

## HOW SOCIETY WORKS<sup>1</sup> (Version 0.2) –

### Social order by self-organization

by Dirk Helbing

*The invention of laws and regulations is celebrated as great success principle of societies and they are, of course, important. However, a major part of social order is based on self-organization, which builds on simple social mechanism. These mechanisms have evolved over historical times and are the basis of the success or failure of civilizations. Currently, many people oppose globalization, because traditional social mechanisms fail to create cooperation and social order under globalized conditions that are increasingly characterized by homogeneous or random interactions. However, I will show that there are other social mechanisms such as reputation systems, which will work in a globalized world, too.*

Since the origin of human civilization, there is a continuous struggle between chaos and order. While chaos may stimulate creativity and innovation, order is needed to coordinate human action, to create synergy effects and more efficiency. Our success also critically depends on the ability to produce "collective goods," such as our transportation infrastructures, universities, schools, and theaters, as well as language and culture.

---

<sup>1</sup> Dear Reader,

thank you for your interest in this chapter, which is thought to stimulate debate.

What you are seeing here is work in progress, a chapter of a book on the emerging Digital Society I am currently writing. My plan was to elaborate and polish this further, before I share this with anybody else. However, I often feel that it is more important to share my thoughts with the public now than trying to perfect the book first while keeping my analysis and insights for myself in times requiring new ideas.

So, please apologize if this does not look 100% ready. Updates will follow. Your critical thoughts and constructive feedback are very welcome. You can reach me via [dhelbing@ethz.ch](mailto:dhelbing@ethz.ch) or @dirkhelbing at twitter.

I hope these materials can serve as a stepping stone towards mastering the challenges ahead of us and towards developing an open and participatory information infrastructure for the Digital Society of the 21st century that would enable everyone to take better informed decisions and more effective actions.

I believe that our society is heading towards a tipping point, and that this creates the opportunity for a better future. But it will take many of us to work it out. Let's do this together!

Thank you very much, I wish you an enjoyable reading,

Dirk Helbing

PS: Special thanks go to the FuturICT community.

According to Thomas Hobbes (1588-1679), civilization started with everyone fighting against everybody else ("homo hominis lupus"), and it required a strong state to create social order. Even today, the merits of civilization are highly vulnerable, as the outbreak of civil wars illustrates, or the breakdown of social order after natural disasters. But we are not only threatened by the outbreak of conflict. We are also suffering from serious cooperation problems, which often result from "social dilemma situations."

### *The challenge of cooperation*

To understand the nature of "social dilemmas," let us discuss the problem of producing "collective goods." For illustration, assume a situation in which you and others engage in creating and using something together (the "collective good"). Then, everyone invests a certain contribution that, in some way or another, goes into a "common pot." If the overall investment reaches a sufficient size, it creates a synergy effect and produces benefits. To reflect this, the total investment, i.e. the amount in the pot, is multiplied with a factor greater than 1. Finally, the resulting overall amount is assumed to be equally split among all contributors.

In such situations, you cannot benefit if you don't invest. In contrast, if everyone invests a sufficiently high amount, everyone gains, as the investment is assumed to create a synergy effect. But if you invest much and others little, you are likely to walk home with a loss. Conversely, a low investment may produce an individual advantage, as long as the others invest enough.

Therefore, cooperation (contributing much) is risky, and free-riding (contributing little) is tempting, which destabilizes cooperative behavior. If the situation occurs many times, cooperation erodes, and a so-called "tragedy of

the commons" results. In the worst case, nobody invests anymore in the end and nobody will get a benefit. This is pretty much the situation one can find in countries characterized by high levels of corruption and tax evasion.

Therefore, cooperative behavior, even though beneficial for everyone, can break down for similar reasons as free traffic flow breaks down on a crowded circular road: the desirable state of the system is unstable (see the movie at [http://www.youtube.com/watch?v=7wm-pZp\\_mi0](http://www.youtube.com/watch?v=7wm-pZp_mi0)). But given that it's possible to stabilize free traffic flows by means of traffic assistance systems, are there any biological or socio-economic mechanisms that can stabilize social cooperation?

