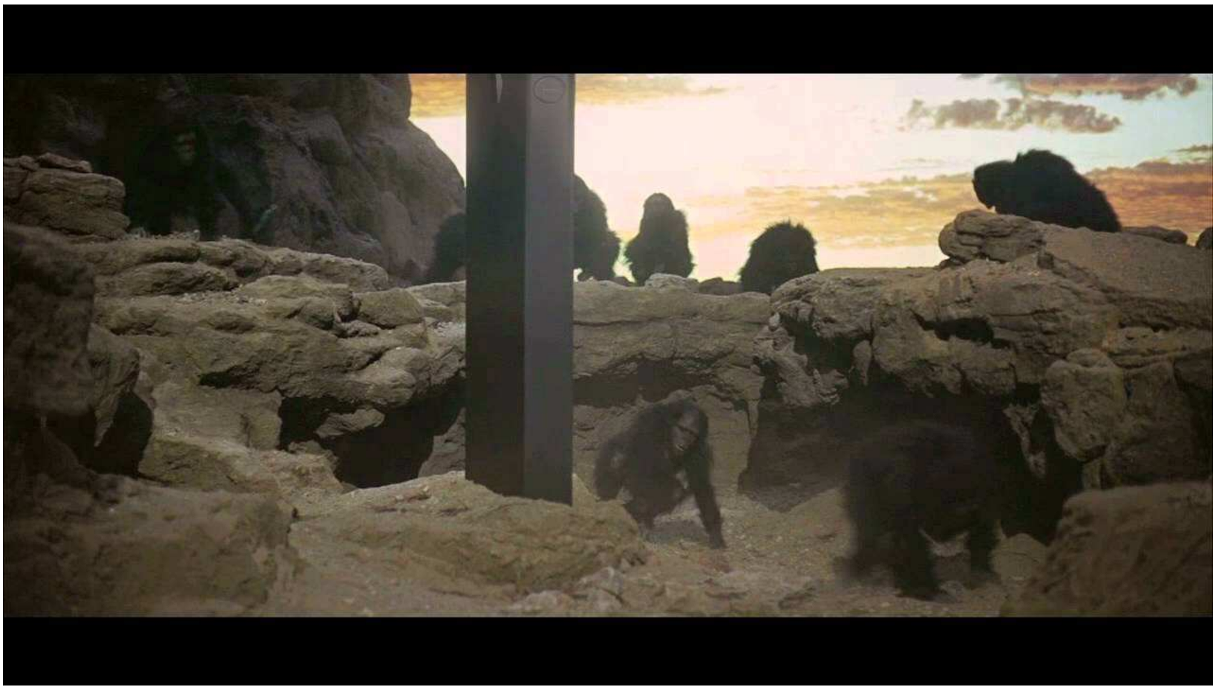
Mapping sociotechnical publics for responsible innovation

Udo Pesch



Contents

1. Responsible innovation

In this chapter we will first discusses some of the presumptions that are often held with regards to the role of technology in society and the way that responsibilities are organized in innovation processes. These presumptions conflict with empirical reality, leading to **moral discomfort** about the way responsibilities are distributed in the process of technology development. The notion of **Responsible Innovation** is introduced in academic literature as a framework that allows the satisfactory uptake of moral issues in technology development. This framework combines elements from the traditions of **technology assessment** (TA) and **engineering ethics**. Two approaches to responsible innovation are distinguished: first, an approach that aims to **include values in the design** of new technology; second, an approach that aims to present a framework for the **governance of innovation** processes. Both approaches are important, but do not yet present a clear-cut understanding of responsibility. To resolve this shortcoming we will develop an alternative approach of responsibility, which is based on the reciprocal relation of a professional actor and a certain 'public' that is able to assess the moral acceptability of the actions of that actor. A problem for innovation is that there is no pre-given public that can be connected to technology development. This implies that for responsible innovation we have to construct a **sociotechnical public** that is specific to a certain technology.

1.1 Why responsible innovation?

Koyaanisqatsi is a 1983 movie directed by Godfrey Reggio. The name of this highly interesting movie is a word from the language of the Hopi Indians, which means 'life out of balance'. The movie contrasts the world created by technology with the world created by nature. The film has no plot and no actors, it only shows images that are edited to fit the pulsating music of composer Philip Glass. We see accelerated shots of how city life is determined by technology, how whole neighborhoods are demolished by dynamite and how airplanes and rockets leave earth, the solid ground under our feet. Conversely we are shown slow motion images of nature in its imposing and harmonious beauty. The suggestion is clear, by creating our own environment by means of technology, we have created a monster that defines our rhythm and that has taken away our freedom.

In an interview, Reggio makes the following statement about his movie:

> What I try to show is that the main event today is not seen by those of us that live in it. We see the surface of the newspapers, the obviousness of conflict, of social injustice, of the market, the welling up of culture, but to me, the greatest event, the most important event of perhaps our entire history, nothing comparable in the past to this event, is fundamentally going unnoticed. And the event is the following: the transiting from all nature or the natural environment as our

host of life of human habitation into a technological milieu, into mass technology as the environment of life. So, these films have never been on the effect of technology, of industry on people, it's been that everyone, politics, education, the financial structure, the nation state structure, language, culture, the religions, all of that exists within the host of technology. So, it's not the effect of, it's that everything exists within. It's not that we use technology, we *live* technology. Technology has become as ubiquitous as the air that we breathe, so we're no longer conscious of its presence. What I decided to do when making this film was to rip out all the foreground of a traditional film, the foreground being the actors, the characterization, the plot, the story. I tried to take the background, all of that that just support it like wallpaper, move that up into the foreground, make that the subject, noble it with the virtues of portraiture, and make that the presence. So, we looked at traffic as the event, we looked at the organization of a city as the equivalent of what a computer chip looks like, we looked at acceleration and density as qualities of a way of life that is not seen and goes unquestioned. Life unquestioned is live lived in a religious state.[1]

Reggio is clearly uncomfortable with this omnipresence of technology, also because he assumes that technological development is an autonomous force which simply overcomes us. Following the work of the philosopher Jacques Ellul (1962), Reggio takes the theoretical perspective of *technological determinism*. Starting point of this perspective is that technological progress revolves around the attempt to look for the most efficient solution of a certain societal problem. As there can be only one solution that is most efficient, the eventual outcome of technological progress is that there is only one form of technology that eventually will be used by everyone. The idea that underlies this deterministic point of view is that technology is nothing more than applied science, in the sense that technological progress simply follows the advances in science.

Reggio and Ellul see such technological determinism as a threat, because it takes away human freedom. Other determinists are less afraid of technology as they expect that, as technology becomes better and better, it will improve the quality of life. In any case, the way we live, the way we communicate, the way we think about ourselves, the way we work, the way we relax, and so on and so on, they are all defined, mediated or made possible by technology. In that sense, we have come to live in a technological *milieu*. But does that also mean that technology is an autonomously developing thing? Policymakers, technology developers, the public-at-large all seem to adhere to a deterministic point of view, which is manifested in the belief that technology cannot be stopped, it can only be slowed down, but in the end it will become effectuated. Moreover, it is also believed that technologies cannot be shaped, so that decision making about a

[1] See: https://www.youtube.com/watch?v=_Mr26_m5rGQ.

technology can only be expressed as a yes or no issue, you are either in favor or against a certain technology, there is no middle ground.

However, technology is much more than the application of scientific findings, it is first and foremost the practical uptake of science-based expertise to solve certain societal problems by designing new artefacts and systems. What is reckoned to be a societal problem or a feasible solution is a question that is fundamentally open, subjected to interpretation and contingency. Technological developments are the outcome of a sequence of choices. In turn, these choices are made by a range of actors: engineers who made design choices in order to solve specific problems; companies, universities or state organizations who have engineers on their payroll that pursue the creation of wealth, knowledge or welfare; and consumers who have to choose what to buy and how to use certain products. In other words, technology is made by actual people (cf. Rip, 1988). The aggregate of all these choices may look like an overwhelming force, but in the end technology is the outcome of discrete human actions: technology is actively shaped by us, it is not a yes or no-phenomenon.

Nevertheless, we live in a technological milieu. Engineers and designers create the world which we live in, they build the world that we share. As such, they have a big responsibility. However, this responsibility is hardly recognizable in the way that engineers and designers deal with their jobs, nor in the way that responsibility is arranged in formal and informal structures. Engineers and designers are used to break down complex problems into manageable portions. It is all about drawing problem boundaries that allow a specific engineering solution. As a professional, the responsibility of an engineer is to come up with the right solutions for the posed problems. He or she is not appointed to question the origin of the problem, nor the application of the solution – these are, respectively, the delegated responsibilities of the client and of the end-user. In other words, there is a *functional distribution of responsibilities* that reduces the need of engineers and designers to take responsibility for aspects outside of their direct span of control (Van de Poel & Royakkers, 2011).

The traditional approach to this engineering morality is captured in the slogan of the National Rifle Association: ‘guns do not kill people, people do’. Still, Mikhail Kalashnikov, the inventor of the AK-47, the most lethal gun in history, stated just before he died that he was suffering ‘spiritual pain’ about the question whether he was responsible for the deaths by the weapons he created. The idea that responsibilities concerning technology development may be functionally distributed is based on the myth of determinism. Only in case of an autonomously developing technology you can say that designers do not share responsibility for the way that the results of their work is used. The remorse of Kalashnikov testifies the moral discomfort with the way responsibilities are distributed in technology development.

The fact that engineers share responsibility for the use of the technologies they have designed does not imply that engineers are to be blamed for each and every negative result of the application of their designs. After all, the use of a technology does not always comply with the original intensions of its designers and engineers (Herring & Roy, 2007). For instance, the Smart car was intended as a cheap and sustainable car. However, people that bought a Smart were people that already had a car, and used it for activities for which they previously did not use a car, such as shopping. It seems that that a technology partially becomes defined in its *use*, implying that also users bear responsibility for technology, also users have to carry a part of the moral burden of technology. With that, the moral ramifications of technology development become even more confounded.

But there is more: the feeling of moral discomfort grows stronger if we look at the *effects* of technology. The functional distribution of responsibilities could be upheld in a time that people widely agreed upon the positive results of technology development. However, over the last decades we have become increasingly aware of adverse effects of technologies. In 1963, Rachel Carson published a book called *A Silent Spring* in which she claimed that because of the use of DDT and other pesticides, all insects were killed so that birds no longer had any food, leading to a spring without the song of birds. The Apollo missions, which, on the one hand, were a main pamphlet of technological progress, showed, on the other hand, pictures of earth as an organic whole, as a unique, blue and green planet that seemed vulnerable as no one had realized before. The Vietnam war of the 1960s and 1970s was another blow for the belief in technology, as it showed that the most advanced army in the world could not win a war of a disorganized bunch of guerilla soldiers in a developing country. In 1973, the oil producing countries in the Middle East reduced their oil production because of the support of Western states for Israel during the Yom Kippur-war. This led to the so-called first oil crisis, and it showed our reliance on finite energy sources. We can consider the explosion of the Chernobyl nuclear reactor in 1986, which showed the potential devastation of invisible radiation, as the final blow for the belief in technology. Other salient adverse effects which have become manifest are for instance forms of air pollution, such as the hole in the ozone-layer, due to CFCs in cooling equipment, the emissions of particulate matter having severe negative health consequences, and also the emissions of greenhouse gasses such as CO_2, leading to climate change and all sorts of disruptive societal, economical, and political effects.

The introduction of new technologies might lead to new forms of use, that make it hard to predict the effects of new technologies, but also make it hard to do something about their negative effects once these have become manifest. This insight has given rise to the so-called *control dilemma*, which states that the possibility to steer a technology is greatest in the early stages of its introduction, whereas the awareness about the effects of this technology is greatest when the technology has become fully embedded in society (Collingridge, 1980).

Ulrich Beck (1992) speaks of the modern society as a *risk society*. He claims that in the course of history, technology has served the goal of controlling *natural* risks, such as earthquakes, floods, crop failure, and disease. It has done so with an enormous success, we have been able to fight the devastating forces of nature; the technological milieu that has been the effect of this success also brings about new risks, that are *created by the technologies themselves*. These technological risks prompt us to rethink the role responsibility of engineering practice – we cannot assume that the work of engineers and designers will, on the whole, yield positive results. Following all these considerations that lead to moral discomfort, Owen et al. (2013) state that we have to develop a new '*social contract for innovation'*, a new way to think about the way that engineers morally relate to their work and about the way that society deals with technology.

It is not just because we have a better understanding of the way that technology works that we need a new social contract for innovation. Also society has changed fundamentally in recent decades, bringing along the necessity to deal differently with technology. People have become better educated, more emancipated, and they have – thanks to social media – more capacity to express themselves towards the rest of society. In the words of Thomas Friedman (2006), the 'world has become flat': we have become subjects of a globalized economy, richer and more egalitarian than ever before. We are no longer bound by nationality, tribe or class, but we are individuals who are above all internationally oriented. The result of these social developments is that authority is not taken for granted any longer, government, companies, and experts are not trusted at face value, but the public-at-large demands to be included in the decision-making processes that concern them. Decisions on technology clearly belong to such processes, which explains the increased appeal for making innovation processes more inclusive and more democratic. The old dogma that engineers were situated outside of society and that their only task was to build something that is to be accepted or not by society cannot be upheld any longer, we have to realize that technology and innovation are fully embedded in society.

These social changes do not just lead to demands that are externally imposed upon engineers. It not intended to burden them with new constraints, on the contrary, it is meant to realize a 'better technology for a better society' (cf. Rip, Misa, & Schot, 1995). If technologies fail to take account of societal concerns, they are prone to fail, which can be illustrated by the many cases in which the introduction of new technologies has encountered societal resistance. Emerging fields like nanotechnology, genetic engineering, big data all face both high expectations and deep anxiety. The implementation of more developed technologies such shale gas extraction or the construction of wind turbines have led to widespread public resistance (Correljé, Cuppen, Dignum, Pesch, & Taebi, 2015; Dignum, Correljé, Cuppen, Pesch, & Taebi, 2015). New technologies also lead to discussions about values we hold dearly, such as privacy that is affected some way or the other by the rise of social network sites, drones, and surveillance cameras (Van den Hoven, 2014). Such societal opposition necessitates further reflection on the relationship between technology and society and their reciprocal impact. It urges us to anticipate on the responsiveness of new technologies to the beliefs, sentiments and values that are present in society.

1.2 Two approaches to responsible innovation

The notion of *responsible innovation* has quickly gained prominence in academic and policy circles, as is testified by an increasing range of book and journal publications, funding schemes, research projects, educational programs, etc. (Guston et al., 2014; Kuhlmann & Rip, 2014; Owen, Macnaghten, & Stilgoe, 2012; Van den Hoven, 2014).[2] The aspiration is that by having innovations and innovation processes that are more responsible, more ethically and socially acceptable technologies will be developed and a new social contract for innovation can be established.

According to Armin Grunwald (2014), the notion of responsible innovation builds forth on the framework of *Technology Assessment* (TA) with the addition of explicit ethical reflections that originate from the field of *engineering ethics*. TA emerged in the 1970s as an early-warning tool to prevent new technologies from having negative effects, but over the years this framework has been rearticulated in line with insights developed in *science and technology studies* (STS), a domain which researches scientific and technological processes as intrinsically socio-cultural events (Smits, Leyten, & Den Hertog, 1995). STS instructs us that technology cannot be seen as separated from society, but as a phenomenon that both shapes and is shaped by society. In turn,

[2] Academic literature and policy programs usually speak of 'responsible research and innovation'. As we will focus on issues related to technological development, we will restrict ourselves to the label of 'responsible innovation'. Surely, this does not mean that responsible research is not important nor that it is unrelated to responsible innovation.

this suggests that the assessment of a technology should be based on the active involvement of society.

