From Theory to Practice: A Path Through Imperfection—Lessons from Evaluation of Policy Oriented Environmental Health Research in Belgium

Hans Keune
Research Institute for Nature and Forest (INBO) & University of Antwerp

Bert Morrens
University of Antwerp

IlseLoots
University of Antwerp

Johan Springael
University of Antwerp

Background: Dealing with complex issues per definition bears the burden of imperfection. Whatever comforting theoretical concepts may promise, real life complexity will take its messy toll once travelling from conceptual ambition to real life practice. We specifically reflect on the social scientific contribution to these inter- and transdisciplinary endeavors.

Purpose: We reflect on how social scientific research has tried to find its way in complex real life research practice in the field of environment and health and specifically focus on the challenges posed by the imperfections we encountered.

Setting: Two case studies in the field of environment and health in Belgium. One case study focuses on policy interpretation of research results. The other focuses on a selection of research priorities.

Intervention: Not applicable.

Research Design: Not applicable.

Data Collection and Analysis: inter- and transdisciplinary research and multi-criteria analysis.

Findings: Resulting from our practical experiences we present a typology of imperfections: imperfect information, imperfect expert assessment, imperfect processing of incommensurable data, imperfect socio-political weighing/deliberation and reflection on imperfection. As easy as it seems to design processes of structuring complex issues and decision making in which involvement of a relevant diversity of actors and factors is considered praiseworthy, as complicated it is, as we have shown, in practice to organize and live up to expectations. In judging the quality of this endeavour, diverse theoretical yardsticks maybe applied that will not necessarily do justice to the practical complications of research practice. The concepts of abduction, intersubjectivity and transferability, offer interesting qualitative notions of pragmatic approaches that seem relevant to our work.

Keywords: *complexity, environment & health, interdisciplinary and transdisciplinary research, social science, imperfection, quality*

We reflect on how social scientific research has tried to find its way in complex real life research practice in the field of environment and health and specifically focus on the challenges posed by the imperfections we encountered. We present a typology of imperfections and illustrate the challenge of dealing with these imperfections with examples of practical experience in two case studies in the field of environment and health. One case study (*action-plan*) focuses on policy interpretation of research results (Keune et al., 2009b). The other (hot spot selection procedure) focuses on a selection of research priorities (Keune et al., 2010). An important characteristic of *both case studies* was the close interdisciplinary cooperation between social and natural scientists and the close transdisciplinary cooperation between scientists and policy representatives. Also participation of both experts and stakeholders was an important part of the research effort. Finally, an important feature of the research was policy orientation: *both cases* had the ambition to produce policy relevant outcomes and to provide a basis for policy making. We will use the two case studies as examples to illustrate how imperfections in research practice imposed research dilemmas. We will do this without going into detailed discussion of case study specifics; the examples are exemplary for the difficulties we encountered in practice and are used to reflect on some fundamental research dilemmas.

First we will introduce the research setting and the two case studies. Next we will highlight how the research approach was designed by discussing the common methodological ambitions of *both case studies*, inspired by the analytical-deliberative approach (Stern and Fineberg 1996), and how this was conceptualized in a practical transdisciplinary approach. In the research practice section the practical work will be discussed. First, we will introduce the phenomenon of imperfection which appeared to be an unavoidable companion along the way. Next, a typology of imperfections will be presented, illustrated by several concrete examples of imperfection from research practice: imperfect information, imperfect expert assessment, imperfect processing of incommensurable data, imperfect socio-political weighing/deliberation and reflection on imperfection. In the discussion section we will focus on some fundamental issues of knowledge assessment: how can we decide on the quality of our work? We will present some ideas from the social scientific literature that may be helpful in this respect.

The Centre of Expertise for Environment and Health (CEH)

First, we will introduce the research setting and the two case studies. From the end of 2001 until the end of 2011 in Flanders (the Dutch speaking part of Belgium) human bio-monitoring research is being carried out. The human bio-monitoring project investigates the very complex relation between environmental pollution and human health by measuring some selected pollutants and certain health effects in human beings, using biomarkers.[1] This project is carried out in the scope of the Flemish Centre of Expertise for Environment and Health (2010), funded and steered by the Flemish government. In the CEH, environmental health experts from all Flemish universities and from two research institutes cooperate. The CEH combines

[1] For details: http://www.milieu-en-gezondheid.be/English/research.html

natural (Bilau et al. 2009; Schroijen et al. 2008) and social scientific research (Keune et al., 2008a).

Two Case Studies

1. Action-plan. Instigated by policy representatives together with medical and environmental scientific experts and policymakers, we developed an *action-plan* for setting policy priorities with regard to the bio-monitoring results: from research results to policy action (Keune et al., 2009b). At first, the discussions mainly focused on environmental and medical scientific interpretation of the monitoring data. It was thought of as a merely scientific quest. While trying to build bridges towards policy interpretation though the limitations of an exclusively scientific endeavor clearly showed: no scientist or group of scientists dared to claim to possess the necessary and overarching knowledge to assess the policy priorities when factors other than medical and environmental (scientific) also had to be taken into account (such as social preferences, feasibility of policy measures; issues introduced by the social scientists). The social scientists proposed a stakeholder jury in addition to experts that would judge relevant data and knowledge and developed a procedure to structure the analytical deliberative process (Stern and Fineberg 1996, see below).