*An outcome that is bad for everyone*

"Tragedies of the commons" are known from many areas of life. Often cited examples are the degradation of our environment, overfishing, the exploitation of social benefit systems, or dangerous climate change. In fact, even though probably nobody wants to destroy our planet, we still exploit its resources in a non-sustainable way and pollute our Earth. For example, a good solution how to safely keep nuclear waste has still not been found. Moreover, nobody wants to be responsible for the extinction of a species of fish, but we are facing a serious overfishing problem in many areas of the world. We also know that public schools, public hospitals, and many other useful investments made by the state require our taxes. Nevertheless, the problem of tax evasion is widespread.

Now, the reader will argue: of course, we can make contracts and establish institutions such as courts to ensure that they will be kept. This is correct, but what institutions shall we establish? How efficient are they? And what are their undesired side effects? In an attempt to address these questions, the following paragraphs will give a short and certainly incomplete overview of

mechanisms that support cooperation – and have played a major role in human history.

### *Family relations*

An early mechanism invented to promote cooperation in social dilemma situations, first discussed by George R. Price (1922-1975), is called "genetic favoritism." This means that, the closer you are genetically related with someone else, the more advantages you would give to him or her as compared to strangers. This principle has led to tribal structures and dynasties, which have been around for a long time – and still are in many countries. Today's inheritance law favors relatives, too. However, genetic favoritism has a number of undesirable side effects, such as a lack of fair opportunities for non-relatives, and also ethnic conflicts or blood revenge. The drama of *Romeo and Juliette* articulates the impermeability of family structures in the past very well. Another famous example is the Indian system of casts, which many people don't find timely anymore.

### *Scared by future "revenge"*

What other options do we have? Repeated interactions, as they already occurred in early human settlements, are an example. In case you interact with the same person time and again, you may play *tit for tat* or some other strategy that teaches your interaction partner that non-cooperative behavior won't pay off. The strategy is thousands of years old and is also known as "an eye for an eye, a tooth for a tooth."

The effectiveness of such revenge strategies was studied in 1981 by Robert Axelrod (\*1943) in his famous computer tournaments. It turned out that, if you

would interact with me just frequently enough, mutual cooperation would beat exploitation strategies. This effect is also known as "shadow of the future." As a consequence, a situation of "direct reciprocity" is expected to result, in which "I help you and you help me." But what if our friendship becomes too close, such that the result is good for us but bad for others? This could lead to corruption and impermeable, inefficient markets.

Moreover, what should we do if we interact with someone only once, for example, a renovation worker? Will we then have to live with a bad experience that we are likely to make? And what should we do, if a social dilemma situation involves many players? Then, a tit-for-tat strategy is too simple, because we don't know who cheated us and who didn't.

### *Costly punishment*

For such reasons, further social mechanisms have emerged, which include "altruistic punishment." As Ernst Fehr and Simon Gächter have shown in 2002, if players can punish others, this will promote cooperation, even when punishment is costly. This effect is so important that people, if faced with the option to choose between a world without punishment and a world offering a sanctioning option, often decide for the second one. The experimental discovery of this was made as late as 2006 by Özgür Gürek, Bernd Irlenbusch, and Bettina Rockenbach. Punishment efforts are mainly needed in the beginning, to teach people "to behave," i.e. to be cooperative. In the end, the punishment option may be rarely used. But for cooperation to persist, the punishment option still needs to be there (or another mechanisms which can stabilize cooperation).

Note that such mutual "peer" punishment is a widespread mechanism. In particular, it is used to stabilize social norms, i.e. behavioral rules. Every single one of us probably exercises peer punishment many times a day –

sometimes in a mild way (e.g. by raising eyebrows), and sometimes more furiously (e.g. when shouting at others).