The field of engineering ethics started with a particular interest in questions for the responsibility of engineers, for instance in the creation of nuclear weapons and products that were bad for the environment. Very much resembling the early stages of TA, this initial question for engineering responsibility assumed that technology developed in isolation from society and that moral values only came into play when technologies were applied. However, also in engineering ethics awareness grew that this is not the case: technologies are not value-neutral, but values are intrinsically embedded in them (Verbeek, 2006; Winner, 1980). With this awareness, questions for responsibility have become accompanied by questions for the role of values in engineering activities and technological design.

In the literature we may distinguish two approaches to the combination of TA and engineering ethics ranked under the label of responsible innovation. On the one hand, we can see authors reasoning in line with engineering ethics focusing on the question how to design values into the technological artefact or system (Taebi, Correljé, Cuppen, Dignum, & Pesch, 2014; Von Schomberg, 2013). On the other hand, there are authors who take STS as a point of departure to develop a framework for the governance of innovation (Stilgoe, Owen, & Macnaghten, 2013). Authors taking this second approach shy away from the focus on the values that are served or affected by the new technology as this often tends to become a form of 'speculative ethics', in the sense that claims are made about the future technology that are based on mere guesswork, ignoring the fundamentally uncertain characteristics of a technology and its future use (Nordmann, 2014; Nordmann & Rip, 2009; Owen et al., 2012; Stilgoe et al., 2013). It is emphasized that technology is always 'work-in-progress' (Stilgoe, 2015, p. 124), not only because the technology itself will be subjected to revision, but also, and predominantly so, because the patterns of interaction between society and technology are always evolving.

Partly, these differences of approach return in the scope of technologies that are focused upon. Researchers that look at the values that are related to a technology tend to address rather concrete domains of engineering, such as ICT, shale gas or chicken husbandry systems (e.g. Dignum et al., 2015; Van de Poel, 2014; Van den Hoven, 2014), while the scholars that emphasize the process of responsible innovation especially look at technologies that are not yet so developed, such as geoengineering or nanotechnology (e.g. Nordmann, 2014; Stilgoe, 2015), making it quite hard to talk about specific design choices. In spite of all these differences, the insights and reflections from both approaches should be taken at heart in any innovation process.

Bringing society into technological development means that the *values of the public* should be attended as much as possible, while at the same time we need a framework that allows the *governance of innovation*. Basically both approaches deal with values, either values that are specified with regards to a certain technology or values that are generic and have to be taken into account in any innovation process. The category of generic values are identified as follows by Owen et al., (2013, pp. 38-39) who claim that responsible innovation entails a collective and continuous commitment to be:

> **Anticipatory** – describing and analyzing those intended and potentially unintended impacts that might arise, be these economic, social, environmental, or otherwise. Alternative pathways should also be explored, for instance by asking scientists and innovators "what if . . . " and "what else might it do?" questions. Future explorations are not intended to come to claims that are plausible, as such methods do not aim to predict, but they should be useful as a space to surface issues and explore possible impacts and implications that may otherwise remain uncovered and little discussed. They serve as a useful entry point for reflection on the purposes, promises, and possible impacts of innovation.
>
> **Reflective** – reflecting on underlying purposes, motivations, and potential impacts, what is known and what is not known; associated uncertainties, risks, areas of ignorance, assumptions, questions, and dilemmas.
>
> **Deliberative** – inclusively *opening up* visions, purposes, questions, and dilemmas to broad, collective deliberation through processes of dialogue, engagement, and debate, inviting and listening to wider perspectives from publics and diverse stakeholders. This allows the introduction of a broad range of perspectives to reframe issues and the identification of areas of potential contestation.
>
> **Responsive** – using this collective process of reflexivity to both set the direction and influence the subsequent trajectory and pace of innovation, through effective mechanisms of participatory and anticipatory governance. This should be an iterative, inclusive, and open process of adaptive learning, with dynamic capability.

These generic values, or *four dimensions of responsible innovation* as they are labeled by Owen et al., are not just criteria that can be taken into account, they can also be seen as guiding principles that have to be taken into account when working on a new technology. In that sense, they figure as codes of conduct that encourage an engineer to take a holistic account of the effects of a technology and reflect on a wide range of public values at stake.

With that, the dimensions of responsible innovation are an extension of professional codes of conduct that have been presented in engineering ethics as conventions agreed upon between professionals that enable engineers to cooperate in serving a shared ideal (Davis, 1991). Codes of conduct help engineers to see and reflect upon the moral impact of their work and, with that, they strengthen their capacity to act responsibly. Still, additional ways to enhance the ability of engineers to develop responsibility have to be found. In many cases, codes of ethics appear to be too weakly institutionalized to figure as regulative frameworks and they also are formulated at too high a level of generality to be applicable in concrete situations (Pritchard, 2009; Van de Poel & Royakkers, 2011). Moreover, codified norms may complicate interpretation in case of the trade-off between incommensurable values instead of facilitating such interpretation (Pesch, 2008). In the next section, we will develop a more comprehensive account of responsibility that may be used by engineers and other technology developers to build a reflective space that underpins the quest for responsible innovations.

1.3 Responsibility, institutions and sociotechnical publics

Though responsible innovation is a topic that receives a lot of attention in scholarly literature and research programs, it still remains unclear what responsibility actually means (Grinbaum & Groves, 2013). In many respects, responsibility is a contested concept that gives rise to many different interpretations. Here, we will propose an understanding of responsibility here that helps to get a more clear-cut idea of how innovation, innovators and society at large relate to each other in moral terms.[3] In this understanding, responsibility is all about individuals making decisions while taking account of the normative evaluations of specific societal contexts. This account is primarily based on a distinction between active and passive responsibility (Bovens, 1998), and the way these are connected to each other.

Active responsibility pertains to a prospective connection between an actor and his or her actions and the moral character of that person (Minogue, 1963; Pesch, 2005, 2015a). Being

[3] This account of responsibility originates in virtue ethics (which finds its basis in the work of Aristoteles) and also uses elements of the work of John Dewey Dewey, J. (1922). *Human nature and conduct*: Courier Corporation. Virtue ethics is not a dominant stream in ethics, like utilitarianism and deontological ethics. These two approaches primarily focus on the structure of moral reasoning, the social context is usually out of consideration (See: Van de Poel, I., & Royakkers, L. (2011). Ethics, technology and engineering. Oxford: Blackwell.)

responsible for something means that you have reflected upon the consequences of an action before you have made that action. You think about what an action will lead to and then decide what kind of action you are going to take. While active responsibility is essentially forward-looking, *passive responsibility* deals with questions only to be asked *after* an undesirable event has occurred. Passive responsibility, which can also be defined as *accountability*, is all about the normative assessment of an action that has ended. To be held accountable entails that one accounts for a specific action in front of an *external body*, a certain audience or a *public*: a person or a group of persons that watch over the outcomes of an action.

Essential for our account on responsibility is the *connection* between active and passive responsibility. The assumption is that, in most cases, active responsibility presupposes some external party, a public, that assesses the outcomes of actions. Such an external party may be an empirical phenomenon, such as the student that will reflect upon his or her answers because there will be a teacher who assesses the quality of the work. It may also be an outer-worldly phenomenon, like the biblical God who will assess your deeds during the final judgment. The issue is not whether something is real or not, but whether we have *internalized* the normative evaluation of this external party – in other words, whether we *think* it is real. This may sound abstract, but it basically says that we continuously think about what others will say about what we do. So, what will the neighbors think of me if I come home drunk, what will my mother say when I get a tattoo, what will my deceased grandfather think of me when I graduate, what will my girlfriend say if I befriend my ex on Facebook, what will the police do if they catch me cycling without light, and so on? In a way, our conscience is nothing more than the internalized representation of the full set of publics.

Passive responsibility is all about punishment. If you do not make your exam well, you'll get a low grade. If your drive your bike without light, you might get fined. Also in case of active responsibility, we get punished. The difference is that we punish ourselves: whenever we think we do something wrong in the eyes of some imagined party, we feel remorse (Nieuwenburg, 2004), we actually hurt ourselves when we do something wrong. It is this propensity to feel pain, to feel bad about yourself, that characterizes active responsibility.

In sum, we can say that active responsibility is based on the presence of certain *publics*. In many cases such a public take on the shape of *accountability structures*, existing social contexts that warrant a reciprocal relation between the individual's moral intuition and the specific public that characterizes that social context (Pesch, 2005, 2008, 2015a; Van Gunsteren, 1994). The presence of such an accountability structure enables people to learn what counts as good behavior: they are punished for doing things wrong and rewarded for doing things right. By internalizing the rules that relate to a certain social context, individuals develop intuitions about which behavior is good or bad. The paradigmatic case of an accountability structure is a court of justice, which determines to what extent a suspect is culpable for a misdemeanor.

Relating this account of responsibility to technology development prompts us to look at which social contexts professional engineers have or should have internalized as the background of their moral reasoning. What are the publics that are relevant for innovation? To answer this question, we will first make a detour to the way that in modernity – which is the era that started in seventeenth century Europe – formalized social contexts have emerged that can be said to figure as comprehensive accountability structures. These formalized social contexts, which can be labeled as *institutional domains*, form the stage upon which all professional and official activities take place. Each of the domains of state, market, and science fulfills a specific role in society (Dahl & Lindblom, 1963; Pesch, 2014a, 2014b; Polanyi, 2001), and each of them comes with a specific 'public' that assesses the moral acceptability of actions taken in these domains, which we can respectively the electorate, consumers and scientific peers. These publics that can be seen as spectators looking at activities performed by actors from the institutional realms as spectators, continuously deciding on the desirability of such performances (cf. Fitzgerald, 2015; Green, 2010).

The domain of the state includes bodies such as government, parliament, and law. In liberal democracies, this domain is designed to enable the establishment and maintenance of collectively binding decisions. Opposite to this public domain is the private domain of the market in which actors try to maximize their personal utility. The state and the market relate to each other in a dichotomous manner: they are both mutually exclusive and complementary. Other than the organization of society (state) or the distribution of welfare (market), scientists deal with the reliability of truth claims so that usable answers about the functioning of the world can be constructed.

The institutional domains are subjected to the scrutiny of their publics: political actors, market suppliers, and scientists have to be prepared to expose their results to an impersonal audience which assesses their desirability. The presence of public scrutiny establishes a feedback loop between individual action and its acceptability in a wider societal context. If values are transgressed by a certain action, the public that constitutes a certain societal context can immediately express its discontent. More concretely, if parliamentarians fail to comply with the wishes of the electorate, they will be voted out of parliament during the next elections (Montesquieu, 2002). In the domain of the market, actors are held accountable mainly by the structure of the market itself. The presence of competition guarantees the functioning of the price mechanism, which, in turn, forces suppliers of products to stick to prices that consumers are willing to pay (Dumont, 1977; Smith, 1998). The domain of science is predominantly characterized by the presence of *explicitness*. Truth claims have to be made an object of external assessment, for instance in the form of the blind peer review system (Merton, 1979).[4]

With that, institutional domains can be seen as *rule-systems* that provide the rules as well as the mechanisms with which these rules are maintained. The double role of institutional domains with regards to rules allows not only the maintenance of values that are held to be important in that domain, it also allows the articulation of values that are specified in reaction to new societal developments.

Institutional domains are a specific case of *institutions*, which can be essentially defined as "rules, enforcement characteristics of rules, and norms of behavior that structure repeated interaction" (North, 1989, p. 1321). Institutions are social contexts which are based on recurring behavior: when certain actions are repeated often, they come to form the self-evident structure of a certain context and they will come to enable and constrain our activities and decisions. They coordinate our social behavior ranging from the very mundane, like saying 'good morning' to your fellow students, to the highly sophisticated, like the activities undertaken in an institutional domain.

Institutions as rule systems are societal contexts that are based on a *shared understanding*: whoever makes part of an institution knows the rules, knows how to behave, knows what is expected of him or her. At the same time, an institution also conveys a set of shared understandings, they are 'webs of significance' that are spun in social interactions (Berger & Luckmann, 1991; Geertz, 1973; Weber, 1972). It is via interaction with other people that one learns to give *meaning* to specific phenomena, it allows people to communicate with each other, and to develop a mental representation of the world that a community can understand. Meanings that are conveyed by institutional contexts determine the way people see the world, and with that they also determine the actions of these people.

The institutional domains of state, market and science are typical cases of institutions, as they are, like few other institutions, 'designed' to figure as accountability structures: they are social contexts in which personalized power relationships are replaced by universalistic and objective

[4] This requirement if explicitness is accompanied by the requirement of reproducibility. Any claim made in a an academic paper has to be supported by the data, a reference or an argument that can be reproduced by the public of readers.

rules, so that no individual is bound by external authority (Weber, 1972).[5] The design of these domains is based on the democratic principle that there is no authoritative ruler that decides what to do, what to buy and what to think. In an institutional domain, every individual is held accountable by a public of equals. By the construction of institutional domains, a societal system could be created that guaranteed individuals to be capable of following their own preferred way of living (Berlin, 2002). This is the basis of the 'social contracts' we have as individuals with the authorities that represent the realms of the state, the market or in science. We, as publics, check whether the functioning of these authorities is agreeable, if not we will vote politicians out of government or parliament, we will make companies go bankrupt, and the articles of scientists will not be published. Earlier, we have seen the plea for a new social contract for innovation, which suggests that processes of technology development are not any longer satisfactorily covered by the prevalent constellation of institutional domains. Or, in other words, there is no designated public that is able to check the moral content of innovation.

This means that we have to be on the lookout for a public or for publics that allow the assessment of innovation, and we have to think about how these publics can be made part of the reflections of technology developers. This urges us to think about possible institutional arrangements, methods, tools and procedures that allow a connection between technology developers and the relevant publics, and that allow the deliberate uptake of the values that these publics find important. In short, to have responsible innovation, we have to find out how technological developments relate to different publics and their values. Later in this paper it will be shown that there is no pre-given public that can be connected to a technology: there is no fixed institutional domain that can be linked to innovation. The upshot of this situation is that for every new technology we have to look for the selection of actors that are affected by it, and relate or involve these actors in the process of technology development somehow. Such a group of actors will be called a *sociotechnical public*.

As a sociotechnical public is case-dependent, in the sense that it only pertains to a singular innovation process, it is really hard to assess the moral load of a technology: we cannot simply turn to a pre-existing institutional domain and we never know the 'real' composition of the group(s) of people that are affected by a new technology. We can only do so by reconstructing a *public-by-proxy* that gives us some idea about how to assess a new technology. Here, we will develop an approach to make such a public-by-proxy upon the basis of insights from theory on technology dynamics. We will try to identify and involve a suitable set of actors, for which we will

[5] In historical reality, there is not someone who has actually designed institutional domains as accountability structures. This is merely a reconstruction, a label that can be attached only in retrospect.

rely on the framework of *Constructive Technology Assessment*. For distilling the right set of values that are to be included in the design of the technology, we will turn to the framework of *Value Sensitive Design*. This combination of frameworks will give rise to the *sociotechnical value map*, a research tool that allows students to identify a sociotechnical public and its values in case of a particular innovation process.