2. Hotspot selection procedure. In the CEH we experimented with the input of a diversity of actors with regard to setting research priorities (Keune et al., 2010). The aim was to select hotspots: between 2009 and 2011 a limited number of local environment and health issues will be investigated by human bio-monitoring. Important questions are: "On which issues to focus our research effort?" and "Which problems are most urgent and in need of human bio-monitoring?" We designed a *hotspot selection procedure* to make an inventory of possibly relevant issues and to select a limited number for research. We opened up discussion and invited (other) scientists, policy representatives and stakeholders to propose research options. In the procedure used in this case study also a combination of expert and stakeholder input was essential. Moreover the final decision was supported by a multi-criteria analysis of expert assessment and stakeholder advice.

Designing the Research Approach

In this section, we will highlight how the research approach was designed by discussing the common methodological ambitions of *both case studies* and how this was conceptualized in a practical, transdisciplinary approach.

Research Ambition

The ambition that is represented by the case study approach can be characterized as an example of the analytical-deliberative approach (Stern and Fineberg 1996; Stern 2005). Similar concepts are the extended peer review (Pereira and Funtowicz 2006) and inclusive risk governance (Renn and Schweizer 2009). Though different in vocabulary and methodology, these approaches have one key feature in common: the opening up of debate to different fields of expertise and opinion, both scientific and non-scientific (Stirling 2008; Funtowicz et al. 1999). Another key feature is that openness is considered to be helpful in dealing with complexity by acknowledging the

relevance of both natural scientific complexity and social complexity by taking into account facts about reality and values of social actors regarding reality (Keune et al., 2009a). Renn and Schweizer (2009) point out that a variety of rationales may drive such ambitions. Analytically they distinguish for example a functionalist rationale (including all relevant knowledge and information) and a deliberative rationale (including all relevant arguments). The analytical-deliberative approach combines both the functionalist and the deliberative rationale (ibid.). The fact that the ambition and rationale behind such approaches may be very different, potentially leading to a diversity of practical procedures, has important consequences for quality assessment: it is difficult to objectify quality criteria when the approach to a large extent is characterized by normative ambitions that may differ in nature depending on the diversity of preferences of those involved (Row et al. 2004; Rauschmayer et al. 2009; Renn 2008). We will return to this issue in the discussion section.

When designing such ambitions, at an idealistic level both natural scientists and policy makers involved relatively easily seem to agree to some of the analytical deliberative features. Partly this can be explained by the popularity of 'melodious bells' such as the public participation trend in policy circles, for example at the European Union level, and ethical issues in health research. This can also be explained by limitations to their own efforts to tackle complexity in their field of expertise, especially when this effort is practice- and/or policy oriented: there is some feeling of uneasiness with regard to priority setting concerning socially sensitive topics such as environment and health policies and hotspots and to a problem solving perspective in general.

Nevertheless, also at this abstract level quite some discussion is needed: different perspectives need to come to some agreement on a joint approach that will allow fruitful collaboration. From a social scientific perspective with a background in the field of science and technology studies, the opening up of work and debate to different perspectives and stakes (Stirling 2008; Stern and Fineberg 1996) and interdisciplinary and transdisciplinary cooperation (Nowotney et al. 2001; Funtowicz and Ravetz 1990) are introduced. Some natural scientists may be reluctant to engage in interpretative gymnastics though: there seems some allergy to the 'complicatedness' of opening up debate to other perspectives. Whereas we (social scientists) believe and proclaim that openness will enrich science and policy with diverse relevant knowledge and opinion, some natural scientific scholars have a different view on relevant knowledge and objectivity. One of their main arguments is practical complicatedness: they fear that the research schedule might be endangered, that things might become too complicated. Whereas our natural scientific colleagues are used to rather complicated and nuanced routines and quality standards within the natural scientific domain, there seems some discomfort with 'other complicatedness' such as opening up debate to other perspectives. The perspective of non-scientists or non-experts (stakeholders) is easily disqualified by natural scientists because of its subjective nature. This supposes their perspective to be solely of the objective kind. Typical characteristics of the natural scientific approach in the CEH are a quest for objective scientific

proof, quantitative analysis (for example the bio-monitoring results), statistical representativeness (and thus large numbers), consensus orientation and a focus on expert opinions. Those characteristics belong to the dominant (post)positivistic paradigm in science (Guba and Lincoln 1994; Morgan 2007). Even though the policy perspective is rather different, the basic reflexes of most policy representatives cooperating with us within the framework of the CEH turned out to be rather similar to the natural scientific one, mainly due to similar scientific backgrounds.