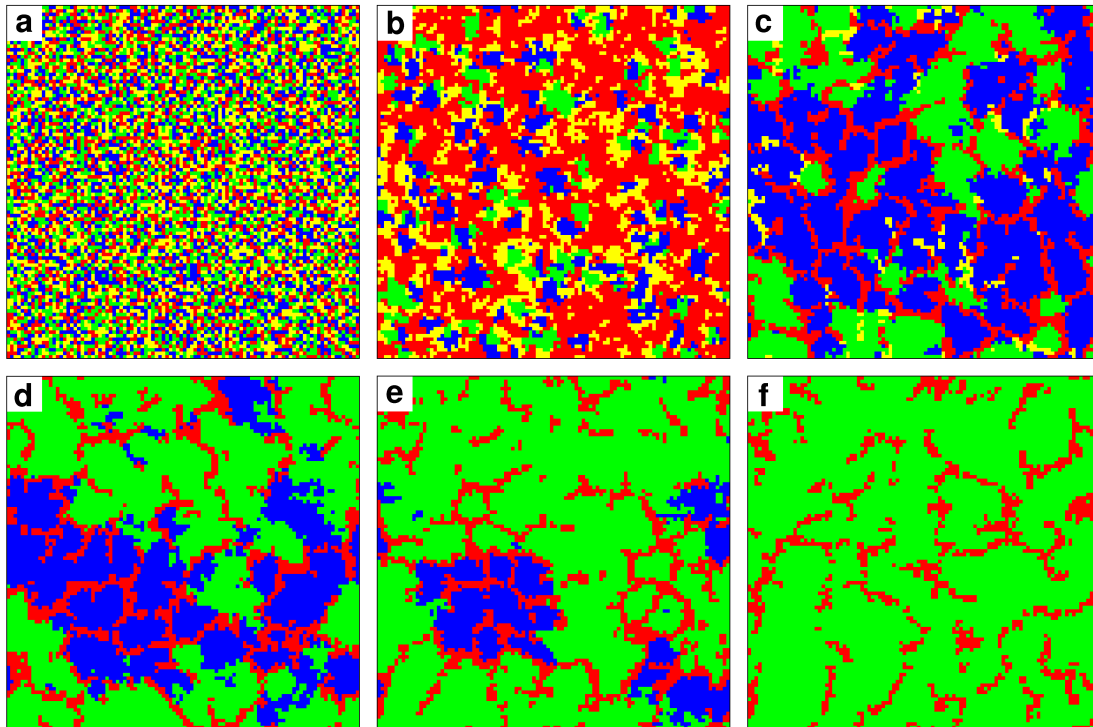
### *The birth of moral behavior*

But why do we punish others at all, if it reduces our own payoff, while others benefit from the resulting cooperation? This puzzle is called the "second-order free-rider dilemma," where "first-order free-riders" mean non-cooperators and "second-order free-riders" mean non-punishers.

To answer the above question, in a study with Attila Szolnoki, Matjaz Perc György Szabo in 2010, I have analyzed a collective goods problem considering four possible behaviors: (1) cooperators who don't punish non-cooperators, (2) cooperators punishing non-cooperators, called "moralists" (green), (3) non-cooperators punishing other non-cooperators, called "immoralists" because of their hypocritical behavior (yellow), and (4) non-cooperators who don't punish (red). We furthermore assumed that individuals imitate the best performing behavior played by their interaction partners.

If the interaction partners were randomly chosen from the entire population, moralists couldn't compete with cooperators – due to the additional punishment costs. Therefore, cooperators ended up in a social dilemma situation with non-cooperators, which they lost, and a "tragedy of the commons" occurred. However, if individuals interacted with a small number of neighbors, we found the emergence of clusters of people, in which the same behavior prevailed (see picture below). Surprisingly, the fact that "birds of a feather flock together" makes a big difference for the outcome of the interactions: it allows moralists to thrive. Assuming all other things to be equal, in the spatial (rather than random) interaction scenario, cooperators lost the battle with non-cooperators in their neighborhood, as expected, but moralists could cope with them (as their punishment of non-cooperators reduced the

success of non-cooperative behavior). Therefore, when individuals interacted with neighbors in geographical space or social networks rather than with random interaction partners, moral behavior could win its way and second-order free-riders were crowded out. As a consequence, "moral behavior," i.e. cooperation and the punishment of non-cooperative behavior thrived.



