

DRAFT - DO NOT QUOTE



Governance of Responsible Innovation

GREAT – 321480



DEL.2.2. Theoretical Landscape			
Deliverable No.		D2.2	
Workpackage No.	WP2	Workpackage Title	Theoretical Landscape
Task No.	T 2.2	Task Title	Theoretical Review
Start Date:	01.03.2013	Revision Date:	31.12.2013
Authors		Sophie Pellé, Bernard Reber	
Contributors (advises, reviews)		Philippe Bardy; GREAT partners from NU, Oxford VTT.	
Status (F: final; D: draft; RD: revised draft)		D	
Distribution			
Document ID / File Name			

DRAFT - DO NOT QUOTE

Table of Contents

Executive Summary.....	5
Introduction.....	24
Chapter 1: The forerunners of RRI.....	31
1.1. Innovation.....	31
1.2. Corporate Social Responsibility (CSR) and Sustainable Development (SD).	33
1.3. Technological and Participatory Assessment.....	36
Chapter 2: Responsibility Turn in Innovation.....	39
Chapter 3: What responsibility?.....	48
3.1. Negative understanding of responsibility.....	49
3.2. Bypass of responsibility.....	51
3.3. Diluted and unlimited responsibility.....	52
3.4. Reduction to accountability.....	54
Chapter 4: Towards a positive Conception of Responsibility.	57
4.1. Task, role and authority.....	57
4.2. Capacity.....	58
4.3. Care.....	59
4.4. Process.....	61
Chapter 5: Normative elements of actual RRI approaches.....	66
5.1. Anticipation.....	67
5.2. Transparency.....	68
5.3. Responsiveness.....	68
Chapter 6: Norms in Contexts.....	72
6.1. RRI Governance.....	73
6.2. Determining the context.....	76
Chapter 7: Conclusion. Defending responsibility as a polysemic concept embedded in a contextual and reflexive governance.	85
7.1. Case Study: the limitations of INDECT project.....	86
7.2. Proposal for a positive participative, interactive and reflective conception of responsibility in context.....	88
References	92
Annex 1.....	97
INDECT project: An assessment.....	97
1. The theoretical background and problem hypothesis of the project.....	97
2. Method used for the analysis - parameters of analysis	98
2.1. Approach to ethics (tools and use of the tools).....	98
2.2. Role of ethical issues.....	99
2.3 Ethical issues in relation to the context.....	99
2.4. Which sort of presupposition?.....	99
2.5. Revision process - reflexivity.....	99
2.6. Which patterns of governance?.....	99
3. Results of the analysis.....	99
3.1. Approach to ethics.....	99
3.2. Role of ethical issues.....	102
4. Synthesis with some interpretation of those results	106
5. Recommendation for the project.....	107
6. Considerations for GREAT	108

DRAFT - DO NOT QUOTE

Annex 2: Paris Workshop, Methodological Meeting (22-24.04.2013).....110

DRAFT - DO NOT QUOTE

GREAT

Theoretical Landscape

Deliverable 2.2.

Sophie Pellé, University Paris Descartes-CNRS (Post-doctoral Fellow)

Bernard Reber, University Paris Descartes-CNRS (Director of
Research)

With the help of **Philippe Bardy**, University Paris Descartes (Lecturer
of English, Paris 1-Panthéon-Sorbonne, PhD Student in Philosophy)¹

¹ We wish to thank Philippe Goujon (Professor), Robert Gianni (Post-doctoral Fellow) from Namur University and the invitees of the Paris Workshop (see Annex 2) for providing help and advice on this text. We also extend our thanks to the reviewers from VTT and Oxford University.

DRAFT - DO NOT QUOTE

Executive Summary

The notion of Responsible Research and Innovation (RRI) emerges from the contemporary articulation between science, technology, economy, industry and society. The development of technology and research and the emergence of new issues in the twentieth century such as climate change, nuclear power, the precedents of Genetically Modified Organisms (GMO's), not to mention the scandal of bovine spongiform encephalopathy (BSE), which triggered a strong public reaction, have increased the need to find new ways for monitoring, controlling, organizing and shaping innovation in science and technology. RRI furthers the tradition which originated in bioethics, in ethical committees focused on various technologies (biomedicine, ICT) and in technology assessment (TA) practices. Furthermore, RRI continues the ethical reflection on technology and research, as framed by Ethical Legal, and Social Impacts or Assessment (ELSI or ELSA) initiatives which emerged in the early 2000s in the height of the controversy over the development of genomics. Added to that, the reflections on corporate social responsibility (CSR) and Sustainable Development (SD) from the last three decades seem to have paved the way for an embodiment of ethics into the shaping of technology. Technology assessment approaches have evolved to the point of integrating participatory approaches with a view to develop technology in a more democratic way, thereby allowing all stakeholders to discuss their moral assessments (moral intuitions, principles, norms, values), and even influence the development of technology. Participation has come to be seen as a way of broadening the set of normative elements that are required to make technology-related decisions. And this approach is in stark contrast with the limited ability of expert committees that usually stand for the values of other society members. The need to involve citizens, end-users, and, more generally, all stakeholders having an interest in technology is at the core of a new 'social contract' between science, society and technology. The fact of explicitly bringing the values of individuals (or other normative elements) into light in order to shape research and economic policies as well as scientific and engineering activities has contributed to raise public awareness on science and technology, and therefore