Research Approach

The research approaches of *both case studies* have important characteristics in common. First, the close interdisciplinary cooperation with natural scientists. This implies that the general approach had to be negotiated between totally different disciplinary backgrounds, and that natural and social scientific data had to be combined in the process. Second the close cooperation with policy representatives: the research had to be policy relevant, which puts totally different demands to research than only scientific ones. Third the involvement of both experts and stakeholders is part of the assessment procedure. Finally the structuring and processing approach is, more or less, similar in *both case studies*. The basic problem that needs to be solved is choosing between options (we call them *cases*) that are rather different in nature. There are different policy options in the *action-plan*, for example policy on asthma incidence and policy on pollution from pesticides (Keune et al., 2009b) and there are different research options in the *hotspot selection procedure*, for example unexplainable higher death rates in one region and pollution from an industrial zone in another region (Keune et al., 2010). The choice is being based on incommensurable[2] assessment criteria: first seriousness of health risks, second policy aspects and third social aspects. An obvious structuring method for this is a multi-criteria method of analysis (Roy 1996; Munda 2004). The procedures are organized as follows: first desk research (such as literature and data search) provides the different options with background information concerning the different assessment criteria. The environmental and health information relevant to assess the health risk is being gathered by natural scientists.[3] The social scientists are responsible for policy-related and social aspects. Second, the desk research information is assessed in an expert consultation. Experts on environment and health assess the health risk criterion, policy experts assess the policy aspects and social experts assess the social aspects. These assessments result in both *quantitative information* (priority rankings of options on different criteria) and *qualitative information* (arguments, difference of opinion, uncertainties). The outcomes of the expert consultation are processed in a multi-criteria analysis as well as in an account of (other) qualifications. Third the results of both desk research and expert consultation are discussed by a stakeholder jury that gives advice on the basis of all information: different than specialized expertise, a societal view deals with the political question of deciding what's important considering all specific aspects together. Finally both procedures

[2] By incommensurability we mean that these aspects do not share likewise measures that make comparison easy.

[3] Not discussed further here as it does not concern social scientific research.

are aimed at a well informed and substantiated decision-making by the CEH and/or policymakers (the Ministers of Environmental and Health policies decide on policy priorities presented in the *action-plan*; the CEH in close cooperation with policymakers decides on research priorities in the *hotspot selection procedure*).[4]

Research Practice

In this section, the practical work will be discussed. First, we will introduce the phenomenon of imperfection, which appeared to be an unavoidable companion along the way. Based on our research experiences, next, a typology of imperfections will be presented and illustrated by several concrete examples of imperfection from research practice:

Imperfect information: Limited data, in number, in type of information
Imperfect expert assessment: Limited number or diversity of experts, limited expertise, assessment uncertainty, difference of opinion
Imperfect processing of incommensurable data: Mathematical limitations, counter intuitive results, imperfect information base
Imperfect socio-political weighing/deliberation: Information overload, role confusion
Imperfect reflection on imperfection: Limited number or diversity of respondents, input constraint

[4] Readers who wonder why chronologically the procedure on research priorities was not instigated first: openness of our natural scientific colleagues and policy representatives to experiments involving 'other' actors end elements in the process only gradually developed during the first years of the CEH (Keune et al. 2008b).

Diabolic Dilemmas: Imperfection

By imperfection, we mean the discrepancy between on the one hand the calm and tidy ideal of ambitions, ideas, concepts, theories, methods, and on the other hand stubborn and messy complex reality blurring our view and troubling the ideal result, resulting not only in an imperfect grip on and understanding of reality, but also causing problems of interpretation. Moreover, the concept of imperfection supposes an idea of what would be perfect, or at least some idea about what would come close to the ideal. This is often problematic in itself and subject of intense scientific debate, let alone satisfying whatever standard that is taken as a yardstick. And if imperfection inescapably is part of the game, an important challenge emerges: how can we deal with it? If we cannot escape imperfection, should we then limit our ambitions? Or should we still try to live up to expectations and ambitions as much as possible? How can we decide if imperfection leads us nowhere or at least somewhere useful? We distinguish several kinds of imperfection that occurred during our case study work. We do not pretend to be exhaustive about imperfection; we mainly want to discuss the most pressing and important issues with which we struggled.

Imperfect Information

Is it legitimate to collect and use *imperfect information* in procedures concerning complicated socially important decisions? One of the main problems we faced during the desk research in *both case studies* was imperfect information, resulting in diabolic dilemmas. How can we decide

whether the information we were able to collect is sufficient as a basis for assessment? How can we balance recommendations from the quantitative and the qualitative research camps? One of the central issues seems to be representation: when may we scientifically conclude that the data we (were able to) gather are representative for the phenomena we want to evaluate? Is the answer (partly) hidden in numbers? Statisticians calculate exactly which numbers (of research participants, data) are needed to draw scientifically valid conclusions, thus 'implying' objectivity and certainty. They more or less refrain from scientific interpretation when for example the number of research participants is well below what they consider to be statistically significant, thus implying that kind of information not to be scientifically relevant or useful. In our social scientific work in the CEH we often continue when numbers are small compared to the standards in the natural sciences or quantitative social sciences. In fact, we do not automatically set out any quantitative standards for our work when we start. Is this wise? Is our mission different from the natural scientific one, in the sense of experimenting with opening up discussion and knowledge, and considering the explorative (thus more qualitative) nature of our work? Is our focus on different (social) perspectives (knowledge and opinions) different from investigating statistical correlations between substances (that behave similar here and elsewhere) and health effects? Are our possibilities thus more limited because of context specificity (people and their opinions and knowledge are context bound, pollutants behave the same everywhere...)?