Nevertheless, punishing each other is often annoying or inefficient. This is one of the reasons why we may prefer so-called "pool punishment" over the above "peer punishment." In such a case, we invest into a common pot paying for a police or court or other institution sanctioning improper behavior. But a problem of this approach is that representatives of the punitive institution may be corrupted. Besides this corruption problem, it is often far from clear, who deserves to be punished and who not.

We must certainly be careful not to sanction the wrong people. This requires high inspection and discovery efforts, which might not always be justified by the success rates. Then, lack of success may lead to less inspection and

eventually to more crime – a strange interdependency pointed out by Heiko Rauhut, which might explain the crime cycles that have often been observed in the past. To our great surprise, the computer simulations of the spreading and fighting of criminal behavior that I performed in 2013 together with Karsten Donnay and Matjaz Perc suggest that more surveillance and greater punishment, i.e. more deterrence, are not able to eliminate crime. In fact, while there is an almost 10 times higher rate of prisoners in the USA than in Europe, crime rates do not seem to be lower. In fact, I expect a crime prevention strategy based on the consideration of socio-economic factors to be much more effective than one mainly based on deterrence.

### *Group selection and success-driven migration*

Things become even more tricky if we have several groups with different preferences, for example, due to different cultural backgrounds or education. The subject has become popular under the label of "group selection," which was promoted by Vero Coppner Wynne-Edwards (1906-1997) and others.

In fact, it seems that group competition can promote cooperation. Compare two groups: one with a high level and one with a low level of cooperation. Then, the more cooperative group is expected to get higher payoffs, such that it should grow more quickly compared to the non-cooperative group. Consequently, cooperation should spread and free-riding disappear. However, what would happen if there existed an exchange of people between the two groups? Then, free-riders could exploit the cooperative group and quickly undermine the cooperation in it.

### *The surprising role of migration*



Can cooperation only thrive in a world without migration and exchange? Surprisingly, the contrary is true if the conditions are right. In computer simulations I performed in 2008/09 together with Wenjian Yu, we have studied success-driven migration. Our simulations made the following assumptions: (1) Individuals move to the most favorable location within a certain radius around their current location ("success-driven migration"). (2) They tend to imitate the behavior of the most successful interaction partner (neighbor). (3) With a certain probability, the individual migrates to a free location, or it flips its behavior (from cooperative to non-cooperative or vice versa). Rule (1) does not change the number of cooperators, while all the other rules undermine significant levels of cooperation. Nevertheless, when all three rules are applied together, a surprisingly high level of cooperation emerges after a sufficiently long time. This is even true when the computer simulation is started with no cooperators at all, as Thomas Hobbes assumed to be the initial state of society. How is this possible, in spite of the fact that we don't assume a "Leviathan," here, i.e. a strong state, which imposes cooperation in a top-down way?

It turns out that migration disperses individuals in space. However, the rare flipping of individual behaviors creates a few cooperators. After a sufficiently long time, some cooperators are randomly located next to each other by sheer coincidence. In such a cooperative cluster, cooperation is rewarding, and neighboring individuals imitate this successful behavior. Afterwards, cooperation spreads quickly all over the system. Cooperative individuals move away from non-cooperative ones and join other cooperators to form clusters. Eventually, the system is dominated by large cooperative clusters, with a few non-cooperative individuals at their boundaries. Therefore, we find that individual behavior and social neighborhood co-evolve, and individual behavior is determined by the behavior in the neighborhood (the "milieu").

In conclusion, when people can freely move and live in the places they prefer, this can largely promote cooperation, given that they are susceptible to the local culture and that they quickly adapt to successful local behaviors. Thus,

migration isn't a problem, while lack of integration may be. But it takes two to be friends. Integration requires efforts on both sides, the migrants and the receiving society.