DRAFT - DO NOT QUOTE

strengthened the legitimacy of policies while increasing the social acceptability and ethical desirability for innovation.

The aim of this deliverable is to analyse and assess the various conceptions of innovation and responsibility that have been used in RRI existing theories. To this end, we first provide an historical account for the emergence of RRI and show the most critical limitations of CSR, SD, PTA approaches (Chapter 1).

Social Corporate Responsibility rests on the idea that private companies should not only take into account the interest of shareholders, but also include the interests of its stakeholders (i.e. employees, customers, suppliers, local communities, potential polluters, but also regulators, non-governmental organizations (NGOs), civil society organizations (CSOs) or ‘the public’ at large). Thus, private companies need to comply with national or international legal regulations, but also ethical norms. We show that the first limitation to Corporate Social Responsibility is that social responsibility is understood in a pure consequentialist way. This approach of responsibility is not satisfactory as it rests on a too optimistic vision of knowledge and rationality. An adequate conception of responsibility also morally engages individuals or organizations by virtue of their actions, regardless of the consequences. Moreover, as we will see, CSR is often reduced to a way of complying with existing norms such as in the case of innovation (like nanotechnology). Finally, there is no co-construction in relation with the context.

Whether economic, social or human, sustainable development (SD) implies that the use of our resources does not compromise the sustainability of natural systems and of the environment. Yet, innovation and SD seem to be in tension. On the one hand, SD is fed by innovation, such as in the case of renewable energy technologies that represent an ecologically sustainable innovation. On the other hand, sustainability already implies a substantive normative content according to which resources or biodiversity has to be maintained. We seek to show that the scope of RRI is much larger: while sustainability relies on existing norms of preservation, RRI aims at shaping the way we innovate and create new things and new ideas, taking into account the ways norms are considered and assessed.

DRAFT - DO NOT QUOTE

Finally, we turn to Technological Assessment and Participatory Technological Assessment that include different approaches (Constructive Technology Assessment (CTA), Real Time Technology Assessment (RTA), value sensitive design, risk assessment, the precautionary principle, New and Emerging Science and Technology approaches (NEST)). Technology Assessment approaches are based on impact assessment, forecasting, scenario analysis or consensus conferences, and can involve around fifty different devices. Despite of their merits, we will highlight the different problems encountered by these approaches: first, the “capacitation of actors” involved (especially the ordinary citizen), second, the obstacles in communication: a) to find the appropriate learning process to face the diversity of the public, b) to be skilled enough, as experts, to translate sophisticated knowledge in interdisciplinary arenas, c) to compare the different assets behind the choice of neutrality or plurality in the selection of citizen and experts. The third problem is related with the confinement of these mini-publics (as it is impossible to include all stakeholders). Finally, PA and PTA raise the issue of the efficiency of the norms settled during the process (i.e. establishing suitable conditions for making the required changes). Indeed, it is not because the expression of a particular norm is sound and justified in an ethical way that it will be adopted and implemented. As we show with the INDECT project case (Annexe 1, chapter 7.1), the identification of the norm does not implicitly describes the methods or mechanisms required for its expression within a project.

Here, a remark about innovation and research might be necessary. In many discourses of decision-makers, research feed innovation in creating new possibilities for industries and job opportunities. However, a closer examination shows that research and innovation differ on two levels. First, they work on different timescales. Research time is slower than innovation time. More precisely, we have to distinguish two different cases: research whose aim is to make innovative processes or products available on the market, and research whose purpose is to assess the impacts of products and processes. For instance, it takes more time for research to evaluate GM organisms than to have a robust long-term knowledge of their impacts on environment or health.

DRAFT - DO NOT QUOTE

Second, research may focus on the assessment of the effects of innovation and, more precisely, on its adverse and dangerous effects. Responsibility will not be the same if research is simply behind innovation processes or if research is broadly opening the scope of medium and long term consequences of innovation.