When elaborating the criteria policy relevance and social aspects in the design stage of the desk research approach, together with scholars from the natural sciences and policy makers, *in both case studies* long lists of relevant aspects were written down with great enthusiasm. In the early conceptual stages of the a*ction plan* even *cost benefit analysis* was part of our ambition, one that received a warm welcome especially from policymakers: objectifying in monetary terms the costs and benefits of complicated socially sensitive issues seems the perfect cure for a lot of political nightmares. Unfortunately, experts on health issues and health care economics awakened us crudely from our dreams: unambiguously quantifying and objectifying such complex issues is unrealistic, let alone with limited resources. This seemingly perfect approach thus was already given up in the first stages of our journey to real life practice.

For the policy relevance criterion in *both case studies* we found an easy way out. This criterion by the way was not naïve of politics: policy relevance of course is largely dependent on political will and power arrangements. What was envisaged here was technical feasibility. What good would it do for public health to set policy priorities to issues for which no technical cure exists? Relieving ourselves from very extensive and complicated (highly specialized) document analysis and/or a long list of expert interviews seemed only fair considering our limited manpower and time horizon. In close consultation with policy representatives we decided to invest in a ranking questionnaire and thereby collecting our desk research material from the experience of the policy experts[5] 'in the

[5] A quick note on the terms policy expert/policy representative; indeed not indicated clearly so far. We work together with three types of (what we call) policy representatives. First the politically

field'. Due to practical limitations, thus a seemingly more perfect approach (extensive investigation of policy aspects) was exchanged for an expert elicitation approach. We will return to limitations of expert elicitation in the section '*Imperfect expert interpretation*'.

This leaves us with social aspects: indeed, quite a challenge for more than a century of social science (for an interesting account of problems and limitations of social science see Phillips 1973). When we started the design of the procedure of the *action-plan*, two key social issues of interest appeared to be relevant: *public concern* (risk perception) and *social priorities* (public support for environment & health policy). We considered these elements of interest for a minister when setting policy priorities: next to health risk and policy issues, what else would be important to be taken into consideration? Whose views does a minister want to take into account before reaching a decision and communicating the decision? This was considered not an exercise in power sharing (often public participation projects seem to have this rationale/promise) but in sharing perspectives and information, creating critical mass as a basis for well informed and substantiated decision-making.

Perhaps rather legitimate as key social issues, the actual capturing of these issues in desk research information is far from obvious, especially with limited resources. Moreover, in order to be adequate to our comparison of apples and oranges (specific research results in one case and specific research issues in another) we needed information on the basis of which differentiation and comparison between *cases* was possible. Ideally we would have had more time and resources in order to perfect our grip on these issues. In a perfect world this would perhaps mean that we could organize extensive public consultations and perception research, fine-tuned to the specifics of our endeavour and context. But even then, complicated choices would have to be made in order to decide how best to move forward. In fact, we may ask ourselves: can we objectively define perfection in this respect?

Ideals from a perfect world can be challenging and inspiring, in our research practice though, this turned out to be something completely different. Limited in time and resources, we had to find our way. Although originally designed for the *action-plan*, also for the *hotspot selection procedure*, more or less the same social scientific focus was used, be it that the concrete elaboration in research practice differed to some extent (Keune et al., 2009b, 2010). Types of information sources we used in the *action-plan* ranged from media coverage, to sources of risk perception: we interpreted the cases under discussion from the perspective of well-known factors that play an important role in risk perception, for example whether children are main victims of risks related to cases, and we used risk perception questionnaire results from the human biomonitoring project (Keune et al., 2008a). Furthermore we used key informants from the Flemish network of

appointed members of staff working directly for the Ministers: cabinet staff. We mainly discuss our work with them in the steering group attached to the CEH, rather than directly cooperate on a more regular basis. When reaching decisions, they are often at the 'end of the pipeline'. Second, the civil servants working for the administrations. We cooperate and discuss with some of them rather directly. Others play a role similar to the third group: experts from governmental expert institutions like monitoring agencies. They are mainly involved as experts in steering groups and expert consultation.

local health and environmental experts[6] (experts working in areas corresponding to cases were asked about their impression of local perceptions). In the *hotspot selection procedure* the same key informants from the Flemish network were used, but they were not the only ones. Different from the *action-plan,* we now directly looked for input from local actors. Furthermore, because of the character of the *hotspot selection procedure*, also at a Flemish level different groups and experts were involved (in proposing research options).

In *both case studies* our desk research results do not contain large numbers of respondents or participants; in fact, in some instances even only a very limited number... How legitimate is it to use such information, to compare cases, when cases sometimes differ enormously in numbers? Furthermore, the natural scientists often asked us with troubled faces, whether the results are representative enough... To be honest about the numbers: we conclude we don't know... and ask for other experts' help in order to create critical mass (see below). Furthermore we mainly do qualitative analysis: we concentrate on content, the types of information (such as types of arguments), rather than on numbers in specific contexts (like the number of certain arguments within certain case regions). Moreover, we include our own expert judgments in the information: the qualitative character not only shows in the information we use but also in our personal role as selectors, constructors and interpreters of information. The qualitative approach also shows in the expert consultation for interpretation of the information we collected.

Imperfect Expert Interpretation

How do you assess whether expert interpretation of imperfect information is robust or valid enough as a basis for decision making? How do you assess for example the quality and representation of the assessors? How do you value the assessment when for example opinions between assessors seem to differ quit substantially, or when the assessors tell you that their assessment has been hampered by (sometimes many) uncertainties? In a perfect world perhaps we are able to consult as many experts we want, from all relevant types of expertise, with all relevant knowledge on board, and without dispute or hesitation in their (preferably common) views. In a perfect world this will lead to crystal clear answers to our research question, leaving us without doubt and without much extra effort to put into in order to realize our ambitions. Of course this is not what happened.