Migration is not always welcomed, even though it has always been part of human history. Many countries struggle with migration and integration. However, the United States of America, known as the "melting pot," are a good example for the positive potential of migration. This success is based on the principle that, in the USA, it is relatively easy to interact with strangers.

Another positive example is the Italian village Riace, in which the service sector was gradually disappearing, as young people were moving away to other places. But one day, a boat with migrants stranded. The mayor interpreted this as a divine sign, and decided to use this as an opportunity for his village. And, in fact, a miracle occurred. Thanks to the migrants, the village was revived. Since they were welcomed, they were grateful and gave a lot of good things back to the old inhabitants of the village. As the migrants were not treated as foreigners, but as part of the community, a trustful relationship could grow.

### *Common pool resource management*

Above, I have described many simple social mechanisms that can be and have been tested for their effectiveness and efficiency in laboratory settings. They are also known to play a role in reality. But what about more complex socio-economic systems? Would self-organization be able to create desirable and efficient solutions? This is what Elinor Ostrom (1933-2012) studied – and what she got the Nobel prize for. It is often claimed that public ("common pool") resources cannot be efficiently managed, and that's why they should be privatized.

Elinor Ostrom discovered that this argument is actually wrong. She studied the way in which Common Pool Resources (CPR) were managed in Switzerland and elsewhere, and found that self-governance works well, if only the interaction rules are suitably chosen (see [http://en.wikipedia.org/wiki/Elinor\\_Ostrom](http://en.wikipedia.org/wiki/Elinor_Ostrom)). One suitable set of rules that provide good conditions for successful self-governance is specified below:

1. There are clearly defined boundaries between in- and out-group parties (effectively excluding external, un-entitled parties).
2. Rules regarding the appropriation and provision of common resources exist that are adapted to the local conditions.
3. The collective choice arrangements allow most resource appropriators to participate in the decision-making process.
4. There is an effective monitoring by people who are part of or accountable to the appropriators.
5. A list of graduated sanctions is applied to resource appropriators who violate community rules.
6. Mechanisms of conflict resolution exist that are cheap and easy to access.
7. The self-governance of the community is recognized by higher-level authorities.
8. In the case of larger common-pool resources, the system is organized in the form of multiple layers of nested enterprises, with small local CPRs at the base level.

As it turns out, public goods can even be created under less restrictive conditions. This amazing fact can be observed in communities of volunteers – from Linux over Wikipedia, Open Streetmap, and StackOverflow to Zooniverse and many other forums.

*The problem of globalization*

One take home message of this chapter is that self-organization based on local interactions is at the heart of all societies in the world. Since the very beginning of ancient societies until today, a great deal of social order emerges in a bottom-up way, based on suitable interaction mechanisms and institutional settings. This approach is flexible, adaptive, resilient, effective and efficient.

The mechanisms enabling bottom-up self-organization discussed before promote the interaction of agents with mutually fitting behaviors: "birds of a feather flock together." The crucial question is: are these mechanisms also effective in a globalized world?

One may say that the process of globalization creates increasingly "well-mixed" interactions: ever more people or companies interact with each other, often in more or less anonymous or random ways. Such conditions, unfortunately, promote an erosion of cooperation and social order.

This undesirable effect is illustrated by a video accessible at <http://www.youtube.com/watch?v=TVExTATGvP0>. It illustrates a ring of "agents" (e.g. individuals or companies), each engaged with their neighbors in the creation of collective goods. The local interactions initially promote a high level of cooperation. But then we add more and more interaction links with other, randomly chosen agents in the system. While the additional links help to increase the level of cooperation in the beginning, cooperation soon starts to drop, as the connectivity increases, and it finally goes to zero. In other words, when too many agents interact with each other, a "tragedy of the commons" results, where everyone is suffering from a lack of cooperation. Data from Andrew Haldane from the Bank of England suggest that, for example, the financial meltdown in 2008 might have resulted from a hyper-connected banking network.