After a brief general presentation of what innovation is and a genealogy of the concept of RRI (chapter 1), we highlight some of the characteristics of current conceptions of innovation in order to show that they help to understand the notion of responsibility in relation with the process of innovation (chapter 2). We identify here a “responsibility turn”. Innovation flourishes out of new technological or scientific possibilities (new ideas, new conceptions, new tools, etc.) and must have the power to satisfy a need in order to generate some benefits. Innovation is thus supposed to feed growth in reducing costs of production, in rising the quality of services or in creating a new demand for product and services. In a first understanding of innovation, responsibility plays a limited role: it can be invoked in cases of **noncompliance** with contracts and more generally in relation with **legal constraints**. However this understanding of responsibility as **liability** is limited and only covers a restricted dimension of the problem.

Indeed, some recent evolutions of the way we innovate could help to integrate responsibility into the processes of innovation. For instance, in new areas of interest, out of the traditional distinction between product and process innovation, the emphasis has been made on the possibility that the “mental models” of the organisation change to propose not only new products, but a whole new conception of the function that products have to serve. This conception – paradigm innovation – opens the possibility of reframing the relation between individuals and products. Moreover, it shows that actors of innovation have a crucial capacity to invent new processes by which a specific function is satisfied. This will have particular consequences for responsible innovation. In the same vein, innovation has been recently conceived as a collective process – a “multiplayer game” – involving a whole network of actors, a complex system of interactions between them and an institutional, social and political environment. Governance approaches of responsible innovation precisely seek to organise the possibility of a collegiate design of innovation. For instance, open innovation

DRAFT - DO NOT QUOTE

raises new organisational issues related with the need to coordinate an important number of more or less active contributors, and to organise how knowledge is shared among them. These recent evolutions of innovation have had the interesting consequence of promoting an early involvement of end-users in the design and conception of products. Participatory design practices, where end-users co-conceive the product with engineers or designers have emerged. Finally, innovation introduces gaps and ruptures in the lifecycle of products. In some cases, and even if some ruptures will always remain unexpected, it is even possible to forecast and integrate the evolution of the product as it has been promoted by sustainable development practices.

Therefore, some of the elements that intrinsically constitute innovation already imply a form of responsibility. Innovation is a dynamic process and intrinsically, it is about change. Successful innovators show a high capacity to adapt to their changing environment and to social, institutional, technical or ethical constraints. This dynamic aspect of innovation involves a key element of RRI: **responsiveness**.

To sum up, we emphasize that innovation is conceived as a complex process that results on the one side, from forces that favour a limited conception of responsibility (when it is **reduced to liability** or when innovation is only driven by the quest for economic benefits without taking other parameters into account). On the other side, innovation practices evolve rapidly, and recently, they relied on the need for a co-shaping of technology and products including the persons to whom they are designed. In addition, current understanding of innovation emphasised the role of **responsiveness** in successful organisations. These elements can be taken as building blocks toward an understanding of responsible innovation governance. In itself, innovation contains the possibility of integrating adaptive processes of conception and production, which is one of the pillar of RRI. Beyond that, these elements show that a possibility of intertwining responsibility and innovation exists, as opposed to traditional conceptions.

To ensure the transition from a potential link between innovation and responsibility to an intrinsically normative concept of innovation in research we conduct a specific inquiry

DRAFT - DO NOT QUOTE

(chapter 4) into the meanings (and their potential relationships) of responsibility when related with innovation. Our aim is to reach the normative conditions needed to ensure a responsible way of innovating. Indeed, it is important to clarify the conflict of interpretations surrounding responsibility, the implicit blind spots in various RRI approaches and the possible ways to articulate different understandings of responsibility.

First, the various interpretations of responsibility - more precisely moral responsibility - that have been developed for decades by moral philosophers are of different relevance to understand the problem of RRI. For instance, the **causal** (logical) dimension of responsibility has to be distinguished from dimensions of **blameworthiness, liability or accountability** (here, some authors oppose causality and moral responsibility, while others defend compatibilist positions). Or, next to definitions that insist on **sanctions**, other understandings claim for a focus on **positive capacities** such as **care** or **responsiveness**. However, chapter 4 shows that these approaches either rely on an external, negative and retrospective conception of responsibility that empty the notion from its content, or promote a positive, internal and prospective conception that rests on a substantive definition of the good.

We first considered the most common meaning of responsibility, which is closely related with someone's wrongdoing and with sanction (the first historical root of responsibility). Responsibility is conceived as imputed to someone for his or her actions, whose negative outcomes or harms have to be compensated or **repaired**. In this **legal-oriented** interpretation of responsibility, it is possible to distinguish between **blameworthiness** (when A can be blamed for an outcome X, for instance, a car accident) and **liability** (A is liable to pay for the damages caused by outcome X). In both cases, someone is held responsible for her actions or decisions that happened to break the law or to infringe a social or a moral norm.