In *both case studies*, we invited experts with as much relevant expertise as possible in order to give meaning to our desk research information from the viewpoint of prioritization. Because the information is neither complete nor perfect, nor self-evident and we lack an unambiguous or objective measuring rod, we strongly rely on the combination of individual expert assessments. This implies that the quality of the expert assessment is dependent not only on the imperfect information we provide, but also on the expertise of experts and their willingness and capabilities to weigh the information. The critical mass of experts consulted also comes in numbers of course: the more diverse and numerous,

[6] Part of the work of these experts is related to networking and contact with local people concerning environment and health issues. They are supposed to be well aware of local perceptions and concerns.

the more quality we expect from the joint effort. This ambition is more or less hampered by the fact that experts are very busy people and thus a lot of them simply do not have the time to cooperate. We tried to compensate this time constraint to some extent by offering a small fee, but our resources were very limited. Moreover, because Flanders is a rather small area, we have to fish in shallow water. We tried to compensate for this by also looking for experts in the Netherlands (same language area). This usually leaves us with rather small groups of experts per assessment criterion: ranging in our experience so far between 5 and 15 experts. With respect to diversity and relevance of expertise we have to be humble. The field of environment and health science and policy contains many highly specialized experts, but none with all overarching expertise needed to fully be able to judge all of our data and questions. Some know more about the environmental issues, some of the health issues. With respect to the social aspects, it's hard to find experts who combine relevant social scientific and environmental health expertise. Another practical limitation is that experts were not given the opportunity to take notice of the input of other experts, like in a proper Delphi round. The reason for this is that in negotiating our procedure in the beginning, our partners from natural science and policy considered a two-round approach too time consuming.

When we look at the inputs of the experts, we notice two other issues relevant for the quality of the assessment: one is the uncertainties expressed concerning their assessments; the other is dispersion of expert assessments. Uncertainties are diverse in character and sometimes are rather important. In the *action-plan* (Keune et al., 2009b), we distinguished the following types of uncertainty: lack of expertise; lack of knowledge within the scientific domain; lack of information in the desk research; lack of interpretability of the biomonitoring results; lack of clear sight on cause – effect relationships. Furthermore the dispersion of expert views is sometimes rather significant. Do these issues further disqualify the quality, or perhaps the robustness of the expert input? And to what extent is that so? Can or could we have done better, otherwise, or, is this it?

Imperfect Processing of Incommensurable Data

In order to process the combination of expert judgments, assessment criteria and cases, in *both case studies* we used a multi-criteria analysis (MCA) (Roy, 1996; Belton & Stewart, 2002;). Based on the expert assessments, individual expert rankings are produced. Pair wise comparison of rankings generates consensus rankings on group level. After this, relative importances/weights are to be attributed to the different criteria. MCA is not a miracle tool that objectively will solve all problems by unambiguously calculating what is best. It is more a kind of sounding board: it will structure and visualize the input of actors and factors involved. As such it will offer a basis for well informed and transparent reflection, learning and deliberation. Why then do we have to consider a MCA to have limitations in supporting human judgment and debate? Because of its dependence on the quality of the knowledge and opinions that fuel it (Dodgson et al. 2000).

In practice unexpectedly, at least for some of us, in *both case studies* another

type of imperfection crossed our path: even when the mathematical outcomes are to be considered dependent on the quality of information used as input and as such imperfect, we did not expect the mathematics itself to produce counterintuitive results to the extent that they even caused concern amongst some of the researchers. Counterintuitive results may be the result of limited information included in the analysis: the intuition of someone judging the results may take into account other information. A stakeholder representative may for example take into account issue relevant to the specifics stakes of his of her organization. Also results may be counter intuitive because someone judging the results may not be able to take into account all information processed in the MCA. Complex issues in which a choice has to be made between different options based on a diversity of facets (or criteria) demands more than the human brain can handle: it cannot assess all complex information at the same time (Miller 1956; Kagan 1988).

An intuitive concern with results may also be purely based on mathematics. For example when a consensus ranking is calculated based on several individual rankings, and the consensus ranking cannot be explained satisfactory when judging the individual rankings, then the calculation itself seems cause for concern. This may of course be caused by the incapability of humans to do complicated calculations without the computer. This may also be caused by the fact that the quality of the output is dependent of the quality of the mathematical input. When the rankings used as a basis for calculating consensus rankings are very different from each other, this may result in consensus rankings that are incapable of satisfying the intuition of the user. This does not mean that mathematical errors were made, but this also does not mean that the outcome is satisfactory or perfect.

Considering the above we may ask ourselves whether it is legitimate to narrow down complex issues and complicated and imperfect information about these issues to imperfect weights resulting in numerical rankings. We may also ask ourselves though whether we do have a better alternative.