Questions about this chapter:

- What is the perspective of technological determinism and what are the flaws of this perspective?
- Why is there moral discontent about innovation?
- What are the approaches that can be recognized in the idea of responsible innovation?
- What is technology assessment and how has it developed over time?
- What are the four dimensions of responsible innovation and explain how these can be used?
- Describe active and passive responsibility and explain how these two forms or responsibility are related to each other in society?
- What are the main institutional domains in society and how can these figure as accountability structures?
- What is a sociotechnical public and why does such a public has to be constructed in order to pursue responsible innovation?

2. The dynamics of technology

This chapter will present the main notions that have to be taken into account when thinking about the development of technology. We will present technology development as a process that is **socially constructed**: different social groups have different understandings of a new technology, and technology development is the outcome over the political struggle over the 'right' understanding. At the same time, technological development also depends upon the linkages with existing technologies, they are part of **technical systems** that structure the problem definitions of the actors involved in technology development. The theory of evolutionary economics gives a comprehensive framework that describes the development of technology as a **confrontation** between a **variation environment** that produces alternative designs and **a selection environment** that selects a number of these designs. If the variation and selection environment are interlinked, we may speak of a **sociotechnical regime**, that if it is too strong, may lead to **lock-in** situations, that are the result of **positive feedbacks**. The notion of regimes emphasizes the difference between the 'hardware' and the 'software' of a technology. With regards to the hardware, we can say that the physical qualities of technological artefacts make it function as a rule system, which may be illustrated by using the notions of **scripts** and **affordances**. The software-side of technology is especially relevant when talking about future technologies, these are discursive entities that are constituted by the **promises** and **expectations** that are forwarded in certain social contexts. With regards to these contexts, we can say that modern technology is usually produced in **innovation systems**, networks of actors from universities, industry and government.

2.1 The social construction of technology and technical systems

Instead of seeing technology as a self-propelling process, which is the starting point of the deterministic perspective, empirical research shows that technology is *socially constructed*. Technology development is a human enterprise and it is based on human choices. Technology revolves around the solutions of certain societal problems, but, unlike determinists assume, there is no pre-given consensus about what the nature of such a problem really is. Different social groups will have different understandings and definitions of a problem, these groups will try to push their own particular problem definition as the driving principle of a technological design (Bijker, Hughes, Pinch, & Douglas, 2012).

This approach of the social construction of technology (or SCOT) can be illustrated by the development of the bicycle (Bijker, 1997). When the pedal-driven bicycle was invented halfway the nineteenth century, it became embraced by young males who wanted to be tough and cool. The main purpose of the bike was to be fast. To make the bike faster, the size of the front wheel had to become bigger and bigger, until it developed into the Penny Farthing which had a front wheel with a diameter of 1.5 meter. The additional benefit of this bike was that it was dangerous and uncomfortable, a good thing if you're trying to act cool. However, this type of bike also raised opposition. Elderly people were annoyed because they were ran over by young guys, women also wanted to be capable of riding a bike, which the design at that time didn't allow them to do. Gradually, you see new regulations and new forms of design that made the bike much more suitable for these other social groups. The front wheel became smaller, the bikes became more comfortable and safe by having breaks, air-filled tires, and chain propulsion. In the end, this led to the dominant design of the bicycle which we still use today, while the Penny Farthings have become part of history.

The presence of interpretative flexibility and of social groups which dispose over different degrees of power deeply influences what becomes of a technology in the end. This suggests that studying technology dynamics becomes the identification of social groups, their problem definitions and their societal power, and that the design of a technology is nothing more than the outcome of the function of these variables. However, this is not the full picture. Societal actors are not unbounded in their capacity to influence the development of a technology. This is because of the *situated character* of technological development (Rip, 1995): the success of a new technology depends on the linkages it can make between existing technologies, infrastructures, institutions and user experiences. Any technology comes with a tight fabric of interdependencies with society and other technologies.

To find out how the situated character of technology relates to socially constructed problem definitions we turn to the notion of *technical systems* as was introduced by Thomas Hughes (1983, 1987). This notion takes the interdependence of technologies as its starting point. Technical systems are especially recognizable in network technologies, such as electricity systems, infrastructures and ICT. These systems are based on the connection between different components that are geared towards a common goal. Because of these connections, the technologies that are part of the system have a strong influence on each other, but also on the system as a whole. In fact, specific technologies or social factors may hamper the further growth of a sociotechnical system, as they lag behind the development of the full system. Such lagging parts are defined as a *reverse salient*, a term that stems from military history in which it is used to signify a section of the front line that lags behind. Often, military people will direct all their efforts towards fixing such a reverse salient. The same goes for the development of sociotechnical system: inventors, engineers, entrepreneurs and others direct their creative and constructive forces mainly at the correction of these reverse salients by redefining them into solvable social or technical challenges – so called *critical problems*. By translating obstructions into a series of solvable problems, the sociotechnical system can grow again.

Hughes's approach seems especially suitable for network technologies that are centrally controlled, as such control allows the directed optimization of the system as a whole. However, as Karel Mulder and Marjolein Knot (2001) show, the systems theory of Hughes is also applicable in forms of industrial production that involve independent corporations. Mulder en Knot describe the history of PVC plastic development as a series of problems that are brought forwards by certain social groups – in case of PVC, these problems mainly pertained to the toxicity of the material – that were taken up by the industry and translated into new products, but also in new ways to tackle the toxic nature of PVC, for instance by developing new forms of waste management.

In other words, the situated character of technologies may trigger specific problem definitions which might give rise to the formulation of new solvable challenges that are taken by a technology developers as the way to move forward. They create a shared understanding of what the problem at hand is and what viable solutions may be. The development of a technology may thus be seen as a *trajectory* in which specific social groups come up with problem definitions that are contingent on an existing set of technologies. In turn, these problem definitions will be translated into new technological designs and new social arrangements that are aimed to resolve the problems at hand. In our analysis of the history of a technology, we may observe a series of forks: moments in which a specific development of the technology is further taken up at the extent of alternative approaches.

2.2 Evolutionary theory and regimes

The trajectory of a technology involves a sequence of confrontations between technology developers and society. This sequence of confrontations can be also be described in terms of *variation* and *selection*, like evolutionary processes in biology, as is done in the theory of evolutionary economics (Dosi & Nelson, 1994). Evolutionary economics is grounded in the work of Joseph Schumpeter (1942/2000) who said that innovation in a market society is a form of 'creative destruction'. The market is a system in which entrepreneurs have to develop new technologies to gain a business advantage over their competitors. According to Dosi and Winter (1994), the aggregate outcome of these entrepreneurial activities resemble an evolutionary process: the entrepreneurs are part of a *variation environment* which further is composed out of actors such as engineers, companies, and so on, who develop variations of new technological designs in order to become successful. On the other side, we have the *selection environment*, composed out of consumers, regulators, and so on, who decide which of the technological variations are chosen, and as such decide which alternatives eventually become successful.

Earlier is has been shown that technologies do not develop autonomously, but that they are characterized by technological and social interdependencies. The evolutionary perspective on technological development reveals that these interdependencies are the result of how the variation and selection environment interact and of the ideas that actors from both environments have about their respective counterparts. Producers of new technologies do not just make new things, they only make things which they *expect* to be successful. Buyers of new products do not just buy new products, but they will only buy products which they *believe* will work. In other words, there are cognitive representations that figure as feedback loops between the variation and selection environment that to a large extent determine the course of technology development. Unlike biological evolution, technological evolution is not a blind process that works via trial-and-error – technologies are chosen that are believed to become successful. Actors make their choices based on convictions, experiences, expectations that guide them along their variation or selection activities. In many cases, choices that have been made earlier become self-reinforcing: actors get accustomed to certain technologies and as such they will continue make the same choices. Such repeated choices can be seen as *rules* that come to allow and constrain actors to make certain decisions. The constellation of rules that are current with regards to a certain technology is called a *sociotechnical regime*:

> A technological regime is the rule-set or grammar embedded in a complex of engineering practices, production process technologies, product characteristics, skills and procedures, ways of handling relevant artefacts and persons, ways of defining problems; all of them embedded in institutions and infrastructures (Rip & Kemp, 1998, p. 338).

Regimes work as rule systems, with that they are institutions *par excellence*. The original articulation of the idea of sociotechnical regimes emphasized their functioning as mental representations that brought forwards the rules that structured the activities of engineers (Nelson & Winter, 1977). However, regimes entail more than the cognitive apparatus of engineers, as the introduction of a new technology always takes place against the backdrop of existing regimes (Rip, 1995, p. 420). New technologies have to be embedded in an existing regime, and with that these regimes are at the same time reproduced by these new technologies. Regimes are furthermore developed against the background of so-called 'sociotechnical landscapes. According to Rip and Kemp (1998), such a landscape pertains to factors that create the boundaries that enable and constrain (individual) processes of decision-making. This understanding of a landscape is a metaphor that refers to a geographical structure that restrains the scope of human activity. Following this metaphor, the level of landscape may be seen as the collection of immovable or otherwise unchangeable phenomena that have to be taken into account in order to engage in action. The notion of landscapes emphasizes the fact that technological development takes place in the context of long-term, stable processes in areas such as demography, economics, environmental degradation, and so on. Besides taking account of the most common interpretation of landscape, which emphasizes long-term, stable social developments, we should also be on the lookout for singular events, such as disasters, media hypes, and so on (Kemp & Van den Bosch, 2006). The impact of these singular events may be quite big and have unanticipated consequences, whereas long-term trends are, by definition, more predictable (Pesch, 2015b).

2.3 Positive feedbacks

The presence of a regime establishes a repetition of similar connections between the variation- and the selection-environment. This makes technological development a process that is basically *path-dependent*: once a certain design choice is taken or once a certain user practice has become prevalent, it becomes almost impossible to change. Alternative solutions and practices become virtually unfeasible, as is paradigmatically illustrated by the dominance of the QWERTY-keyboard, which is in fact the slowest configuration of keys possible. This slowness was deliberately pursued because otherwise the hammers of the nineteenth century typewriter would get jammed. Obviously modern keyboards allow far greater typing speed, but because everybody is used to the QWERTY-keyboard, more efficient designs have not been successful (David, 1985). We can speak of a *lock-in situation* (Arthur, 1989): a certain combination of design and user practices have become fixated to such an extent that alternative configurations become unfeasible – even if such alternatives are more desirable. Think for instance about the way that our society and economy are completely reliant on fossil fuels and the necessity to use renewable forms of energy because of climate change and resource depletion. The use of fossil fuels is deeply embedded in economic, political, and technological infrastructures, we can speak of a *carbon lock-in* (Unruh, 2000). If it has proven to be impossible to successfully implement a faster keyboard, it does not take a lot of imagination to realize how immensely difficult it is to overcome this carbon lock-in.

Lock-ins are the result of *positive feedbacks*, a notion which characterizes a situation in which an advantage in market share reinforces itself, leading to market dominance. An example of such positive feedback is the process in which the VHS system outcompeted the Betamax and

Philips V2000 systems and became the standard in home videotape recording – until digital forms of recording took over in the late 1990s.

Taping television programs only became possible in the 1970s, when three systems were introduced on the consumer market: tThe Betamax developed by Sony (1974), V2000 developed by Philips (1976), and VHS developed by JVC (1976). While Sony and Philips focused on image quality in developing their systems, JVC focused on a low price. In 1977, the American media company RCA wanted to ally with one of the companies, but they wanted to have a lengthy recording time of up to 4 hours. As this would mean a degradation of picture quality, Sony refused to cooperate with RCA. JVC, however, agreed to develop a system that could handle lengthy recordings. Especially in the USA, the length of recording time proved out to be the most important criterion for consumers, more important than picture quality and also more important than price (though the lower cost of the VHS system did help). The length of the tapes was convenient for the American market, as it allowed the recording of movies and American Football games instead of just the one-hour television programs that could be recorded with the Betamax system. Becoming the more popular system, JVC could further decrease its prices, making it even more attractive for new consumers. This proved to be especially beneficial on the European market where length was no decisive criterion, because the PAL-system that is used to broadcast television programs allows lengthy recordings anyway. The American success of JVC allowed them to reduce the price of VHS recorders even more, and gave them a big competitive advantage, so it could outcompete both the Betamax and the V2000 systems up to the point that these latter two systems disappeared completely.

This story of the 'videotape format war' contradicts the idea of diminishing returns, which is usually assumed in economic theory and which holds that economic actions eventually create a negative feedback that leads to a predictable equilibrium for prices that marks the most efficient use and allocation of resources. Brian W. Arthur (1989) states that there is no guarantee that a particular economic outcome is the 'best' one, as positive feedbacks may contribute to cause small market advantages that cause non-linear and irreversible outcomes. Being non-linear entities, positive feedbacks make it impossible to predict the future of a technology, you never know which small advantage might be a decisive factor in the end, you never know the point of entrance of the feedback loop.

Positive feedbacks may be based on the expectations about the success of a new technology, creating self-fulfilling prophecies. They may also be based on the way a new technology connects to our experience or existing technologies, as was shown in the case of the QWERTY-keyboard. Economies of scale and learning processes are other factors that contribute to a situation in which a specific technology has an advantage over another technology: selling more products might lead to the decrease of prices and to consumers getting used to work with a certain product or system. Another factor that contributes to positive feedbacks is the presence of network effects, for the fact that it is advantageous for a technology to be associated with existing networks of users and other technologies.

This last factor raises our attention as our current society can be characterized as a *network society* in its very essence (Bauman, 2000; Castells, 2011): we have a market that is based on global computer networks, allowing to do financial transactions with everyone, everywhere, anytime at all; information is distributed around the globe as fast as the speed of light; organizations have transformed from command-and-control bureaucracies into self-steering networks; economic success is on access to resources, but on access to information. In all, the traditional economy has fundamentally changed from an economy that could be characterized by the law of diminishing returns into an economy that is characterized by positive feedbacks. Think, for instance, how inevitable it is to use Facebook to keep in touch with other people, you cannot choose by yourself for alternative platform – even though that alternative may be more appealing –, because you would be all alone.

This does not mean that technologies that are characterized by positive feedback are immune to the forces of creative destruction. On the contrary, who remembers VHS, who remembers Myspace? Winning technologies do not have to be survivors on the long run, however, as winners they will have set off a certain technological trajectory that cannot be avoided – they have created their own sociotechnical interdependencies that will figure as the basis for new successful technological developments. As a simple example we can return to the video format war. It has been an urban myth that VHS won this war because, unlike respectable companies as Philips and Sony, JVC allowed porn movies to be prerecorded on VHS. Despite the fictitious character of this story, no company would ever prevent porn to be distributed on their new system since then. Path-dependency has led a myth to become reality in the creation of new technologies.

2.4 Sociotechnical scripts

Above it has been said that we may feature 'regimes' as a conceptual extension of the notion of 'institutions'. At the same time, the concept of 'regimes' stresses that the meanings and practices which enable and constrain technology development are not only embedded in *institutions*, but, as an expansion of the concept of institutions, also in *technological infrastructures* – or technologies –

themselves. With that, the notion of 'regimes' helps us to accentuate the difference between 'software' and 'hardware', so to speak, in matters related to technology. In the hardware that is constituted by technological artefacts there are certain rules embedded, which makes these artefacts a manifestation of a regime that, because of their sheer materiality, are hard to change.