*Age of coercion or age of reputation?*

In fact, citizens and politicians all over the world have noticed that our economy and societies, and the foundations they are built on, are destabilizing. In an attempt to stabilize our socio-economic system, governments all over the world have undertaken attempts to establish social order in a top-down way, based on surveillance and powerful institutions such as armed police. However, this approach is destined to fail due to the high level of systemic complexity, as I have pointed out in a previous chapter. In fact, we are seeing a lot of evidence for this failure: signs of economic, social and political instability are almost everywhere. Will our globalized society just collapse and break into pieces, thereby re-establishing a decentralized organization? Or is there any chance to live in a globalized world, in which cooperation and social order are stable? Could we perhaps build something like an assistance system for cooperation?

In fact, it is known that reputation systems can promote cooperation through "indirect reciprocity." Here, the principle is that someone helps you, while you help somebody else. Reputation mechanisms support people and companies with compatible preferences and behaviors in finding each other. In a sense, such a reputation system can also serve as a kind of "social immune system," protecting us from harmful interactions.

*Pluralistic, community-driven reputation systems*

These days, reputation and recommender systems are quickly spreading all over the Web, which stresses their value. People can rate products, news, and comments, and they do! There must be a good reason that people undertake this effort. In fact, they get useful recommendations in exchange, as we know it from *amazon*, *ebay*, *tripadvisor* and many other platforms. As

Wojtek Przepiorka and others have found, such recommendations are beneficial not only for users, who tend to get a better service, but also for companies. A higher reputation allows them to sell products or services at a higher price.

But how should reputation systems be designed? It is certainly not good enough to leave it to a company to decide, what recommendations we get and how we see the world. This would promote manipulation and undermine the "wisdom of the crowd," leading to bad outcomes. It is, therefore, important that recommender systems do not reduce socio-diversity. In other words, we should be able to look at the world from our own perspective, based on our own values and quality criteria. Otherwise, according to Eli Pariser, we will end up in a "filter bubble," i.e. in a small subset of the information society that fits our taste. But we may lose our ability to communicate with others who have different points of views. In fact, some analysts think that the difficulty in finding political compromises between republicans and democrats in the USA is related to the fact that they use increasingly different concepts and words to talk about them. In a sense, they are living in different, largely separated worlds.

Therefore, reputation systems would have to become much more pluralistic. When users post ratings or comments on products, companies, news, pieces of information, and information sources, including people, it should be possible to assess not just the overall quality (as it is often done on a five-star scale or even just by a thumb up or down). The reputation system should support different quality dimensions such as the physical, chemical, biological, environmental, economic, technological, and social qualities. Such dimensions may include popularity, durability, sustainability, social factors, or how controversial something is.

Moreover, users should be able to choose from diverse information filters (such as the ones generating personalized recommendations), and to generate, share, and modify them. I want to call this approach "social filtering."

A simple system of this case has been implemented in Virtual Journal, <http://vijo.inn.ac/>. Then, we could have filters recommending us the latest news, the most controversial stories, the news that our friends are interested in, or a surprise filter. So, we could choose among a set of filters that we find useful. To consider credibility and relevance, the filters should also put a stronger weight on information sources we trust (e.g. the opinions of friends or family members), and neglect information sources we do not want to rely on (e.g. anonymous ratings). For this purpose, users should be able to rate information sources as well, i.e. other raters. Then, spammers would quickly lose their reputation and, with this, their influence on the recommendations made.

Altogether the system of personal information filters would establish an "information ecosystem," in which increasingly good filters would evolve by modification and selection, thereby steadily enhancing our ability to find meaningful information. Then, the pluralistic reputation values of companies and their products (e.g. insurance contracts or loan schemes) would give a pretty differentiated picture, which could also help the companies to develop better customized and more successful products. Hence, reputation systems can be good for both, customers and producers. Customers will get better offers, and producers can take a higher price for better quality. This is serving both sides.

In summary, the challenge of creating a universal, pluralistic reputation system might be imagined as transferring the principles, on which social order in a village is based, to the "global village," i.e. to the conditions of a globalized world. The underlying success principle is the matching of people or companies sharing compatible interests. A crucial question is, how to design reputation systems in a way that would make them resistant to manipulation and provide enough freedom for privacy and innovation? Information Box 1 offers some related ideas.