Imperfect Socio-political Weighing/Deliberation

Desk research and expert assessments potentially provide a rich information basis for a stakeholder jury. One might expect (as we did) that offering full transparency is the obvious thing to do: in the *action-plan* step by step we provided information in individual contacts before the group meeting. The jury members of the *action-plan* in the end wanted to know everything (Keune et al., 2009b). Most of them even wanted to take into account complicated information such as uncertainties and dispersion of expert assessments. Our expectations of them addressing it all during the group discussion, nevertheless, were not fully rewarded as such. Articulating opinions on complex information interactively appeared problematic. The discussion was partly paralyzed by an overload of information and nuances. During the discussion they indicated it felt as if they had to *replay the expert* instead of responding from a social perspective on expert assessments. Based on this experience we took a different approach in the stakeholder jury of the *hotspot selection procedure* (Keune et al., 2010). Instead of focusing on full transparency while providing information, we focused

the discussion on a concrete advisory proposal, which could be commented upon. Some jury members who also took part in the *action-plan* jury were more satisfied, because they had to deal with less detailed information. Some 'new' jury members nevertheless as well as some 'old' members kept having questions about more or less the same type of complicated and nuanced information that was presented to the former (*action-plan*) jury . . . for example they asked for more detailed information on toxicological issues or were interested to learn more about scientific uncertainties and diversity in expert opinions. We may thus ask ourselves whether reducing complexity is wise while at the same time dealing with full complexity is hard to do. And again here we may ask ourselves: can we objectively define perfection in this respect?

Reflection on Imperfection

In *both case studies*, we invited participants in our research to reflect on the project. Moreover, we tried to evaluate the research experiences ourselves, with our colleagues from natural sciences and policy. This type of evaluation does not objectify any judgment of quality of the project, but it does collect a diversity of viewpoints on the projects workings and outcomes that can be informative. Here also we are dependent on the quality of the collection of individual evaluations, both in diversity in backgrounds, number of respondents and the effort they put into it. Ideally when asking for evaluations from participants, we would ask a lot of questions about a lot of information and detail, demanding quite some time of the busy schedules, and of course, they would be happy to do this all. In practice we cannot expect participants to do all of this. In fact, we may be happy with all the input we get. And per individual this mainly concerns part of the process they have been involved in, perhaps with some general questions on the project. We try to be open in our evaluative questioning but also guide them towards topics we are interested in from the project's perspective. Overall most participants are quite positive about the project, but some do indicate critiques and concerns on specific issues (Keune et al., 2009b, 2010). To some extent, the imperfections mentioned in this paper are addressed by some participants.

Discussion: Importance, Imperfection and Quality

In this section we will focus on some fundamental issues of knowledge assessment: how can we or others decide on the quality of our work? We will present some ideas from social scientific literature that may be helpful in this respect.

When the lead author of this paper presents his work at conferences, people often comment that the presentation was very honest. This always makes him wonder what a dishonest presentation would be like, and what the reasons behind such dishonesty would be. Being transparent on details about the work that has been done is merely normal for him. And this includes things that were difficult or imperfect. This is just part of the job. This may not be as comforting as supposedly perfect processes or crystal clear theories or models. But if dealing with complex reality per definition is imperfect and for that reason difficult, why avoid that which is an essential part of dealing with reality? It seems that dealing with imperfection is dealing with

complex reality. From a problem solving perspective, dealing with imperfection not only is bold, but necessary. An intriguing diabolic question remains, though: does the fact that reality is very complex imply that anything goes; that we and our directly involved partners are the only ones to hold us accountable for the quality of our work? The question that follows from this: on which basis do we judge quality? When we have the ambition of dealing with important but complex issues, and in practice information/knowledge about these issues is imperfect, we have to decide on the quality of imperfect information/knowledge with respect to our ambition: to what extent does the information/knowledge live up to the expectation of problem solving? Our experiences in *both case studies* lead us to some fundamental questions. How can we judge the quality of our way of dealing with imperfect but important information? When do we know enough and for which purpose? Who decides? Based on which knowledge, expertise, criteria?

Bryman (2006) concluded from interviews with social scientists that (with regard to the use of mixed methods) epistemological and ontological debates are not necessarily guiding principles in research practice. He proposes the research questions to be of main guiding importance, methods and philosophical debates may follow in the slipstream. Morin (2008) states[7]: 'the method emerges from the research'; here the word method is used in its original meaning as path, indicating that only in travelling the right method appears.

Still, this does not dismiss us of accountability to quality standards that critically assess whether we did our work well. Seale (1999a, b) takes up the gauntlet of quality standards but states research practice to be autonomous from abstract methodological and philosophical notions. Nevertheless, he does not degrade the abstract level to be useless, it may inspire (like the history of art may inspire artists), but must not suffocate the development of one's own style. This plea against literal application of theoretical notions in daily life does not deny any relevance of theories to practice-oriented improvisation, nor to the articulation of practice in theoretical terms afterwards. It does dismiss researchers of literally sticking to theory while finding their way through complicated reality. Perhaps this nuanced relation between theory and practice is an example of complexity itself.