The quality of technologies to function as rule systems has been recognized in STS and the philosophy of technology, especially with regards to the normative behavior that is invited by an artefact or system (Correljé et al., 2015). A clear example is that of a speed bump ('the dead policeman') which urges a driver to take up a cautious driving style. But there are also less obvious applications of value-laden functionalities in the design of technological artefacts and system, as was illustrated in Langdon Winner's article *Do artifacts have politics?* (1980). In this article, Winner claims that the urban planner Robert Moses designed low overpasses over the parkways on Long Island so that buses from New York City could not access the beaches. As a result, the urban poor, primarily African Americans, could not reach the shore, as they were dependent on the buses for transport. The beaches were therefore only accessible for the white, upper and middle classes. Winner concludes that values (or in his words 'politics') are deliberately designed into technological artifacts.

Another example of such politically biased artefacts can be found in the way that modern cities are rife with 'unpleasant designs' or 'silent agents', that manipulate behavior in public places.[6] San Francisco, the birthplace of street skateboarding, was the first city to design solutions such as 'pig's ears' – metal flanges added to the corner edges of pavements and low walls to deter skateboarders. These periodic bumps along the edge create a barrier that would send a skateboarder tumbling if they tried to jump and slide along. Indeed, one of the main criticisms of such design is that it aims to exclude already marginalized populations such as youths or the homeless. Preventing rough sleeping is a recurring theme. Any space that someone might lie down in, or even sit too long, is likely to see spikes, railings, stones or bollards added. In the Canadian city of Calgary, authorities covered the ground beneath the Louise Bridge with thousands of bowling ball-sized rocks. This unusual landscaping feature wasn't for the aesthetic benefit of pedestrians walking along the nearby path, but part of a plan to displace the homeless population that took shelter under the bridge. In recent years, public benches too have been redesigned to stop somebody sleeping there.

However, the value-laden aspects of technology do not always have to be included on purpose. Often these aspects are the outcome of *implicit* design: designers and producers have an implicit world view that drives their technological design, which is the core of the idea developed

[6] See: http://www.bbc.com/future/story/20131202-dirty-tricks-of-city-design

by Akrich (1992) that technological artefacts embody certain 'scripts'. In the design of a new technology, designer use certain images or representations of their 'target audience'. Often these images or representations are only held unconsciously by the designers, but they have the effect that certain tastes, competences, motives, aspirations, and prejudices become inscribed in the artefact. In their turn, these scripts steer, guide and limit the behavior of the user. Van Oost (2003) and Oudshoorn *et al.* (2002) show how different artifacts, such as electric shavers, bicycles, and microwave ovens, were specifically designed with a definite idea about how male and female users relate to such technological artifacts. Male shavers are grey and black, contain dials and screws, and can be opened up by the user to take a look 'under the hood'. Female shavers, by contrast, are smooth, come in pastel colors, have no dials and screws and cannot be taken apart. Moreover, they are sold as cosmetic devices, not as electrical appliances.

A similar concept to script is that of 'affordances', a concept originally coined in psychology, which is used to describe the actions that are allowed or made possible by the combinations of physical properties of objects (Gibson, 1977). Whereas the concept of scripts directs us to implicitly held ideas that are designed into a technological artefact, the concept of affordances directs us to the way a user is confronted by the practices that are allowed by a certain technological artefact (Van den Berg & Leenes, 2013). In both cases, the conclusion reads that technologies are not value-neutral, they are not merely instrumental objects. In many ways, the objects of technology are strongly value-laden, as they incorporate certain (often dominant) values while failing to represent others. Furthermore, they may also give rise to new types of behavior, and with that they also lead to new expectations and new sets of values. In other words, technologies *mediate* perceptions, experiences, practices, and norms (Verbeek, 2006).

Technologies steer our behavior, but that does not mean that users are without power to resist the non-negotiable nature of technology. For instance in case of the unpleasantly designed urban furniture, designers have reacted by coming up with playful ideas to make their city more comfortable again. At first glance their creations are almost silly, but they are based on the serious point that unpleasant design can create exclusions in a city, and divisions between the rich and poor. One German artist, Oliver Schau, devised a simple solution to reclaim the unforgiving architecture of Hamburg, by wrapping bright yellow flexible plastic pipe around bicycle racks and bridge struts to create impromptu resting places that would be impossible to sit on otherwise. In other words, users and other technology developers may design their own scripts and counter-scripts (Van den Berg & Leenes, 2013). In the end, technology is an open-ended process that instigates its own reality and its own, unpredictable trajectories.

2.5 Expectations as self-fulfilling prophecies

We tend to think of technology as things that are essentially tangible. Instruments, tools, artefacts, infrastructures, and so on. Also when we talk about future technologies, we first seem to think about its physical appearance. At least to some extent this is rather weird as future technologies, by definition, exist only as thought constructs, be it in mind, on paper or in words. Though future technologies do not exist in empirical reality, they do exist as conceptual entities, they are words that carry a certain meaning – they are *discursive entities* only relating to the software part of technology (Pesch, 2015b). In the words of Jasanoff and Kim, they are 'sociotechnical imaginaries'. Imagination is not just a fantasy or an illusion, but an important resource that enables new ways of living by projecting goals and seeking to attain these. Imaginaries can be seen as authoritative descriptions of futures that are considered to be attainable and desirable. They revolve around the construction and reproduction of expectations about the qualities and benefits of certain scientific developments and technologies.With that, imagination helps to produce systems of meaning that enable collective interpretation of social reality (Jasanoff & Kim, 2009, p. 122). In the realm of technology, future visions and expectations shape trajectories of research and innovation and with that they become *performative* (Brown & Michael, 2003): they create their own reality like self-fulfilling promises. The awareness that technological development starts with the creation, articulation and dissemination of expectations, prompts us to look at the software of technology development, instead of looking at the hardware. Talking about responsible innovation is talking about things that are not there, and in most cases it is talking about things about which we are unsure whether they will ever exist at all, and if they will come to exist, what they will look like, how they will be used, and what the impact of their use will be.

This is testified, for instance, by Moore's law, which states that every 18 months chips will become twice as fast. This law, formulated in 1965 by Intel founder Gordon Moore, has been relatively accurate over the last 50 years, which suggests that technology progresses without external interference. Though Moore's law looks like a natural law, in reality this prediction has been playing a role in the strategic game that is played between the manufacturers of memory chips. They take the prediction of Moore's Law as a yardstick for their own progress and for further investments. When the promised specifications run the risk of not being met, additional measures are needed, such as entering into strategic alliances. Companies use the prediction of Moore to decide on the R&D goals and the size of the investments. They regard this as the right strategy because they assume that others will do the same: self-preservation implies to obey Moore's Law as the authoritative view of the future (Van Lente, 2012).

Expectations do not emerge from thin air, they are the result of human agency. In technology development, the construction of expectations often is a deliberate form of activity. Actors strategically raise expectations by *promising* that a new technology allows to resolve current or future practical problems, they do so in order to mobilize resources for their work. Engineers, scientists and technology developers try to appeal to an audience of actors who can provide the financial means, time, policy support, and/or organizational capacity to substantiate the technology. As Van Lente and Rip (1998) claim, such strategic behavior opens up a *rhetorical space* that allows a technology to become a reality. Moreover, technological promises and the construction of visions helps to *coordinate* actions by various stakeholders (Dierkes, Hoffmann, & Marz, 1996; Grin, 2000; Quist, 2007). If people share expectations about the future, including the role that a technology-to-be will play, they can adjust their activities to that imagined future.

2.6 Innovation systems

Earlier is has been said that contemporary society can be denoted as a network society. This does not only relate to globalized computer networks, but also to the way that we engage in professional activities. Until recently, our economy could be characterized by the presence of bureaucratic organizations that were based on hierarchical management structures that, in turn, could be straightforwardly designated to either one of the institutional domains of state, market or science. Nowadays we have an economy that is based on organizations that are flexible and that conveniently cross the boundaries between state, market and science. One of the most salient manifestations of the network society may be found in the domain of technology development: literature emphasizes that innovation is a product of the linkages between science, markets and governments (Etzkowitz & Leydesdorff, 2000). Modern technology cannot be developed without having a fertile synergy between the resources of knowledge, money and regulation. To an increasing extent new technologies are produced by networks of industries, universities, and governments. The complexity of modern technologies and of modern society implies that innovation has become a process that crosses boundaries between the domains of knowledge production, private investment and regulatory arrangements. Innovations are the outcomes of *innovation systems*, not of isolated inventors.[7]

Typical of innovation processes is that they span the institutional constellation, and with that they can be denoted as 'hybrid' in their essence (Avenel, Favier, Ma, Mangematin, & Rieu, 2007; Gibbons, 2000). Modern technology has become a very specialized, multi-actor undertaking (Swierstra & Jelsma, 2006) that link activities in science, industry and the market (Elzen, Enserink, & Smit, 1996), and as such they involve the simultaneous presence of a plurality of institutional domains that shapes the actions of engineers, bringing together both entrepreneurial and scientific elements.

Hekkert et al. (2007) describe how such innovation systems have different functions that drive them towards technological change. The first of these functions is that of *entrepreneurial activity*, which states that the role of the entrepreneur is to turn the potential of new knowledge, networks, and markets into concrete actions to generate – and take advantage of – new business opportunities. Function two is that of *knowledge development*, usually in the form of R&D activities. Function three concerns *knowledge diffusion* through networks. The fourth function pertains to *guidance of search*, such as long term policy goals. Function five, that of *market formation*, relates to the necessity that new technologies have to fit in an existing market, they have to be connected to the technologies that are already known. The sixth function is that of *resource mobilization* and it highlight the fact that resources, both financial and human capital, are necessary as a basic input to all activities within the innovation system. The seventh and final function is called the *creation of legitimacy*. As innovation boils down to creative destruction, there will always be some party who loses from the introduction of a new technology, which suggests that a technology will usually meet with resistance that has to be overcome by creation a fair amount of support – either from market parties, governmental actors or the public-at-large.

In short, we can see an innovation system as a hybrid constellation of market, government and science from which different capacities emerge that drive the innovation system as a whole, such as resource mobilization, entrepreneurial activities, knowledge diffusion, policy support and legitimacy creation. The innovation system allows the concerted establishment of problem definitions and expectations, the connection between variation and selection environment, and, with that, the formation of sociotechnical regimes. However, as will be argued in the next chapter, an innovation system cannot straightforwardly figure as a sociotechnical

[7] Unfortunately, the label of 'system' is used in different ways in innovation studies, possibly leading to confusion. We have already seen 'technical systems', here we introduce 'innovation systems'. Furthermore, literature also speaks of 'sociotechnical systems' and 'system innovations'.

public, it is an empirical phenomenon that might give rise to certain moral concerns that need to be addressed in order to have responsible forms of innovation.

Questions for this chapter:

- What does the approach of the social construction of technology entail?
- What are reverse salients and what are critical problems? How do these notions relate to the problem definitions that are held by the actors that are involved in innovation processes?
- Characterize the evolutionary approach to technology development?
- How does the interaction between the variation and selection environment lead to the establishment of sociotechnical regimes?
- What are lock-in situations and how do they relate to positive feedbacks?
- Why are network technologies vulnerable for positive feedbacks?
- What are sociotechnical scripts and explain how this notion relates to the idea of rule-systems?
- Why can expectations about technology be seen as self-fulfilling prophecies?
- What is an innovation system and how does such a system relate to the formation of expectations?
- Why is technology development a hybrid activity?

3. Learning for responsible innovation

Whereas the previous chapter looked at the empirical nature of technology development, this chapter focusses on the ways to pursue patterns of responsible innovation by actively intervening in the course of technology development. The problem with regards to responsibility is that technology development cannot be connected to a distinct institutional domain that figures as a coherent rule-system: engineering is a hybrid activity which makes the responsibility of an engineer always a diffuse issue. By developing proxies that represent the publics and their values at stake, we may enhance the moral intuitions of an engineer. These proxies, or sociotechnical publics as they are called here, are based on the frameworks of **Constructive Technology Assessment** (CTA) and **Value Sensitive Design** (VSD). Constructive technology assessment aims at creating **learning effects** by bringing actors from the variation and the selection environment together in **bridging events**. It is important to include **outsiders**, especially members from the public-at-large, as they will be affected by the future technology. As manifestations of CTA, we will present **scenario workshops** as a way to organize bridging events and we will present **strategic niche management** (SNM) as a method for stimulating **sustainable transitions**. VSD aims to include **public values** in the design of technologies and their **institutional context**.

3.1 Hybrid responsibility

Having touched upon the question how technologies come into being, we can now create a better understanding of the roots of the moral discomfort that has been introduced earlier. These roots relate to the *mediating role of technological artefacts*: technologies make us do things while we usually consider technologies to be value-neutral. Moreover, *the presence of regimes* make it hard to adjust the prevalent position of already used technologies, taking away our ability to counteract any negative effects. Evolutionary theory also shows us that *design and use cannot be easily separated*, which makes a functional distribution of responsibilities quite problematic. Another complicating factor here is that *innovation processes are not the work of isolated individuals* or groups of people that can be straightforwardly designated.

With regards to the last point we can say that technologies are the result of the activities of actors that come from different institutional backgrounds. This means that for technology developers it is hard to identify a clearly defined accountability structure. Engineers relate to different activities, both entrepreneurial and scientific, and the societal context to which they respond bears the character of a network. This means that in order to actively reflect on societal values, engineers have an ambiguous point of reference with respect to the nature of their activities, and it also means that the moral rules upon which actions are based are contingently formed at the level of a network, and not at the level at which wider societal values are produced. As there is no separate institutional realm that can figure as an accountability structure, technology developers might have to deal contradictory moral demands: it can happen that in a single engineering activity, moral appeals from different accountability structures are made at the same time.

Reflecting on the network-like character of engineering practices, Swierstra and Jelsma (2006) find that it gives the opportunity of shifting moral responsibility to others (also see Van de Poel, Nihlén Fahlquist, Doorn, Zwart, & Royakkers, 2012). In other words, the presence of network-based innovation systems has some moral repercussions, which does not mean that we have think negative about these innovation systems. They are simply a fact of life. What we do have to think about is how we can align the reality of innovation systems with a satisfactory accountability structures. As engineering practices are fundamentally hybrid and courses of action are predominantly derived from regimes, any direct articulation of societal values does not take place. As a legitimate accountability structure is not only absent but also impossible, we have to use alternative methods that allow us to have at least some of the functionalities of an accountability structure: we have to use 'proxies' of accountability structures which allow the articulation of publics and of public values, so that, on the one hand, engineers can reflect on these values in an active way and, on the other hand, we may intervene in the technology development with the aim of having better technologies.

This paper will propose a combination of the frameworks of Constructive Technology Assessment (CTA) and Value Sensitive Design (VSD) to develop insights about the preconditions for having suitable proxies for accountability structures. Following CTA, we will assert that engineers have to be able to interact directly with societal stakeholders. Accountability structures figure as rule systems that provide a feedback loop as to adjust individual behavior to societal norms. CTA instructs us to establish so-called 'bridging events' in which producers and users can learn about future technologies and their consequences. These bridging events should target the specification of societal values in relation to a new technological design in line with the approach of VSD, so that engineers can anticipate the moral connotations of their work. The bridging event should figure as the venue that allows questions to be asked such as: how will people relate to the new technology, which values are or might be triggered by the new artifact or system, which stakeholders will become affected negatively or positively by a new technology, etc.?

3.2 Constructive technology assessment

As its name indicates, Constructive Technology Assessment is a form of Technology Assessment.[8] The denominator of 'constructive' is used as CTA is focused on influencing the *creation* of a new technology. It is a method that actively intervenes in the beginning the innovation process and hopes to resists lock-ins by opening up the process of regime formation. This is done by bringing stakeholders together so that more societally desirable technologies can be constructed.