*Social Information Technologies*

Reputation systems are just one possibility to promote social order and favorable outcomes of social interactions. As we have seen before, many problems result when people or companies don't care about the impact of their decisions on others. This may deteriorate everyone's situation, as in "tragedies of the commons," or cause mutually damaging conflicts. How to overcome such problems? How to promote more responsible behaviors and sustainable systems? The classical approach to this is to invent, implement and enforce new legal regulations. But people don't like to be ruled, they can't handle many laws, and they often find ways around them. As a consequence, laws are often not very effective. Nevertheless, with Social Information Technologies it is possible to create a better world, based on local interactions and self-organization. It's actually easier than one might think.

Today, smartphones are increasingly becoming assistants to manage our lives. They guide us to find fitting products, nice restaurants, the right travel connection, and even a new partner. In the future, such personal assistants will be less and less focused on self-centered services. They will pay attention to the interactions between people and companies, and they will produce benefits for all involved parties and the environment, too. Note that, when two people or companies interact, there are just four possible outcomes, among them coordination failures and conflicts of interest.

In the first case, the interactions would be lossful for both sides, i.e. it would be a lose-lose situation. It is pretty clear what to do in such situations: one should avoid the interaction in the first place. For this to happen, we need information technologies that make us aware of the negative side effects of the interaction. Similarly, if we knew the social and environmental implications of our interactions on the environment, we could take better decisions. Measuring the externalities of our actions is, therefore, an important precondition for avoiding damage. In fact, if we had to pay for the externalities



caused by our decisions and actions, then individual and collective interests would become more aligned. As a consequence, we wouldn't run so easily into traps where individual decisions cause overall damage.

The second case is that of a bad win-lose situation, i.e. one side would have an advantage of the interaction, while the other side would suffer a disadvantage, and altogether the interaction would be damaging. In this situation, one side is interested in the interaction, but the other side would like to avoid it. Again, increasing awareness may help, but we would also need social mechanisms that would protect the potential loser from exploitation.

The third case concerns good win-lose situations. While the interaction would again be favorable for one side and unfavorable for the other, overall there would be a systemic benefit of that interaction. Consequently, one side would be interested in the interaction, but the other side would want to abstain from it. It is possible though to turn the win-lose situation into a win-win situation, namely by a value transfer. In this way, the interaction becomes profitable for both sides, which would hence engage in it.

Finally, in the fourth situation, the interaction would create a win-win situation. There are nevertheless two things one might do: balance the gains in a fair way, and create awareness of opportunities one would otherwise miss. In fact, every day, we are walking by hundreds people, who might share some interests with us, but we don't even know about it. These may be a hundreds of missed opportunities. If we had social information technologies helping us to interact with each other more successfully, this would unleash unimaginable social and economic potentials. If we had suitable tools to assist us, the large diversity of people with different cultural backgrounds and interests wouldn't be a problem anymore. Rather than producing conflict, diversity would increasingly turn into an opportunity.

In summary, Social Information Technologies will help us to avoid bad

interactions, to discover opportunities for good interactions, to engage in them successfully, and to turn bad interactions into good ones. In this way, coordination failures and conflicts can be considerably reduced. I, therefore, believe that Social Information Technologies could produce enormous value – be it material or immaterial. Just remember that *Facebook* is worth more than 50 billion dollars, even though it is based on a very simple principle: social networking. How much more valuable would Social Information Technologies be? But I don't want to argue for big business here. In fact, if we created these technologies in a crowd-sourced way for the public good or just for fun (as Linus Torvalds, the initiator of the Linux operating system, said) – even better!

*Towards distributed cybersecurity, based on self-organization*

Since the Arab Spring, governments all over the world have become worried about "twitter revolutions." Are social media destabilizing political systems? Do governments therefore have to censor free speech or influence at least the way tweets or facebook posts are distributed to followers? I don't think so. Biasing free speech will rather affect a society's ability to detect problems and address them early on.