Morgan (2007) puts forward three interesting qualitative notions of pragmatic approaches that seem relevant to our work. He builds on the work of Patton (1975, 2002) who (more extensively) introduced similar notions. A first notion, *abduction*, relates to the above mentioned theory – practice relation: instead of betting on either induction or deduction, there is rather a movement back and forth between the two that is more pragmatic to research reality. This relates well to our work in which abstract ideals are tested and nuanced and adjusted in practice all the time. Another relevant notion is the concept of *intersubjectivity*, which tries to overcome the dichotomy between objectivity and subjectivity. This notion is also of key importance in critical rationalism (Popper 1945) and in action research approaches (Boog et al. 2008). Examples of *intersubjectivity* from our work are the extended peer review

[7] In the words of Mara Selvini Palazzoli, quoted by Alfonso Montuori in his foreword to Morin's book on complexity.

approach in which we try to organize critical mass (also in evaluating our way of working) as well as the close interdisciplinary and transdisciplinary cooperation with our colleagues from natural sciences and policy making. Moreover does peer review in a self-critical research community (Seale 1999a) contribute to the quality of the research. The peer-communication beyond the context-specific level of case studies brings us to the third relevant notion, that of *transferability*, developed by Lincoln and Guba (1985): this concept does not conclude lessons learned in practice to be context-bound or generalizable per se, but looks for opportunities for knowledge-transfer.

Specifically with respect to evaluation of analytical deliberative (or likewise) participatory processes it was already stated that objectifying quality criteria is very difficult. Renn and Schweizer (2009) point out that the diversity of concepts and background philosophies is one of the reasons for this. Rowe et al. (2004) conclude that the complexity of participatory processes make it difficult to identify clear benchmarks for evaluation. Rauschmayer et al. (2009) stress the fact that such processes involve a diversity of actors, and as such a diversity of preferences, also from the point of view of process evaluation. This may lead to the fact that process outcomes are valued differently from different actor perspectives. They propose the use of participatory evaluation.

Navigating through complexity is the task social scientists picked up in the cases discussed here. As easy as it seems to design processes of structuring complex issues and decision making in which involvement of a relevant diversity of actors and factors is considered praiseworthy, as complicated it is, as we have shown, in practice to organize and live up to expectations. In judging the quality of this endeavor, diverse theoretical yardsticks maybe applied that will not necessarily do justice to the practical complications of research practice. The alternative offered by Morgan is one example that offers enough room for context specific manoeuvre for social science to balance theory and practice pragmatically. The three notions perhaps do not represent unambiguous quality criteria, but both the process and action orientation do provide tools to discuss and enhance the relevance of the research approach to the research questions as well as more general, methodological and philosophical issues. They connect well to the practical difficulties that we experienced in messy research practice. They do not resign to either dismissal of objective unambiguous quality criteria or to 'anything goes'. They look for a pragmatic balance between what seems both incommensurable and of value. And they look for *opening up debate*, of which this paper is hopefully a fruitful example, and so, we hope, is our work.

Acknowledgments

The Centre of Expertise for Environment and Health is being financed and steered by the Ministry of the Flemish Government (Department of Economics, Science and Innovation; Flemish Agency for Care and Health; and Department of Environment, Nature and Energy). We thank our colleagues from the CEH, both natural scientists and policy representatives for their fruitful cooperation. Also thanks to Ortwin Renn (University of Stuttgart, Germany), Silvio Funtowicz (Institute for the Protection

and Security of the Citizen (IPSC) European Commission, Ispra, Italy), Adri Smaling (University for Humanistics, Utrecht, The Netherlands) Harry Kunneman (University for Humanistics, Utrecht, The Netherlands), Lou Keune (University of Tilburg), Ann Van Herzele (Research Institute for Nature and Forest, Brussels, Belgium) and Michael Quinn Patton (Utilization-Focused Evaluation, Saint Paul, MN, USA) for their insightful comments on the text. And last but not least we thank two anonymous reviewers for their helpful comments.

References

Belton, V., & Stewart, T. (2002). *Multiple criteria decision analysis: An integrated approach.* Dordrecht, Netherlands: Kluwer.

Bilau M., Matthys C., Baeyens W., Bruckers L., de Backer G., den Hond E., Keune H., Koppen G., Nelen V., Schoeters G., van Larebeke N., Willems J., de Henauw S. (2009) The relation between the estimated dietary intake of PCDD/Fs and levels in blood in a Flemish population (50-65 years). *Environment international,* 35:1, p. 9-13

Boog, B., Slagter, M., Zeelen, J., Preece, J. (Eds.). (2008). *Towards quality improvement of action research. Developing ethics and standards.* Rotterdam, Netherlands: Sense Publishers.

Bryman, A. (2006). Paradigm peace and the implications for quality. *International Journal of Social Research Methodology*, *9*(2), 111–126.

Centre of expertise for Environment and Health. (2010). Retrieved May, 22, 2011, from http://www.milieu-en-gezondheid.be/English

Dodgson, J., Spackman M., Pearman A., & Phillips L. (2000). *Multicriteria analysis manual.* United Kingdom: Department for Environment, Transport and the Regions.

Funtowicz, S. O., Martinez-Aler, J., Munda, G., & Ravetz, J. R. (1999). *Information tools for environmental policy under conditions of complexity*. Copenhagen, Denmark: Environmental Issues Series 9, EEA.

Funtowicz, S., & Ravetz, J. (1990). *Uncertainty and quality in science for policy*. Dordrecht, Netherlands: Kluwer.

Guba, E. G., & Lincoln, Y. S. (1994). Competing paradigms in qualitative research. In N. K. Denzin & Y. S. Lincoln (Eds.), *Handbook of qualitative research* (pp. 105-117). Thousand Oaks, CA: Sage Publications.

Kagan, D. M. (1988). Measurements of divergent and complex thinking. *Educational and Psychological Measurement, 48,* 873-884.