CTA tries to open up regimes by developing new ways of connecting the variation environment of technology development to the selection environment (Schot & Rip, 1997). The presence of regimes obstructs the variation environment to develop alternative technologies and the selection environment to express its needs and desires. The idea is to bridge the gap between the variation and selection environment so that the obstructing effects of regimes can be overcome: technology developers are able to design more desirable artefacts and systems as needs and desires are directly expressed to the variation environment; the selection environment will be better informed about the impacts that technologies will have on their lives. In fact, CTA makes use of the propensity that technological evolution is not blind, so that it makes sense to talk of a *quasi-evolutionary model of technological development* (Rip & Kemp, 1998; Van Lente, 1993) in which the variation and selection environments can be actively adapted to each other. Actors from the variation and selection environment are brought together in so-called 'bridging events' (Garud & Ahlstrom, 1997; Robinson, 2009) in which participatory methods are used to engage in learning about and anticipating to possible future technologies.

If we want to relate CTA to the account of responsibility that we have presented earlier, we can see that socio-technical regimes are rule-systems that might share *some* traits with accountability structures, but which are *not fully* legitimate as accountability structures. The reason for this is that regimes do not provide a feedback loop to adjust individual behavior with values held and articulated at a wider societal level. In other words, we have to construct a societal context functioning as an accountability structure that should enable active responsibility. We will use CTA in order to be able to identify a sociotechnical public that can figure as a proxy for an

[8] CTA is not the only new form of TA. Other forms for instance are **participatory TA**, which focuses on the inclusion of the public sphere (see: Joss, S. (2002). Toward the public sphere—Reflections on the development of participatory technology assessment. *Bulletin of Science, Technology & Society, 22*(3), 220-231.) and **Real-time TA** which unlike CTA is not involved in the experimentation with new technologies, but which focusses especially on the process of knowledge creation (see: Guston, D. H., & Sarewitz, D. (2002). Real-time technology assessment. *Technology in Society, 24*(1–2), 93-109.). In many regards, all these forms of TA overlap, especially with regards to their theoretical starting points.

accountability structure. For that we have to find out how a diversity of actors relates to a new technology and how these actors can be brought together.

An important starting point that we can derive from CTA is that we have to look at the 'software' side of technology development. When talking about a future technology, we have to especially target the meanings and expectations that different actors have developed with regards to the technology, which in essence is a future technology. Such meanings and expectations differ substantially over the people that relate to the new technology. Actors directly involved in the creation of the technology usually draft highly optimistic scenarios, anticipating success. They focus on immediate problems to be overcome, and not on the generic obstacles. Actors not directly involved typically take a more global and generalistic stance, they like to broaden the discussion by including a variety of issues that are easily ignored by the technology creators, such as costs, distribution problems and ethics (Garud & Ahlstrom, 1997). Bridging events have to enable the different actors in the acknowledgment that their own expectations are not the only possible. They have to become aware of other people's realities, and they have to be able to take this along in their assessment of and their work on the new technology.

Such processes that are aimed to make actors probing each other's realities, or that aim to contribute to the construction of common future orientations are seen as *learning processes* (Geels, 2002; Pesch, 2014b; Verbong, Geels, & Raven, 2008). Learning is considered to be essential for opening up the ingrained perceptions, values, practices, and mental routines, and allowing the establishment of a collective of body of meanings, including long-term orientations. Stimulating stakeholder learning can be seen as an attempt to integrate different bodies of meanings and problem definitions that are scattered over different institutions. The assumption is that by actively bringing actors together, barriers between institutions that produce different bodies of meanings are taken away. In other words, the development of collective future orientations is considered to be a prerequisite for the development of sustainable 'hardware'.

A general distinction that is made in learning theories concerns the difference between lower-order and higher-order learning. Lower-order learning refers to the acquisition of new facts or insights with regard to fixed organizational of societal objectives or fixed problem definitions. Higher-order learning refers to redefining of such objectives and new insights on relevant norms and values, adjusting problem definitions and strategies. Whereas lower-order learning takes place within relatively stable discursive structures, higher order learning implies changing discursive structures. With regards to CTA, we may distinguish three levels of learning that are relevant – each of them are in fact forms of higher-order learning. The first level is that in which actors that have different backgrounds come to *understand the meanings, values, and problem definitions of other actors*. So, it is the realization that other people think differently about the world were in. The second level is that in which actors *learn to relate their own position to those of others*. In other words, it concerns the way that people adjust their own behavior to the acknowledgement that other people think differently. The third level of learning is attained if *a shared understanding* of the technology is established. Instead of living in different worlds, actors now come to form an institutional context of some sorts that is related to a specific technology. In other words, the ultimate goal of learning is to establish a social context that can figure as a *sociotechnical public* that brings forwards the conditions for innovating responsibly.

A selected group of stakeholders may function as some kind of accountability structure, a sociotechnical public that allows the sharpening of the moral intuitions of engineers and of forwarding the values that are relevant. However, a participatory setting can *only* lead to the creation of a legitimate public if the selection of stakeholders represents the full spectrum of actors that will be affected by the new technology. Many participatory methods, however, include only actors that represent the innovation system, the 'insiders', thus representing the main institutional domains. It is important to stress that innovation is not just a game of insiders, but also of 'outsiders' that are not represented in the innovation system, but which will be affected by the new technology. Van de Poel (2000, p. 384) defines such outsiders as actors that are outside the system of interaction or network in which technical development takes place, these outsiders do not share the rules that are characteristic of a particular system of interaction. They are to be found outside of the circle of actors that reproduce the regime, the dominant rule systems. Van de Poel identifies the following categories of outsiders: outsider firms; professional engineers and scientists; and societal pressure groups. The first two categories consist of actors that are not part of the main networks that make up the innovation system, the category of societal pressure groups, most notably NGOs, is relevant as these groups may mobilize public opinion or influence government and users.

However, if we only look at societal pressure groups, we still might miss out an important factor in the societal dynamics concerning technology development (Pesch, 2015b): social groups and movements can only be formed around an articulated interest, whereas, social change can just as well be facilitated (or obstructed) by interests, values, and opinions that are not (yet) articulated. In other words, a significant range of societal dynamics is neglected by focusing on formal organizations. Hence, in order to construct an appropriate sociotechnical public we have to be on the lookout for other actors and other voices than those who are represented in the groups and organizations that traditionally promote certain societal interests.[9]

This urges us to think about the future roles of actors from the public-at-large (Pesch, Cuppen, & Di Ruggero, 2014; Walker & Cass, 2007). These roles are not just confined to those of *users* or *consumers*. Increasingly we see that people form the general public take on the role of *protestors*, for instance if they contest the implementation of technologies like wind power, shale gas or carbon capture and storage – at times leading to the termination of technology projects (Correljé et al., 2015; Cuppen, Brunsting, Pesch, & Feenstra, 2015). With regards to this role of the public-as-protestors, we can see a persistent belief of policymakers, companies and technology developers that such behavior are 'not in my backyard' (or NIMBY) responses that follow out of the lack of knowledge, the risk-averseness and the self-interest of the public. However, this NIMBY-label is in many cases incorrect, and it disturbs the dialogue between the public and the officials (Bell, Gray, & Haggett, 2005; Walker, Cass, Burningham, & Barnett, 2010; Wolsink, 2000). To have a sociotechnical public assumes the *symmetry* between all parties involved (Roeser & Pesch, 2015). Another role of the public is that of the *public-as-producers* of new technology. In the field of energy production, for instance, we can observe the emergence of so-called prosumers: citizens that produce their own energy, not only by making use of existing technology such as solar power or heat pumps, but also by developing new sociotechnical arrangements. A similar role of the public can be retraced in the notion of *open innovation* (Chesbrough & Crowther, 2006; Von Hippel, 2009), which highlights the capacity of actors outside of the main arenas of technology development to contribute to innovation.

In short, members from the public-at-large have to be included because: innovation affects their lifeworld and as such have to be consulted from a democratic point of view; resistance of the public can lead to the cancellation of new technologies that may benefit society as a whole; and because the public can contribute to innovation. But how can you identify the public? How can you find actors and voices if they are not organized? This is a fundamental conundrum that cannot be solved, but only circumvented, for instance by looking at social controversies that urge a diversity actors to articulate their interests, knowledge, values, and so on (Rip, 1986). In any case, we have to be open to include new, emergent, claims made in society in order to be able to construct a sociotechnical public.

---

[9] This does not mean that a sociotechnical public has to include a representative sample of society. It means that we have to have a diversity of voices that covers as many viewpoints as possible (see: Cuppen, E. (2012). A quasi-experimental evaluation of learning in a stakeholder dialogue on bio-energy. *Research Policy, 41*(3), 624-637.).

Including more actors than just insiders allows the opening-up of the decision making process by including a heterogeneity of alternatives, viewpoints, and interests. In this vein, bridging events that are connected to CTA provide an effective way to bring actors from the different societal angles together, to have these actors learning about each other, and to create a shared understanding so to construct a sociotechnical public. This is represented in
Figure 1.

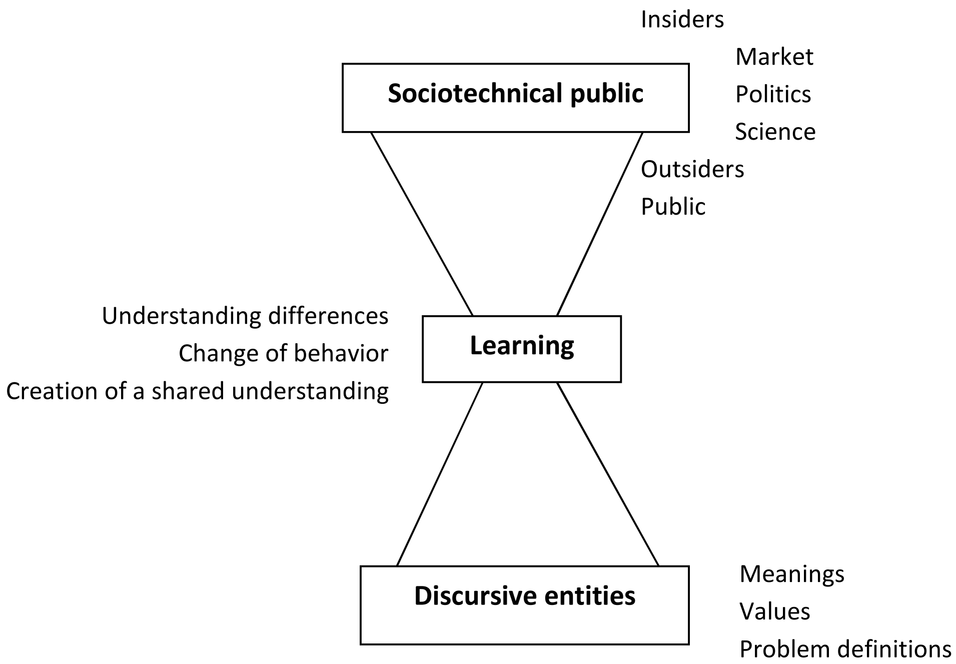


Figure 1: Framework for learning for responsible innovation

As said earlier, a 'real' sociotechnical public cannot be established, we deal with future technologies and with the public-at-large, which is a fundamentally intangible entity. As such the sociotechnical public only is an approximation that, whatever form it is given, is a methodical reconstruction that is intended to inform about the *possible* values, interests and worldviews that are at stake in the future technology. In the following sections, we will present a number of approaches that can be used to methodically reconstruct such publics. First, we will address scenario workshops, a method that is especially targeting uncertainties. In such a workshop a diversity of actors is asked to reflect on different stories about the future. Second, we will present strategic niche management (SNM) as a method to stimulate sustainable innovations. Third, we will introduce value sensitive design (VSD) as a method that aims to identify the values that are relevant in the context of a certain technology, and which intends to design these values into the technologies. Earlier it has been indicated that VSD is used in the STVM, unlike a scenario workshop or SNM, it is a desktop exercise that can be done as a student project. Its assets are that it allows to chart the values at stake, while it may also be used to detect emergent public values – giving rise to a broader proxy of a sociotechnical public.

3.3 Scenario workshops

A bridging event can be organized in the form of a 'scenario workshop' (Mulder, Petrik, Parandian, & Grondahl, 2012; Parandian, 2012; Schot, 2001).[10] Scenarios in which sets of different assumptions of possible future states are assessed belong to the most used approaches to cope with fundamental uncertainties. Scenarios are predominantly used to unfold future realities by using different initial conditions, or different dynamics. Such stories of emerging realities are not intended as the only options for the future to unfold, but as ideas that could give a clue of possible future changes and their implications, and an idea of the territory on which the future would unfold. Hence, scenario analysis is not primarily aimed at scientific rigor of its content, but at creating a sense of the range of future possibilities.

Scenarios might create an interesting storyline that allows for a far better quality of interaction with, and between stakeholders. Often stakeholders have problems in imagining a consistent future: they often have immediate demands or problems to be solved, and it might be hard to make them think in a long term perspective. With that, it is hard for many actors to articulate the way they relate to a future technology. In other words, the use of scenario's figures as a potential tool for improving the quality of interaction with stakeholders, it allows them to reflect on their future roles and it helps them to specify their future values. Moreover, the use of scenarios will help us not to overemphasize a specific future form of technology, which is prone to give rise to speculative ethics – as has been claimed earlier. The different possible futures of a technology may give rise to different values that are at stake. By contrasting these futures in multiple scenarios we can develop a more nuanced overview of the values that might come into play.

Scenarios are a form of *forecasting*: they are a method to predict future developments. As the pun – attributed to at least fifty people – goes: 'prediction is very difficult, especially if it's about the

[10] Alternative approaches are for instance **backcasting** (see: Quist, J., & Vergragt, P. J. (2006). Past and future of backcasting: The shift to stakeholder participation and a proposal for a methodological framework. *Futures, 38*(9), 1027-1045.) or **dilemma workshops** (see: Parandian, A. (2012). *Constructive TA of Newly Emerging Technologies. Stimulating learning by anticipation through bridging events.* Delft University of Technology, Delft.).

future'. Many forms of forecasting underplay this difficulty by looking for ways to reduce the range of uncertainties about the future. To a large extent, the dominant approach to technological forecasting follows the perspective of technological determinism: technologies are believed to progress autonomously, making predictions above all an activity of extrapolating the right parameters. This take on forecasting has been highly unsuccessful in the past, the list of bad predictions about technology is long and notorious. Curiously, such bad predictions have never led to the waning of the belief in the capacity of technology, they are usually apprehended as demonstrating the limitations of human imagination, not of the limitations of technologies, nor the limitations of our predictive methodologies. Moreover, there is a tendency to forget failed technologies, while emphasizing the success stories. Finally, if a technological failure is acknowledged at all, it is usually attributed to human error or mismanagement, almost never to the technology itself.

In the perspective on technology development that is used here, technology is never a thing on its own. It is always embedded in wider social and technological contexts that have qualities that are fundamentally unpredictable. With scenarios, we are able to embrace uncertainty, instead of seeing it as something that could be resolved or what could at least be predicted within a given range of uncertainty. The uncertainties that need to be taken into account find their origin in the way that a technology is embedded in a technical system, in the way that actors will interact with each other and with the technology at stake, and in the way that the sociotechnical landscape figures as a background.

Scenario workshops can be used to bring actors together, and make them reflect upon their future views and values. Such workshops take some distance of the traditional use of scenarios which take a rather 'static' approach, and often a more 'dynamic' approach, in which stakeholders actively participate in the planning process, is desirable. Scenario workshops can be seen as a micro-cosmos in which societal interactions are simulated, so that practices which would arise in societal reality are incorporated and accommodated, which can lead to a more effective design, development, or implementation of the new technology.