But wouldn't a system based on the principle of distributed bottom-up self-organization be insecure? Not necessarily so! Let me give an example. One of the most astonishing complex systems in the world is our body's immune system. Even though we are bombarded every day by thousands of viruses, bacteria, and other harmful agents, our immune system is pretty good in protecting us for usually 5 to 10 decades.

Our immune system is probably more effective than any other protection system we know. And what is even more surprising: in contrast to our central nervous system, the immune system is organized in a decentralized way. This is not by chance. It is well known that decentralized systems tend to be more

resilient to disruptive events. While targeted attacks or point failures can make a centralized system fail, a decentralized system will usually survive the impact of attacks and recover. In fact, this is the reason for the robustness of the Internet. So, why don't we build information systems in ways that protect them by "digital immune systems"? This should also include a reputation system, which could be called a "social immune system."

### *Managing the chat room*

Information exchange and communication on the Web have quickly changed. In the beginning, there was almost no regulation in place. These were the days of the Wild Wild Web, and people often did not respect human dignity and the rights of companies when posting comments. However, one can see a gradual evolution of self-governance structures over time.

Early on, public comments in news forums were published without previous screening, and this spread a lot of low-quality content. Later, comments were increasingly assessed for their lawfulness (e.g. for respecting human dignity) before they went online. Then, it became possible to comment on comments. Now, comments are rated by readers, and good ones get pushed to the top. The next logical step is to rate commenters and to rate raters. Thus, we can see the evolution of a self-governing system that channels the free expression of speech into increasingly constructive paths. I, therefore, believe it is possible to reach a responsible use of the Internet mainly on the basis of self-organization.

In the end, the great majority of malicious behaviors will be handled by crowd-based mechanisms such as the reporting of inappropriate content and a reputation-based display of user-generated Web content. A small fraction will have to be taken care of by a "chat room master" or moderator, and there will be a hierarchy of complaint instances to handle the remaining, complicated cases. I expect that, only a few cases will have to be taken care of by courts

or other institutions, while most activities will be self-governed by social feedback loops in terms of sanctions and rewards by peers. In the following chapters, I will elaborate in more detail, how information technologies allow top-down and bottom-up principles, but also people and companies, to come together in entirely new ways.

#### INFORMATION BOX 1: Creating a trend to the better

For reputation systems to work well, there are a number of further things to consider: (1) the reputation system must be resistant to manipulation attempts; (2) people should not be terrorized by it, or by rumors; (3) to allow for more individual exploration and innovation than in a village, one would like to have the advantages of the greater freedoms of city life – but this requires sufficient options for anonymity or pseudonymity (to an extent that cannot challenge systemic stability).

First, to respect the right of informational self-determination, a person should be able to decide what kind of personal information (social, economic, health, intimate, or other kind of information) it makes accessible for what purpose, for what period of time, and to what circle (such as everyone, non-profit organizations, commercial companies, friends, family members, or just particular individuals). These settings would, then, allow selected others to access and decrypt selected personal information. Of course, one might also decide not to reveal any personal information at all. However, I expect that having a reputation for something will be better for most people than having none, if only to find fitting people who have similar preferences and tastes.

Second, people should be able to post the ratings or comments entered in the reputation system either in an anonymous, pseudonymous, or personally

identifiable way. But pseudonymous posts would have, for example, a 10 times higher weight than anonymous ones, and personal ones a 10 times higher weight than pseudonymous ones. Moreover, everyone who posts something would have to declare the *category* of information: is it a fact (potentially falsifiable and linked to evidence allowing to check it), an advertisement (if there is a personal benefit for posting it), or an opinion (any other information)? Ratings would always have the category "opinion" or "advertisement". If people use the wrong category or post false information, as identified and reported by, say, 10 others, the weight of their ratings (their "influence") would be reduced by a factor of 10 (of course, these values may be adjusted). All other ratings of the same person or pseudonym would be reduced by a factor of 2. This mechanism ensures that manipulation or cheating does not pay off.