Keune, H., Loots, I., Bruckers, L., Bilau, M., Koppen, G., van Larebeke, N., Schoeters, G., Nelen, V., & Baeyens W. (2008a) Monitoring environment, health and perception: an experimental survey on health and environment in Flanders, Belgium. *International Journal of Global Environmental Issues*, Special issue on Risk Perception and Social Trust, 8(1/2), 90-111

Keune H., Morrens B., Croes K., Colles A., Koppen G, Springael J., Loots I., Van Campenhout K., Chovanova H., Schoeters G., Nelen V., Baeyens W., Van Larebeke N. (2010) Open the research agenda: participatory selection of hot spots for human biomonitoring research in Belgium.

Environmental Health 2010, 9:33 doi:10.1186/1476-069X-9-33

Keune H, Morrens B., Loots I. (2008b) Risk communication and human biomonitoring: which practical lessons from the Belgian experience are of use for the EU-perspective? *Environmental Health* 7(Suppl 1):S11 doi:10.1186/1476-069X-7-S1-S11

Keune H, Morrens B., Loots I. (2009a) Natural scientific complexity from a social scientific perspective: environment & health research and policymaking in Belgium. *International Journal of Risk Assessment and Management*, Vol. 13, Nos. 3/4, 2009, pg 242-259.

Keune H., Morrens B., Springael J., Koppen G., Colles A., Loots I., Van Campenhout K., Chovanova H., Bilau M., Bruckers L., Nelen V., Baeyens W., Van Larebeke N. (2009b) Policy interpretation of human biomonitoring research results in Belgium; priorities and complexity, politics and science. *Env. Pol. Gov*. 19, 115–129

Lincoln Y., & Guba, E. (1985). *Naturalistic inquiry*. Beverly Hills, CA: Sage Publications.

Miller, G. A. (1956). The magical number seven, plus or minus two: Some limits on our capacity for processing information. *The Psychological Review, 101*(2), 343-352.

Morin, E. (2008). *On complexity*. New York, NY: Perfect paperbacks.

Morgan, D. L. (2007). Paradigms lost and pragmatism regained: Methodological implications of combining qualitative and quantitative methods. *Journal of Mixed Methods Research, 1*(48). doi: 10.1177/2345678906292462

Munda, G. (2004). Social multicriteria evaluation: Methodological foundations and operational consequences. *European Journal of Operational Research, 158,* 662–677.

Nowotny, H., Scott P., & Gibbons, M. (2001). *Rethinking science: Knowledge and the public in an age of uncertainty*. London, UK: Polity.

Patton, M. (1975). *Alternative evaluation research paradigm*. Grand Forks, ND: University of North Dakota Press.

Patton, M. Q. (2002). *Qualitative research & evaluation methods, 3rd ed*. Thousand Oaks, CA: Sage Publications.

Pereira, G., & Funtowicz, S. O. (2006). Knowledge representation and mediation for transdisciplinary frameworks: Tools to inform debates, dialogues & deliberations. *International Journal of Transdisciplinary Research, 1*(1): 34-50.

Phillips, D. L. (1973). *Abandoning method.* San Francisco, CA – Washington D.C. – London, UK: Jossey-Bass Publishers.

Popper, K. R. (1945). *The open society and its enemies, Vol II: The high tide of prophecy: Hegel, Marx and the aftermath*. London, UK: Routledge.

Rauschmayer, F., Berghöfer, A., Omann, I., & Zikos, D. (2009). Examining processes or/and outcomes? Evaluation concepts in European governance of natural resources. *Environmental Policy and Governance*, *19*(3), 159 – 173.

Renn, O. (2008). *Risk governance. Coping with uncertainty in a complex world*. London, UK: Earthscan.

Renn, O., & Schweizer P. (2009). Inclusive risk governance: Concepts and application to environmental policy making. *Environmental Policy and Governance*, *19*(3), 174 – 185.

Rowe, G., Marsh, R., & Frewer, L. J. (2004). Evaluation of a deliberative

conference. *Science, Technology, and Human Values, 29*(1): 88–121.

Roy, B. (1996). *Multicriteria methodology for decision aiding*. Dordrecht, Netherlands: Kluwer.

Schroijen C, Baeyens W., Schoeters G., Den Hond E., Koppen G., Bruckers L., Nelen V., Van De Mieroop E., Bilau M., Covaci A., Keune H., Loots I., Kleinjans J., Dhooge W., Van Larebeke N. (2008) Internal exposure to pollutants measured in blood and urine of Flemish adolescents in function of area of residence. *Chemosphere* p. 1317-1325.-doi:10.1016/j.chemosphere.2007.11.053

Seale, C. (1999a). Quality in qualitative research. *Qualitative Inquiry, 5,* 465-478.

Seale, C. (1999b). *The quality of qualitative research*. London, UK: Sage Publications.

Stern, P. C. (2005). Deliberative methods for understanding environmental systems. *BioScience, 55*(11), 976-982.

Stern P. C., & Fineberg, H. V. (Eds.) (1996). *Understanding risk: Information decisions in a democratic society*. National Academy Press, Washington D.C.

Stirling, A. (2008). “Opening up" and "closing down": Power, participation, and pluralism in the social appraisal of technology. *Science, Technology & Human Values, 33*, 262-294.