The reasons for using scenarios, opposed to more traditional forms of forecasting, are that the dynamics of various important factors are non-linear, which implies that small changes at specific moments lead to irreversible pathways in a development. Moreover, not all changes are external: we create the world and we are not passive spectators. So our foresight partly depends on our own actions. Hence, in thinking about the future, it is useful to make a distinction between changes that are outside our reach (that are just happening, and we just need to adapt our plans to them) and the changes that we are creating by our plans. For both changes, we might use scenarios, but they are of a different nature:

- *External scenarios* span a 'future space' in which the plans that we make should be effective and efficient. Exploring this space makes sense in order to find (all) options open to the planner. Workshops on external scenarios lead to discussions regarding 'robust' options and precautions for 'the extreme'.
- *Internal scenarios* represent the main comprehensive strategies that could be implemented. These scenarios can be evaluated for their consistency, and lead to discussions regarding their 'success in meeting predetermined targets', and for their 'other impacts' under the various external scenarios.

Although there is no fundamental reason to choose for a certain range of scenarios, the most convenient number is four, which allows a clear plotting on two axes and keeps the number of alternatives manageable. Moreover, a juxtaposition in a coordinate system helps to illustrate the antagonism that is pertained by certain choices – which, especially in relation to internal scenarios, can be seen as an important asset. One can represent different alternative internal scenarios as *dilemmas* with which actors have to deal. Subsequently, the main ingredients of the construction of the scenario storylines are given by the *uncertainties* that are at stake at a given sustainable alternative. The forks of future developments can be found in several domains of uncertainties, which include changes in the domains of demography, technology, culture, economy and/or politics or institutional contexts. The identification of uncertainties in these domains can be based on desk research, brain storm sessions, and participation of symposia and conferences, but particularly on expert interviews.

The juxtaposition of different storylines based on dilemmas shows that certain choices might create specific types of future dependencies between actors and technologies, that, in turn, might raise problems with respect to the overall goal of sustainable development. By showing these dependencies via scenario storylines, decision makers can be made more aware of the effects of their choices.

3.4 Sustainable innovation and strategic niche management

The involvement of outsiders is not only important to create a public that can figure as an accountability structure. According to Van de Poel (2000), outsiders are also important to stimulate radical innovations. Such radical innovations may be the most designated way to accommodate one specific value, namely that of sustainable development, which according to René von Schomberg (2013) is one of the salient values that needs to be pursued in order to accommodate responsible innovation. However, the chances for a radical innovation to become successful are, by definition, counteracted by the existing regimes, which prompt us to think how we can overcome the stronghold of regimes.

Continuing the use of current technology will further contribute to environmental degradation, such as global climate change, the depletion of resources, and the pollution of our life world. It is estimated that in order to realize a society that can sustain itself, technologies that are a factor 20 more eco-efficient have to be in place (Mulder, 2006; Weaver, Jansen, Van Grootveld, Van Spiegel, & Vergragt, 2000). We need radical innovations that are part of a so-called 'transition', which is defined as a radical, structural change of society that is the result of a coevolution of economic, cultural, technological, ecological, and institutional developments (Rotmans, Kemp, & Van Asselt, 2001; Van den Bergh, Truffer, & Kallis, 2011). The *multi-level perspective* (MLP) developed by Geels (2002) sees transitions as the result of the interplay of developments at three analytical levels: that of the *niche*, which is the locus for radical inventions; and those of the, already described, sociotechnical regime and the sociotechnical landscape. Niches are 'protected spaces' such as R&D laboratories, subsidized demonstration projects, or small market niches where users have special demands and are willing to support emerging innovations (Kemp, Schot, & Hoogma, 1998; Schot & Geels, 2008).

Niches are considered to be breeding grounds for innovation, safe places where innovations can be tried, tested, and mature. Whereas large, often long-term changes are difficult to design, manage, and control, niches do promise a certain level of influence and control. Aim is to create a level playing field for sustainable innovations; once they flourish, they can compete with alternative, mainstream technologies. The deliberate creation of niches has been given the name of *strategic niche management* (SNM), an approach that aims to use the niche to instigate various *learning processes*, that will create a stable sociotechnical configuration that challenges the dominant regimes (Kemp et al., 1998; Schot & Geels, 2008). Learning is thus at the heart of the three core processes by which niches are usually characterized:

- The *articulation* (and adjustment) of *expectations or visions*, which provide guidance to the innovation activities, and aim to attract attention and funding from external actors.
- The *building* of *social networks* and the enrolment of more actors, which expand the resource base of niche-innovations.

- *Learning* and *articulation* processes on various dimensions, e.g. technical design, market demand and user preferences, infrastructure requirements, organisational issues and business models, policy instruments, symbolic meanings (Geels, 2011, p. 28).

Comparable with CTA, strategic niche management is intended to stimulate learning. Unlike CTA, however, strategic niche management focusses on the 'hardware' side of a technology. It introduces an environment in which actors can actually experience the new technology. This does not mean that strategic niche management has a short-term perspective, on the contrary, it aims to instigate the first steps that are necessary for large-scale sociotechnical transitions – the regime have to be opened up by creating small points of leverage. The tangible experimental context of a niche is still very much intended to allow the development of shared visions (Raven, 2005; Schot & Geels, 2008; Späth & Rohracher, 2010). As such, the niche itself can be seen as a manifestation of a sociotechnical public.

3.5 Value sensitive design

The rationale of VSD is that since technologies should primarily serve societal needs, the various societal complexities and associated ethical problems should be anticipated at the earliest possible stage. Responsible innovation requires the identification of the relevant public values. Ideally, such an analysis is an ongoing activity that starts *ex ante*. With that, VSD directly builds forth on the consideration that technologies are intrinsically value-laden. By proactively identifying relevant societal values, the design of a new technology can be adjusted so that its social acceptability is increased. Though VSD in itself does not allows us actively involve a group of actors, it allows a reconstruction of a sociotechnical public and the identification of its values. Moreover, it directs us to the opportunities to include the concerns and values of this sociotechnical public in the design of the technology at stake, and its institutional context.

VSD was first introduced to incorporate public values into human computer interactions (B. Friedman & Kahn Jr, 2002; Van den Hoven, 2007), and later the methodology was elaborated to address the inclusion of values in other domains of technological design (e.g. Nissenbaum, 2005; Taebi et al., 2014). VSD aims to create a technological design that adequately incorporates the relevant public values, seeking solutions through technological *adjustment*, i.e., design changes.

The methodology that is proposed in VSD is to have an iterative tripartite process consisting of conceptual, empirical, and technical investigations (Manders-Huits, 2011; Nissenbaum, 2005). The conceptual investigations includes the identification and articulation of the central values in a particular design context and the identification of stakeholders that are affected by this design. Subsequently, the findings from the conceptual investigations are used to find out how stakeholders experience technologies with regard to the values they consider important. The technological investigation aims to contribute directly to the design and performance of the technology in question by focusing primarily on the question how the technology can support the human moral values that are found to be relevant. With that, VSD can be seen as a form of 'front-loading ethics' (Van den Hoven, 2005), allowing a preemptive stance with regards to ethics and technology.

In order to translate values into design requirements, we will distinguish three different levels in a 'value hierarchy' (Van de Poel, 2014). At the highest – most abstract – level, there are fundamental *values* someone may hold paramount, such as safety, environmental friendliness, economic efficiency and so forth. Contestations do not (often) arise from what constitutes a value. Everybody will supposedly endorse abstract values like safety, equity, and efficiency. Rather, controversy arises from how the value is specified into *norms*. Norms are located at the second level of hierarchy and form prescriptions for or restrictions on actions. Such norms may include *objectives* (like 'maximize safety', 'safeguard environment' or 'minimize costs' without a specific target), *goals* that specify a more tangible target, and *constraints* that set boundary or minimum conditions. The bottom level of the value hierarchy, which is also the most concrete one, indicates the technical and institutional *design requirements* that are derived from the norms.

Dignum et al. (2015) observe that the public debate on a new technology often addresses values or norms in the form of *arguments,* which are put forward in the public debate. Such arguments comprise normative statements about *how* the world should be. From the arguments used in the public debate we can gain insight into the values at stake and the associated norms. As such the identification of arguments can be helpful to also identify the values that are specified outside of the circle of insiders.

Figure 2 shows Van de Poel's value hierarchy, including the position of the arguments.

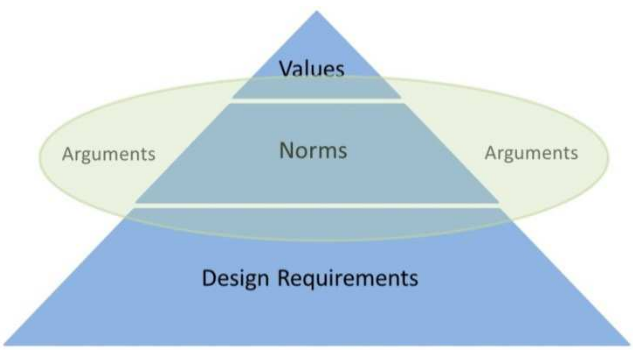


Figure 2: The three levels of the value hierarchy

The value hierarchy can be used both as an analytical tool and as a design tool. As an analytical tool, it can help to analyze *why*, or *for the sake of what*, something is being done or preferred by someone. It can help to explicate the values that underlie certain decisions or characteristics of a design and it can help to illuminate controversies when values and/or norms were specified in the design process but not incorporated in the design. As a design tool, the value hierarchy can be used to come up with a design that is robust in the sense that it can bring together divergent values and norms into a coherent set of design requirements.

VSD has been initially developed to allow the design of technological artefacts, however, it seems sensible to extent our scope of design. As said earlier, technological development is embedded in specific institutional contexts, which incorporate two main categories: (i) formal institutions such as laws, standards, regulations and contracts; and (ii) informal institutions such as customs, traditions and routines. Many of these institutions, especially the formal ones, may be subjected to redesign to accommodate divergent values (Correljé & Groenewegen, 2009; Taebi et al., 2014). The redesign of institutions is especially important if we take account of the role of the public-as-protestors, which was discussed earlier. In many cases, such protests cannot be seen as solely directed against the technology and its risks, but also as a manifestation of broader forms of discontent (cf. Rosanvallon, 2008). The legitimacy and transparency of decision making procedures, the way that authorities interact with the public-at-large (which is often based on the assumption of NIMBY-behavior), the feeling of not being taken seriously, and so on, are all factors that contribute to opposition from society (Cuppen et al., 2015; Dignum et al., 2015; Roeser & Pesch, 2015). This implies that the adaption of the technology at stake will not suffice for making it acceptable for the public. Instead, we could better redesign the institutional context, taking account of all these underlying institutional factors that motivate societal protest. Here we have to think of effective forms of participatory decision making, based on dialogue, compensation and ownership arrangements, and so on.

Questions about this chapter:

- Why is innovation characterized by hybridity and to what kind of moral problems does this hybridity lead?
- In which respects is constructive technology assessment based on the evolutionary approach to technology development and in which respects do these frameworks differ?
- Why is it necessary to establish bridging events and how do such bridging events relate to sociotechnical publics?

- Why is a sociotechnical public always a proxy?
- Who are the outsiders and how can these be included in the innovation process?
- Which different roles can you distinguish for the public-at-large?
- What is the difference between internal and external scenarios?
- How can a scenario workshop trigger actors to reflect on future technologies?
- How can we stimulate radical innovation needed for sustainable development?
- To which extent is a niche a sociotechnical public?
- What are the levels of values in the value hierarchy?
- How can the identification of arguments help us in reconstructing a sociotechnical public?
- Why is it important to apply VSD in case of the redesign of institutions?

4. Making a sociotechnical value map

This chapter will present the **sociotechnical value map** (STVM) as a method to map out a sociotechnical system of a certain technology as well as the values that are related to that system. The STVM combines elements from the frameworks of constructive technology assessment and Value Sensitive Design. CTA helps us to identify the relevant societal stakeholders and to bring these actors together, VSD helps us to design values into a technology – taken together allows us to reconstruct a sociotechnical public. The chapter will present the elements that make up a STVM.

4.1 The sociotechnical value map

In order to make new technologies more responsive to society, we need methods and frameworks that on the one hand allow us to analyze technologies against the backdrop of their future societal reception and uptake, while on the other hand present suggestions about how to adjust the technologies or their processes of societal uptake in such a way that the chances for undesirable impacts are reduced. In this paper, we will propose a method that aims to map out a technology based on its embeddedness in a sociotechnical system – as was proposed by Harald Rohracher (2002) – and which, at the same time, will explicitly try to include reflections on public values as part of the analysis. This method will be called a *sociotechnical value map* (STVM). Mapping out sociotechnical values serves responsible innovation by endorsing of the relevant public values during the innovation process (Taebi et al., 2014) and by establishing a sociotechnical public. The STVM is based on the combination of constructive technology assessment (CTA) and value sensitive design (VSD).

Rohracher's original idea of sociotechnical mapping was predominantly aimed at informing strategic policies for the stimulation of new environmentally friendly technologies (Rohracher, 2002, p. 474). In first place, the STVM is intended to build the reflexive capital of future technology developers, by compelling students to think about the way that a technology interacts with society, and about the moral issues that are pertained by technology. The STVM is a desktop exercise, not aimed at the direct intervention in the innovation process or the engagement of stakeholders. In line with Rohracher, we will base our method on the inclusion of insights from STS in order to analyze the development of the technology, to forecast the eventual hindrances for the further development of the technology, and to think about how to intervene in the development of the technology – in case of the STVM, this is done by identifying the relevant values and by giving suggestions about how to design these values into the technological system.

4.2 Choosing the technology

Any STVM starts with making the choice for a certain technology. The only restriction for choosing a technology is that the STVM has to address a technology that is *in* development. Though this criterion is very much open for interpretation, it will in general be clear whether the societal uptake of a technology has reached a stage in which any further intervention in its trajectory seems fruitless.

Technologies come in many different shapes and formats. We may discern a number of axes here. Technologies range from the very concrete, thinking about tools and products, to infrastructures, services and systems, to methods or just ideas. Some technologies basically convey a way to label an already existing set of practices. We may think here for instance of 'smart grids' or the 'internet of things'. Such a new label functions helps to organize expectations and coordinate activities, leading to new innovations, so they are also a suitable topic for a STVM. Another axis have technologies that are owned or managed by one singular company on one end, and technologies that are available for everyone on the other end. Talking about company-specific products, it is important to focus on the technological characteristics and not to look at the product from a business perspective. One may also use the axis that plots a technology on a scale that ranges from local to global. Another axis is that from simple to very complicated technologies. Obviously more axes may be identified. The key message here is that the nature of the technology that is addressed determines which of the elements of the STVM described below are relevant to discuss. The attention that is given to the aspects that make up a STVM is case specific.

The STVM is based on desktop research, which means that you have to make use of websites of companies, public organizations, NGOs, social movement organizations. A good STVM includes quite some detail, for instance about the activities and stated viewpoints of the stakeholders. To describe a technology, you can make use of academic journals, but also of professional literature and popular media. In any case, writing a STVM is a highly reflective job, which demands both rigor and creativity.

