential mates. It's the same mechanism, but too scattered and not broad enough. The more dispersed these applications are, the more effort it takes to use them and the more costly the information becomes. We need it all in one place.

Concretely the feedback loop would work like this:

1. A user creates a personal priority map. In the future this may be done by a brain scan or by analyzing information transmitted from neural implants. Presently, questionnaires and other records must stand in. The questionnaires would cover for example personal values, various aspects of health and social life, political attitudes and personal taste. This should also include users' tolerance for risk and uncertainty because this is relevant to assess how good a match will be. This priority map is personal data that the user can update and expand, and share or make public selected parts of it.
2. Institutions that gather knowledge about the system (statistics, trends, predictions) make it available to users as correlations between actions and individual priorities. In return they use the shared parts of users' maps to obtain better information about the system, notably tensions that arise when priorities conflict, which can indicate problems with the current organization of the system.
3. Whenever a user takes a decision whose impact is likely to exceed the natural human ability to foresee consequences of individual actions he or she consults the priority map. The user can then tell how well a decision matches their recorded priorities and take this into account without having to bother with the details in every single case. The decisions serve to adapt the system.

This consultation of the priority map thus remedies the lack of intuition humans have for the behavior of highly connected and dynamic social networks. The goal is that users are able to make informed decisions with snap judgments: Simple, intuitive, immediate. Only then will we change the ways of the bulk of people on the planet.

In the future, information about matches with personal priorities may be delivered wirelessly to brain implants, constituting an upgrade of humanity for global interactions. I only discussed here the evaluation of selected decisions, not how to find the best possible course of action according to certain criteria. The latter is a much harder problem. We can note in the passing though that it constitutes an optimization problem and thus lends itself for adiabatic quantum computing.

With presently existing technology we have to settle for visualizing a match or mismatch rather than feeling it. The visualization of big data sets and the possibility to manipulate them interactively

is rapidly improving, and such interaction with data will already serve to make information dramatically cheaper. And it really has to become cheap.

We do not get anywhere with bemoaning that most people do not understand climate models or do not read information brochures about genetically modified crops. It is time to wake up. We've tried long enough to educate them. It doesn't work. The idea of the educated and well-informed citizen is an utopia. It doesn't work because education doesn't please people. They don't like to think. It is too costly and it's not the information they want. What they want is to know how much an estimated risk conflict with their priorities, how much an estimated benefit agrees with their values. They tolerate risk and uncertainty, but they don't tolerate science lectures. If a webpage takes more than 3 seconds to load it'll lose 40% of visitors. Split-second looks at photos. That's the realistic attention span. That's what we have to work with.

In some regards we are already on the way to close this feedback loop. Many scientific institutions share information and take science communication seriously. However, presently this information is still too much, too unclear, and not available to individuals at the right moment.

In other regards we have a long way to go. We do not presently use people's priorities in any systematic way to discover shortcomings in the system and improve it. The economic system to some extent does what we want. After all, it's not like we've been total losers at steering the future of humanity. But the standard theory of the economic system assumes that consumers have full access to relevant information, that they take it into account, and that their decisions reveal their true preferences. However, monetary value is a one-dimensional measure that inevitably disregards the multi-valued reasons people have to invest money, and this projection on a one-dimensional scale means that information is lost.

Concretely, imagine how much more useful book reviews would be if you knew the reviewers' priorities compared to yours, if you knew what they consider a "good book". Imagine how much more useful sales numbers would be if companies knew how important economic and social engagement are for their customers. The economic system alone doesn't give us this information.

Moreover, emotions can capture problems that do not result in actions at all (are not "revealed"). Take the 2008 mortgage crisis as an example. If you read reports from back then, many people clearly felt something was wrong. "Something about that feels very wrong," a banker said, "It makes me sick to my stomach the kind of loans that we do," a mortgage broker was quoted with [6]. But these feelings didn't register in the system. Imagine we could have measured the tension in priorities between, say, keeping their job and acting morally right. This could have been an early warning sign. How many warning signs do we currently miss?

So, what we need for humans to interact intelligently on global scales is a simple and intuitive way for them to tell how much their priorities – long-term as well as short-term, locally as well as globally – match with decisions they can take. That might strike you as a very abstract idea. Let me tell you then where to start making it reality.

SCIENCE MATTERS.

Making information cheap does not make it correct, and as they say: Garbage in, garbage out. Information about the system is only useful to steer the future if it's accurate. There will generically be several information providers for the same correlation and their findings might disagree. This temporary disagreement is in the nature of research which brings us to the process of knowledge discovery and to the system that it operates with.

In the following I refer to the process of knowledge discovery as being executed by the academic system, by which I mean scientific research that is not conducted for profit. Scientific research is of course also conducted for profit, but I will not discuss this here because mixing in the economic system makes things more complicated without making them more insightful.

The academic system plays a pivotal role for establishing the feedback loop that allows our systems to integrate and process globally dispersed information. If we can close the feedback loop and the system can learn, all other problems are self-correcting. Thanks to the scientific method, we need not be afraid of conflicting information and uncertainties. These improve over the course of time, provided the process of knowledge discovery works as desired. Unfortunately, it presently doesn't. The reason is that the academic system too isn't able to learn. But do not despair, because we already know what to do. We need priority maps for scientists.

Personnel in administrative academic positions and scientists are faced today with many complex decision tasks. Everybody agrees that personal assessment of research projects and researchers is the best possible judgment. But not everybody can possibly assess everything and everybody. That is the scientists' problem of too costly information – it would take too much time to read all these papers. There is also other information about the system that scientists do not readily have access to. How good, for example, is the reputation of some university in a country you're not even sure where to find on a world map? Is this research area blooming or in decline? Is that a typical number of coauthors and publications in this field? How do I judge the enthusiasm of this referee?