4.3 Mapping the technology

The *technology map* looks at the technology itself against the background of its technological context. The main purpose of this part of the STVM is to create an initial understanding of the technology at stake and to explore the uncertainties that characterize the technology. The idea is to identify the factors that may cause the trajectory of a technological development to branch of in a certain direction. First, we have to present the technology, its development and its characteristics. Second, possible technical alternatives have to be described, based on the consideration from evolutionary economics that a technology is developed in a variation environment that comes with a range of alternatives that, in turn, are presented to a selection environment. Third, the technological system which forms the background of the technology at stake will have to be described. Fourth, and finally, the landscape factors are to be explored.

> <u>A description of the technology</u>
> First it has to be shown what the character of the technology is. Is it a product or a service, an artefact, a system, or is it a concept that combines different technological developments?
> How far is the technology in its development, is it just an idea, is there a prototype? What is the history of the technological development? Also give an indication of the various performance standards, so how much will it cost, what are environmental (dis)advantages, how safe is it?

Subsequently, we will turn to the other elements that make up the variation environment and address technologies that are under development that may figure as competing alternatives.

> <u>An oversight of technical alternatives</u>
> The technical alternatives that are available on the market or under development have to be presented here. If possible, a hierarchy of technical alternatives should be determined and an indication of what determines the selection of alternatives should be presented.

Now we turn to the future trajectories of the technology. The possible branching points have to be described here. This means that the mechanisms that might determine technological change have to be sketched out.

> <u>The technical system</u>
> How can the technology be seen as part of a technological system? What is the speed of development of this system, and what are the reverse salients that hamper its further development?
> Apart from the connection with other technologies, also the connection to existing or likely regulation and legal arrangements can be made explicit. For instance, are there or will there be laws that prevent or stimulate the further success of the technology? Also think about appropriation: will a company decide to apply for patents so that its specifics have to be

> published; will the company try to keep its technology secret; or will it make its findings accessible for everyone without further ado?

A technology develops in a landscape of relatively stable long-term trends and events with big impacts. As such, the most relevant trends and most disruptive possible future events will have to be explored.

> The landscape
>
> Which long term trends are relevant for the further development of the technology? Are there foreseeable economic, geopolitical or cultural trends that might have an impact on the uptake of the technology? Are there technological changes that underpin or challenge the technology at stake? The situation with regards to singular events is more difficult to explore, as such events are, in their very essence, not based on extrapolation of observable trends. Still, in line with scenario thinking it would be good to think about a range of possible events that will significantly influence the development of the technology at stake. One may, for instance, think about geopolitical stability, disasters such famine, droughts or floods, new disruptive technologies, and so on.

All of the factors that have been identified in this part, relate to the uncertainties that may have a big effect on the course of a future development. By identifying these factors, more insight is created about the scope of possible changes, with that enhancing our capacity to anticipate intended and unintended impacts.

4.4 Mapping the sociotechnical public

The second step of the STVM concerns the *stakeholder map*, in which we will focus on the actors that play or can play a role in the innovation process. Of these stakeholders – which is the full set of actors that are possibly affected by the new technology –, we want to explore their interests, values, capacities and worldviews in order to create an understanding of how these actors may contribute to the further development of the technology at stake. A heterogeneity of societal actors plays a role in the assessment of a technology and may have the ability to stimulate or hamper the further uptake of a technology. In other words, a new technology becomes a subject of contention in a social arena that impacts the development of a technology in each and every aspect.

The goal of this part is to develop a sociotechnical public, which is an analytical reconstruction of a social context that figures as an accountability structure. This reconstruction follows the framework that is represented in figure 1, in which a public that is composed out of insiders, outsiders and the public is able to learn about the variety of meanings, values and problem definitions over which these actors dispose. In the context of the STVM, there is no actual stakeholder learning. The job to develop a systemic account of the meanings, values and problem definitions at stake is for the analyst. In this part of the STVM we will identify the relevant actors and their problem definitions. In the next step, the values will be deducted and analyzed in order to allow suggestions for value sensitive design.

<u>The technology developers</u>

Here, the actors that are directly involved in the creation of the technology have to be identified. So, who are the scientists that have worked on the knowledge-base of the technology, which companies or state organizations are involved in its development? What are the activities that are embarked on? We also have to address the promises and problem definitions that are held by the technology developers: what are the promises that have been raised for this technology, and by whom are these championed; what is the problem or need the new technology is intended to resolve; what are the expectations that vigor with regards to the technology?

<u>The innovation system</u>

The activities of the actors described above take place against the background of an innovation system that figures as the general climate in which the new technology is produced. To depict the innovation system, we have to sketch out the characteristics of industry, policy and science and we have to indicate their relationships in the context of the technology development at stake. Which parties can you recognize, what do these parties do and how do these parties interact? What are the capacities for resource mobilization? Does the innovation system give rise to the concerted creation of expectations and problem definitions, for instance by the development of supporting policies, business strategies, or scientific programs? In all, the question has to be answered whether the climate that is produced by the innovation systems supports or hinders the further development of the technology.

<u>The outsiders</u>

The innovation system presents the actors that can be seen as insiders, but as stated it is also important to include the outsiders in our analysis. Not only because it is essential for responsible innovation to include a wide range of actors and a diversity of voices in the decision making process on the technology, but also because outsiders may have a significant impact on the technology-to-be, for instance by presenting alternative problem definitions that challenge the problem definition of the insiders. As such, we are first urged to identify the NGOs, competing firms and outsider engineers and scientists that forward alternative problem definitions and solutions with regards to the technology at stake. We have to reflect on the way these competing definitions and solutions can have an impact on the further development of the technology at stake, for instance by taking their legitimacy into account, but also by looking at the respective powers of these parties. Do they have the leverage to change the process of technology development?

The emergent publics

As we have presented earlier, the general public can be seen as a special category of outsiders. Unlike the parties shown above, the public is fundamentally intangible, you just never know whether a new group of actors emerges that will try to influence the development of a technology. The identification of the innovation system and of the outsiders can be based on a retrospective empirical research. As we are fundamentally ignorant about the public, we can only be explorative here, for instance by looking at potential societal challenges, and by taking account of the roles that the public can play as user, protestor or producer.

4.5 Mapping the values

In the *value map*, we will try to identify the values of the actors described in the previous part. The first two parts of the STVM are primarily empirical exercises, aimed at gathering and organizing the right material from social reality. In the value map, we have to make a coherent interpretation of these empirical results – in other words, we have to reconstruct a sociotechnical public here. The values that are affected by the technology, and the values that are forwarded by the stakeholders have to be deduced so that the author of the STVM comes up with an analysis that resembles a learning process, in the sense that the heterogeneous set of values is rephrased in concrete recommendations for design, which will be the goal of the next step.

The values of the technology

Based on the analysis of the functional characteristics of the technology, an analysis in terms of values can be made. It has to be addressed which values are intended to be effectuated by the new technology or which values have already become embedded in the design – for instance in the form of sociotechnical scripts.

The values of the public

Based on the stakeholder map the public values can be charted out. The problem definitions, viewpoints, arguments of all stakeholders have to be rearticulated in values – making use of the value hierarchy. Not only have these values to be rearticulated, it also has to be presented how the different stakeholders relate to these values, how different stakeholders *understand* the different values, and whether there are any conflicts between the values themselves or between different understandings of the values.

4.6 Design for values

In the fourth step of the STVM, it will have to be explored how the relevant values can be designed into the technology or into the institutional framework. The main question that needs to be addressed is how the new technology can be implemented responsibly by overcoming potential value conflicts, first by incorporating the values in the design of the technology and second by the redesign of institutional frameworks, such as regulations, alternative forms of ownership or compensation schemes. Clearly this step is a highly creative one, like every design activity.

This means that we have to descend the value hierarchy, approaching the level of the design requirements. Surely, it would be too much to expect the presentation of a full set of concrete design requirements here – the recommendations given are meant to be guiding not conclusive.

<u>Values in the technological design</u>

Here, you have to explore whether there are any values and/or norms missing in the current technological design, and you have to ask yourself how these can be specified into design requirements.

<u>Values in the institutional context</u>

Are there any values and/or norms missing in the current institutional context? How can these be specified into design requirements? The institutional context include the processes in which the different groups of stakeholders interact, which may not only be seen as contexts that are open for redesign, but also as processes in which the stakeholders can articulate their values even further.

The full STVM can be represented in the Figure 3.

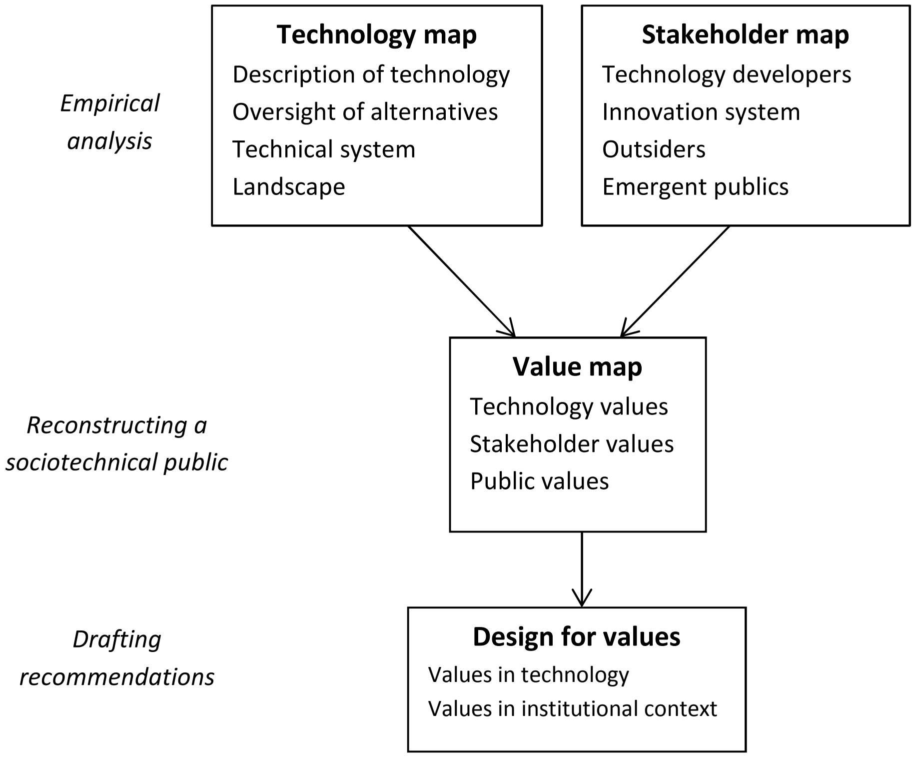


Figure 3: The Sociotechnical Value Map

The STVM is a snapshot: it pertains to only one moment in time. At the same time it should be emphasized that technology is always work-in-progress. There is no finite design, not only because technologies will evolve further, but also, and perhaps even more so, because society will always be subject to change. This means that the connection of values and technology is fundamentally volatile. Reconstructing a sociotechnical public should be a reiterative process, only then it does right to the nature of responsible innovation, which is like any form of innovation, a process that needs continuous maintenance.

References

Akrich, M. (1992). The de-scription of technical objects. *Shaping technology/building society*, 205-224.

Arthur, W. B. (1989). Competing technologies, increasing returns, and lock-in by historical events. *The economic journal, 99*(394), 116-131.

Avenel, E., Favier, A. V., Ma, S., Mangematin, V., & Rieu, C. (2007). Diversification and hybridization in firm knowledge bases in nanotechnologies. *Research Policy, 36*(6), 864-870.

Bauman, Z. (2000). Liquid modernity. *Polity, Cambridge*.

Beck, U. (1992). *Risk society. Towards a new modernity.* London: Sage Publications.

Bell, D., Gray, T., & Haggett, C. (2005). The 'social gap'in wind farm siting decisions: explanations and policy responses. *Environmental Politics, 14*(4), 460-477.

Berger, P. L., & Luckmann, T. (1991). *The social construction of reality. A treatise in the sociology of knowledge*. London: Penguin.

Berlin, I. (2002). *Liberty*. Oxford and London: Oxford University Press.

Bijker, W. E. (1997). *Of bicycles, bakelites, and bulbs: Toward a theory of sociotechnical change*: MIT press.

Bijker, W. E., Hughes, T. P., Pinch, T., & Douglas, D. G. (2012). *The social construction of technological systems: New directions in the sociology and history of technology*: MIT press.

Bovens, M. A. P. (1998). *The Quest for Responsibility: Accountability and Citizenship in Complex Organisations*. Cambridge: Cambridge University Press.

Brown, N., & Michael, M. (2003). A sociology of expectations: retrospecting prospects and prospecting retrospects. *Technology Analysis & Strategic Management, 15*(1), 3-18.

Carson, R., Darling, L., & Darling, L. (1962). *Silent spring*. Boston: Houghton Mifflin; Riverside Press.

Castells, M. (2011). *The rise of the network society: The information age: Economy, society, and culture* (Vol. 1): Wiley. com.

Chesbrough, H., & Crowther, A. K. (2006). Beyond high tech: early adopters of open innovation in other industries. *R&d Management, 36*(3), 229-236.

Collingridge, D. (1980). *The social control of technology*: Pinter London.

Correljé, A. F., Cuppen, E., Dignum, M., Pesch, U., & Taebi, B. (2015). Responsible innovation in energy projects: Values in the design of technologies, institutions and stakeholder interactions *Responsible Innovation 2* (pp. 183-200): Springer International Publishing.

Correljé, A. F., & Groenewegen, J. P. (2009). Public values in the energy sector: economic perspectives. *International Journal of Public Policy, 4*(5), 395-413.

Cuppen, E. (2012). A quasi-experimental evaluation of learning in a stakeholder dialogue on bio-energy. *Research Policy, 41*(3), 624-637.

Cuppen, E., Brunsting, S., Pesch, U., & Feenstra, C. (2015). Seeing things differently: the role of frames in societal debate on a carbon and capture storage project in the Netherlands. *Environment and Planning A, 47*(in press).

Dahl, R. A., & Lindblom, C. E. (1963). *Politics, Economics, and Welfare. Planning and Politico-Economic Systems Resolved into Basic Social Processes*. New York: Harper & Row.

David, P. A. (1985). Clio and the Economics of QWERTY. *The American Economic Review, 75*(2), 332-337.

Davis, M. (1991). Thinking like an engineer: The place of a code of ethics in the practice of a profession. *Philosophy & Public Affairs, 20*(2), 150-167.

Dewey, J. (1922). *Human nature and conduct*: Courier Corporation.

Dierkes, M., Hoffmann, U., & Marz, L. (1996). *Visions of technology: social and institutional factors shaping the development of new technologies*. Frankfurt & New York: Campus Verlag/St. Martin's Press.

Dignum, M., Correljé, A., Cuppen, E., Pesch, U., & Taebi, B. (2015). Contested Technologies and Design for Values: The Case of Shale Gas. *Science and Engineering Ethics*, 1-21.
Dosi, G., & Nelson, R. R. (1994). An introduction to evolutionary theories in economics. *Journal of evolutionary economics, 4*(3), 153-172.
Dumont, L. (1977). *From Mandeville to Marx. The genesis and triumph of economic ideology*. Chicago and London: Chicago University Press.
Ellul, J. (1962). The Technological Order. *Technology and Culture, 3*(4), 394-421.
Elzen, B., Enserink, B., & Smit, W. A. (1996). Socio-technical networks: How a technology studies approach may help to solve problems related to technical change. *Social Studies of Science, 26*(1), 95-141.
Etzkowitz, H., & Leydesdorff, L. (2000). The dynamics of innovation: from National Systems and "Mode 2" to a Triple Helix of university–industry–government relations. *Research Policy, 29*(2), 109-123.
Fitzgerald, S. (2015). Is There a Role for Spectators in Democratic Politics? A Reflection on the Theater Metaphor in Green's "Ocular Democracy". *Constellations*.
Friedman, B., & Kahn Jr, P. H. (2002). *Human values, ethics, and design.* Paper presented at the The human-computer interaction handbook.
Friedman, T. L. (2006). *The world is flat: The globalized world in the twenty-first century*: Penguin London.
Garud, R., & Ahlstrom, D. (1997). Technology assessment: a socio-cognitive perspective. *Journal of Engineering and Technology Management, 14*(1), 25-48.
Geels, F. W. (2002). Technological transitions as evolutionary reconfiguration processes: a multi-level perspective and a case-study. *Research Policy, 31*(8–9), 1257-1274.
Geels, F. W. (2011). The multi-level perspective on sustainability transitions: Responses to seven criticisms. *Environmental Innovation and Societal Transitions, 1*(1), 24-40.
Geertz, C. (1973). *The Interpretation of Cultures*. New Yotk: Basic Books.
Gibbons, M. (2000). Mode 2 society and the emergence of context-sensitive science. *Science and Public Policy, 27*(3), 159-163.
Gibson, J. J. (1977). The theory of affordances. *Hilldale, USA*.
Green, J. E. (2010). The eyes of the people: democracy in an age of spectatorship.
Grin, J. (2000). Vision Assessment to Support Shaping 21st Century Society? Technology Assessment as a Tool for Political Judgement. In J. Grin & A. Grunwald (Eds.), *Vision Assessment: Shaping Technology in 21st Century Society*. Heidelberg: Springer.
Grinbaum, A., & Groves, C. (2013). What is "responsible" about responsible innovation? Understanding the ethical issues. *Responsible Innovation: Managing the Responsible Emergence of Science and Innovation in Society. Wiley, London*, 119-142.
Grunwald, A. (2014). Technology assessment for responsible innovation *Responsible Innovation 1* (pp. 15-31): Springer.
Guston, D. H., Fisher, E., Grunwald, A., Owen, R., Swierstra, T., & van der Burg, S. (2014). Responsible innovation: motivations for a new journal. *Journal of Responsible Innovation, 1*(1), 1-8.
Guston, D. H., & Sarewitz, D. (2002). Real-time technology assessment. *Technology in Society, 24*(1–2), 93-109.
Hekkert, M. P., Suurs, R. A. A., Negro, S. O., Kuhlmann, S., & Smits, R. E. H. M. (2007). Functions of innovation systems: A new approach for analysing technological change. *Technological Forecasting and Social Change, 74*(4), 413-432.
Herring, H., & Roy, R. (2007). Technological innovation, energy efficient design and the rebound effect. *Technovation, 27*(4), 194-203.
Hughes, T. P. (1983). *Networks of power: electrification in Western society 1880-1930*. Baltimore: John Hopkins University Press.
Hughes, T. P. (1987). The evolution of large technical systems. In W. E. Bijker, T. P. Hughes & T. J. Pinch (Eds.), *The social construction of technological systems: new directions in the sociology and history of technology*. Cambridge: MIT Press.

Jasanoff, S., & Kim, S.-H. (2009). Containing the Atom: Sociotechnical Imaginaries and Nuclear Power in the United States and South Korea. *Minerva, 47*(2), 119-146.
Joss, S. (2002). Toward the public sphere—Reflections on the development of participatory technology assessment. *Bulletin of Science, Technology & Society, 22*(3), 220-231.
Kemp, R., Schot, J. W., & Hoogma, R. (1998). Regime shifts to sustainability through processes of niche formation: The approach of strategic niche management. *Technology Analysis & Strategic Management, 10*(2), 175-198.
Kemp, R., & Van den Bosch, S. (2006). *Transitie-experimenten: Praktijkexperimenten met de potentie om bij te dragen aan transities*. Delft: Kenniscentrum voor Duurzame Systeeminnovaties en Transities.
Kuhlmann, S., & Rip, A. (2014). The challenge of addressing Grand Challenges.
Manders-Huits, N. (2011). What values in design? The challenge of incorporating moral values into design. *Science and Engineering Ethics, 17*(2), 271-287.
Merton, R. K. (1979). *The sociology of science. Theoretical and empirical investigations*. Chicago and London: The University of Chicago Press.
Minogue, K. (1963). The moral character of liberalism: The Liberal Mind, Indianapolis, in: Liberty Fund.
Montesquieu. (2002). *The Spirit of the Laws*. Cambridge: Cambridge University Press.
Mulder, K. F. (2006). *Sustainable development for engineers: A handbook and resource guide.* Sheffield: Greenlead.
Mulder, K. F., & Knot, M. (2001). PVC plastic: a history of systems development and entrenchment. *Technology in Society, 23*(2), 265-286.
Mulder, K. F., Petrik, O., Parandian, A., & Grondahl, F. (2012). Scenario based learning regarding contested articulations of sustainability: The example of hydropower and Sweden's energy future. *International Journal of Sustainable Water and Environmental Systems, 4 (1), 2012*.
Nelson, R. R., & Winter, S. G. (1977). In search of useful theory of innovation. *Research Policy, 6*(1), 36-76.
Nieuwenburg, P. (2004). The Agony of Choice Isaiah Berlin and the Phenomenology of Conflict. *Administration & Society, 35*(6), 683-700.
Nissenbaum, H. (2005). Values in technical design. *Encyclopedia of Science, Technology and Society, ed. by C. Mitcham, MacMillan, New York*.
Nordmann, A. (2014). Responsible innovation, the art and craft of anticipation. *Journal of Responsible Innovation, 1*(1), 87-98.
Nordmann, A., & Rip, A. (2009). Mind the gap revisited. [10.1038/nnano.2009.26]. *Nat Nano, 4*(5), 273-274.
North, D. C. (1989). Institutions and economic growth: An historical introduction. *World Development, 17*(9), 1319-1332.
Oudshoorn, N., Saetnan, A. R., & Lie, M. (2002). On gender and things: Reflections on an exhibition on gendered artifacts. *Women's Studies International Forum, 25*(4), 471-483.
Owen, R., Macnaghten, P., & Stilgoe, J. (2012). Responsible research and innovation: From science in society to science for society, with society. *Science and Public Policy, 39*(6), 751-760.
Owen, R., Stilgoe, J., Macnaghten, P., Gorman, M., Fisher, E., & Guston, D. H. (2013). A framework for responsible innovation. *Responsible innovation: managing the responsible emergence of science and innovation in society*, 27-50.
Parandian, A. (2012). *Constructive TA of Newly Emerging Technologies. Stimulating learning by anticipation through bridging events.* Delft University of Technology, Delft.
Pesch, U. (2005). *The Predicaments of Publicness. An Inquiry into the Conceptual Ambiguity of Public Administration.* Delft: Eburon.
Pesch, U. (2008). Administrators and Accountability: The Plurality of Value Systems in the Public Domain. *Public Integrity, 10*(4), 335-343.
Pesch, U. (2014a). Sustainable development and institutional boundaries. *Journal of Integrative Environmental Sciences, 11*(1), 39-54.

Pesch, U. (2014b). Sustainable Innovation, Learning and Responsibility *Responsible Innovation 1* (pp. 199-218): Springer.
Pesch, U. (2015a). Engineers and Active Responsibility. *Science and Engineering Ethics, 21*(4), 925-939.
Pesch, U. (2015b). Tracing discursive space: Agency and change in sustainability transitions. *Technological Forecasting and Social Change, 90, Part B*(0), 379-388.
Pesch, U., Cuppen, E., & Di Ruggero, O. (2014, 27-29 August 2014). *The future of the energy transition and the public.* Paper presented at the The 5th International Conference on Sustainability Transitions, Utrecht.
Polanyi, K. (2001). *The Great Transformation. The Political and Economic Origins of our Time.* Boston: Beacon Press.
Pritchard, M. S. (2009). Professional standards in engineering practice.
Quist, J. (2007). *Backcasting for a sustainable future : the impact after 10 years*. Delft: Eburon.
Quist, J., & Vergragt, P. J. (2006). Past and future of backcasting: The shift to stakeholder participation and a proposal for a methodological framework. *Futures, 38*(9), 1027-1045.
Raven, R. (2005). *Strategic niche management for biomass*. Eindhoven: Eindhoven University of Technology.
Rip, A. (1986). Controversies as Informal Technology A ssessment. *Science Communication, 8*(2), 349-371.
Rip, A. (1988). Technologie als mensenwerk. Enschede: University ot Twente.
Rip, A. (1995). Introduction of new technology: making use of recent insights from sociolcmgy and economics of technology. *Technology Analysis & Strategic Management, 7*(4), 417-432.
Rip, A., & Kemp, R. (1998). Technological change. In S. Rayner & E. L. Malone (Eds.), *Human Choice and Climate Change* (Vol. 2, pp. 327-399). Columbus: Battelle Press.
Rip, A., Misa, T. J., & Schot, J. W. (1995). *Managing technology in society*: Pinter Publishers London, New York.
Robinson, D. K. R. (2009). Co-evolutionary scenarios: An application to prospecting futures of the responsible development of nanotechnology. *Technological Forecasting and Social Change, 76*(9), 1222-1239.
Roeser, S., & Pesch, U. (2015). An Emotional Deliberation Approach to Risk. *Science, Technology & Human Values*.
Rohracher, H. (2002). A sociotechnical mapping of domestic biomass heating systems in Austria. *Bulletin of Science, Technology & Society, 22*(6), 474-483.
Rosanvallon, P. (2008). *Counter-democracy: Politics in an Age of Distrust*: Cambridge University Press.
Rotmans, J., Kemp, R., & Van Asselt, M. (2001). More evolution than revolution: transition management in public policy. *Foresight, 3*(1), 15-31.
Schot, J. W. (2001). Towards New Forms of Participatory Technology Development. *Technology Analysis & Strategic Management, 13*(1), 39-52.
Schot, J. W., & Geels, F. W. (2008). Strategic niche management and sustainable innovation journeys: theory, findings, research agenda, and policy. *Technology Analysis & Strategic Management, 20*(5), 537-554.
Schot, J. W., & Rip, A. (1997). The past and future of constructive technology assessment. *Technological Forecasting and Social Change, 54*(2–3), 251-268.
Schumpeter, J. A. (2000). *Capitalism, Socialism and Democracy*. London and New York: Routledge.
Smith, A. (1998). *An Inquiry into the Nature and Causes of the Wealth of Nations. A Selected Edition*. Oxford and New York: Oxford University Press.
Smits, R., Leyten, J., & Den Hertog, P. (1995). Technology assessment and technology policy in Europe: new concepts, new goals, new infrastructures. *Policy Sciences, 28*(3), 271-299.
Späth, P., & Rohracher, H. (2010). 'Energy regions': The transformative power of regional discourses on socio-technical futures. *Research Policy, 39*(4), 449-458.
Stilgoe, J. (2015). *Experiment earth: Responsible innovation in geoengineering*: Routledge.

Stilgoe, J., Owen, R., & Macnaghten, P. (2013). Developing a framework for responsible innovation. *Research Policy, 42*(9), 1568-1580.
Swierstra, T., & Jelsma, J. (2006). Responsibility without moralism in technoscientific design practice. *Science, Technology & Human Values, 31*(3), 309-332.
Taebi, B., Correljé, A. F., Cuppen, E., Dignum, M., & Pesch, U. (2014). Responsible innovation as an endorsement of public values: The need for interdisciplinary research. *Journal of Responsible Innovation, 1*(1), 118-124.
Unruh, G. C. (2000). Understanding carbon lock-in. *Energy Policy, 28*(12), 817-830.
Van de Poel, I. (2000). On the Role of Outsiders in Technical Development. *Technology Analysis & Strategic Management, 12*(3), 383-397.
Van de Poel, I. (2014). Translating values into design requirements. In D. P. Michelfelder, N. McCarthy & D. E. Goldberg (Eds.), *Philosophy and Engineering: Reflections on practice, principles and process* (pp. 253-266). Dordrecht: Springer.
Van de Poel, I., Nihlén Fahlquist, J., Doorn, N., Zwart, S., & Royakkers, L. (2012). The Problem of Many Hands: Climate Change as an Example. *Science and Engineering Ethics, 18*(1), 49-67.
Van de Poel, I., & Royakkers, L. (2011). Ethics, technology and engineering. Oxford: Blackwell.
Van den Berg, B., & Leenes, R. E. (2013). Abort, retry, fail: scoping techno-regulation and other techno-effects *Human law and computer law: Comparative perspectives* (pp. 67-87): Springer.
Van den Bergh, J. C. J. M., Truffer, B., & Kallis, G. (2011). Environmental innovation and societal transitions: Introduction and overview. *Environmental Innovation and Societal Transitions, 1*(1), 1-23.
Van den Hoven, J. (2005). Design for values and values for design. *Information Age, 4*, 4-7.
Van den Hoven, J. (2007). ICT and value sensitive design. *The Information Society: Innovation, Legitimacy, Ethics and Democracy In honor of Professor Jacques Berleur sj*, 67-72.
Van den Hoven, J. (2014). Responsible innovation: A new look at technology and ethics *Responsible Innovation 1* (pp. 3-13): Springer.
Van Gunsteren, H. (1994). *Culturen van besturen*. Amsterdam & Meppel: Boom.
Van Lente, H. (1993). *Promising Technology: The dynamics of expectations in technological developments*. Enschede: University of Twente.
Van Lente, H. (2012). Navigating foresight in a sea of expectations: lessons from the sociology of expectations. *Technology Analysis & Strategic Management, 24*(8), 769-782.
Van Lente, H., & Rip, A. (1998). The Rise of Membrane Technology. *Social Studies of Science, 28*(2), 221-254.
Van Oost, E. (2003). Materialized gender: how shavers configure the users' femininity and masculinity'. *How Users Matter: The Co-construction of Users and Technology, eds N. Oudshoorn, 8*, 193-208.
Verbeek, P.-P. (2006). Materializing morality design ethics and technological mediation. *Science, Technology & Human Values, 31*(3), 361-380.
Verbong, G., Geels, F. W., & Raven, R. (2008). Multi-niche analysis of dynamics and policies in Dutch renewable energy innovation journeys (1970–2006): hype-cycles, closed networks and technology-focused learning. *Technology Analysis & Strategic Management, 20*(5), 555-573.
Von Hippel, E. (2009). Democratizing Innovation: The Evolving Phenomenon of User Innovation. *International Journal of Innovation Science, 1*(1), 29-40.
Von Schomberg, R. (2013). A vision of responsible research and innovation. *Responsible Innovation*, 51-74.
Walker, G., & Cass, N. (2007). Carbon reduction,'the public'and renewable energy: engaging with socio-technical configurations. *Area, 39*(4), 458-469.
Walker, G., Cass, N., Burningham, K., & Barnett, J. (2010). Renewable energy and sociotechnical change: Imagined subjectivities of'the public'and their implications. *Environment and Planning A, 42*(4), 931-947.

Weaver, P., Jansen, L., Van Grootveld, G., Van Spiegel, E., & Vergragt, P. J. (2000). *Sustainable technology development*. Sheffield: Greenleaf Publishers.
Weber, M. (1972). *Wirtschaft und Gesellschaft. Grundriss der verstehenden Soziologie*. Tübingen: J.C.B. Mohr.
Winner, L. (1980). Do artifacts have politics? *Daedalus, 109*(1), 121-136.
Wolsink, M. (2000). Wind power and the NIMBY-myth: institutional capacity and the limited significance of public support. *Renewable energy, 21*(1), 49-64.