Because of pure need scientists use whatever means are available to select the relevant pieces of information. Typical problems they face are finding relevant new publications or the most interesting candidates on lists with hundreds of applicants. Presently, they rely heavily on personal connections and some existing measures of scientific success. That is not good for many reasons. Personal connections are inevitably subjective and existing measures are too rough, too inflexible and also too streamlined.

Imagine how much a priority map could help scientists and how much time they would save. Their personal map would include their own research topics and judgment of these, interests in other fields of research and assessments of these, how relevant they believe certain traits to be for the scientific success of candidates, how relevant they think it is that a candidate's research topic matches the local research, and so on. You can add your wishlist here.

The information providers would be scientometric measures that indicate how research areas grow, co-authorship networks, co-citation links between research areas, and other existing measures about the connectivity and impact of scientific research.

These priority maps become truly useful when many people use and (partially) share them, for example their assessment of research results or students. But importantly the maps would be useful already for single users because much relevant information is publicly available. It is for example technically feasible to extract keywords from candidates' publication lists (or research plans) and match them with that of members of a committee. This would be a much better way to find potentially interesting candidates than looking for familiar names. This is just one concrete example of which I could list many, but this isn't the place for lots of details so let me return to the big picture.

The main point is that priority maps have the power to make a difference because they are designed to be useful for scientists. This is in contrast to existing measures for scientific success that are designed to be useful for administrative and external uses and are thus widely met with rejection and cynicism by researchers. The existing measures are used just because they are available and wide-spread, not necessarily because scientists think they are good. With the pressure of having to make a decision, saving time and energy scores higher than other, more idealistic, values. This conflict between individual short-term benefits and collective long-term benefits that turn into individual long-term benefits is the same tension we already discussed earlier. The root of the problem is the same: An accurate assessment of individual consequences from collective trends is time-intensive and requires too much personal investment.

Much effort has gone into devising better measures for scientific success. There is no lack of proposals, but none has caught on because scientists have no good reason to use them. On the other hand, there are many proposals for how to fix the academic system, normally top-down solutions designed to change individual incentives. In neither case however the feedback loop is being closed.

The priority maps close the loop. Now scientists can set their values for certain properties they find desirable of research project or candidates and then they can use their *individual* metrics for assessment. From what the scientists regard important will then naturally arise an aggregated measure that can be used externally. But note that this measure now will automatically adapt as scientists come to regard certain properties (say, number of publications in certain journals) more or less relevant. These measures, importantly, are also non-universal and naturally counteract streamlining. The system can now learn.

## A FIVE STEP PLAN.

Science matters for steering the future of humanity because it is the tool to obtain knowledge about the systems that govern our lives. Before scientists can use their knowledge to improve social, economic or politic systems, they have to solve their own problems; the academic system is thus the natural starting point. Roughly, the vision I discussed here can be realized in the following 5 steps:

### **1. Individual Priority Maps for Scientists.**

This is possible with presently existing technology. Scientists record their priorities for interesting research and its potential, for characteristics of good science and scientists. The information that their maps are matched to comes from the Science of Science,

Scientometrics and Bibliometrics and other sources of knowledge about knowledge discovery. Matches are visualized and can be manipulated to be inspected by the user.

**2. Close the feedback loop in the academic system.**

The priority maps for scientists will free time for research and make science more efficient. At the same time they provide information about the system itself that can be fed back into the system. Using this knowledge allows the system to learn and creates a naturally adaptive measure for scientific success. The academic system also serves as a case study that reveals difficulties when realizing this feedback loop.

**3. Individual priority maps for everybody.**

With the knowledge and, hopefully, the success story from the academic system one can now generalize priority maps for general purpose by adding social, political and personal values. Individuals generate their priority maps and institutions provide correlations.

**4. Close the feedback loop for social systems.**

Now we can feed back the knowledge about the system into the system and obtain more knowledge by recording the reaction to this feedback. Imagine how much social, political and economic discourse can be shortcut by this. Imagine the creative and engineering potential that is freed.

**5. Upgrade priority maps to brain extensions.**

In the long run, we should avoid using the visual cortex as pathway to display matches with priority maps. The potential of cheap information will be fully realized when information about our social systems is directly fed back into our brain and we can truly feel the consequences of certain decisions.

## SUMMARY.

I assume most people care about the future of the planet as I do, care as you did with this essay contest asking how humanity should steer the future. I assume most people want to solve our ecological, political and economic problems, and that we just have to make it easier for them to convert caring into action. I assume humans are intrinsically good and mean well, they just don't always get it right.

I may be naïve and I may be wrong. If in fact most people do not regard it relevant to get the plastic out of the oceans and to prevent children in the developing world from dehydration, then lowering the cost of information will not make a difference. However, in that case who am I to tell others that they have to share my values?

I have addressed here the question how humanity can steer the future. I can't tell you which course we will take if we enable our social systems to learn, but at least we will not be drifting any more.

## References

[1] [www.fhi.ox.ac.uk](http://www.fhi.ox.ac.uk), retrieved April 8, 2014

[2] [www.unglobalpulse.org](http://www.unglobalpulse.org), retrieved April 8, 2014

[3] [www.futurict.eu](http://www.futurict.eu), retrieved April 8, 2014

[4] [radicalism.milcord.com](http://radicalism.milcord.com), retrieved April 8, 2014

[5] C.C. Ballew & A. Todorov, PNAS 104, 17948-17953 (2007).

[6] [www.npr.org/templates/story/story.php?storyId=90327686](http://www.npr.org/templates/story/story.php?storyId=90327686), retrieved April 8, 2014