Although this conception of responsibility is a bedrock of social order, it encounters several limits that come from its general **neglect of the internal capacities** of individuals to mobilise their will to act in a responsible way. It is backward looking and relies on a norm coming from outside that has the potential of influencing someone's action through the threat of

DRAFT - DO NOT QUOTE

sanctions. This leads to a misconception of responsibility in at least three ways. Responsibility is (a) bypassed, (b) diluted, and (c) amalgamated with accountability:

a) Focusing on the possibility to impute future damages on the basis of the available knowledge contributes to build a perspective of responsibility that is purely instrumental. There is no normative involvement of actors as the only driver of their behaviour will rely on the **fear of financial or legal penalties**.

b) The second type of problems derives from the **individualistic** overtone of negative and backward looking interpretation of responsibility, based on a strong linkage between individual and outcomes. While considering innovation or research, it is often difficult to isolate who is cause of what. This problem, sometimes labelled as the “**many hands**” issue also results from the future being uncertain or ambiguous, the consequences of emerging technologies being often impossible to forecast. Again, the purely **consequentialist** approach of responsibility collides with its own frame: in seeing responsibility as the result of a calculus (the assessment of the outcomes), one is confronted with the time, space and interactions limits that seem reasonable to assess guilt.

The complexity of the problem is illustrated by the shift that occurred in our use of terms such as ‘Responsible Research’ and ‘Responsible Innovation’, which relate to the field of technology, applied science and engineering. From the purely individualistic interpretation of responsibility, we moved to a conception where the adjective “responsible” is now also ascribed to **the complex network** of actors, institutions, public policies that is entailed in an innovation process.

c) The **sanction-oriented interpretation of responsibility** shows that there is a conceptual displacement from imputation to risk by which responsibility ends up as conflated to accountability. Indeed, the idea of solidarity against risk that led to the advent of insurance systems in the 19th century and to 20th century’s welfare state contributed to alter the pure understanding of responsibility as implying **obligation and repair in the case of fault**. The institutionalisation of the management and prevention of social risks (by means of insurance

DRAFT - DO NOT QUOTE

and social-security systems) replaced the reparation of an individual fault. As for the dilution and avoidance of responsibility, the conceptual reduction of the analysis into a paradigm of accountability and risk prevention, implies a consequentialist framework that is confronted with the limited possibility of evaluating outcomes.

This would call for a **more “flexible” approach of responsibility focused on adaptive processes (“responsiveness”)** as suggested by the Latin word *respondere*.

To sum up, although liability and blameworthiness are essential to uphold social order, the problem with the negative understandings of responsibility is that at no time, the positive ability of individuals to act in a responsible way is called on. They reduce the **epistemological** relevance of such a conception of responsibility but also its **normative** validity and finally its **practical** power to influence individuals’ action. Indeed, neglecting the possibility that individuals engage their responsibility in a prospective way would raise an epistemological issue since it rests upon a misconception of the whole range of attitudes and behaviours actors can adopt towards their actions. Moreover, the normative or ethical validity of such a conception can be questioned as it excludes the ontological link existing between our actions and our responsibility. Yet, engaging freely and with a relevant set of knowledge in a course of action makes me ontologically responsible for this action. Finally, the practical relevance of a purely negative conception of responsibility is also challenged since the threat of actions is by no mean the only driving force of human action. The hope and want to act in a responsible way can be strong incentives that have to be analysed and favoured.

For such reasons we have considered responsibility in a broader meaning (chapter 4), to avoid reducing it to liability and blameworthiness. The condition of wrongdoing, for instance can be extended to other understandings of responsibility as we can also be held responsible for positive outcomes. The same kind of reasoning applies for the condition of moral capacity or freedom. In negative understandings of responsibility, individuals are supposed to exert **a moral capacity and to be free to act**. The negative perspective never draws out of them all the potentialities they contain to conceive individuals’ decision and actions in a closer connection with ethics. In this respect, **liability, blameworthiness or risk prevention do not**

DRAFT - DO NOT QUOTE

offer a comprehensive approach of the practices, activities and capacities that lies behind the concept of responsibility.

In this context, the current conceptions of responsibility that focus on **responsiveness, care, or moral capacities**, offer a way to overcome some of the difficulties we pointed out with negative meanings of responsibility. More positive and prospective understandings of responsibility assume that individual not only pay for the (possibly wrong) things they did but engage in a process through which they take care of others (other human beings, future generations, non-human beings or the environment). In this sense, positive meanings of responsibility will provide with relevant foundations to a conception of RRI. However, we do not imply that negative understanding of responsibility should be discarded. Only that they have played a prominent role up now, when positive meanings should also be called on. In this respect, both types of understanding of responsibility, in answering different questions, plays a complementary role.

Several meanings lie behind positive understandings of responsibility in RRI. First, there is a kind of responsibility involved when somebody is given a **specific task** or a **specific role**. Individuals are assigned to specific activities and to ensure that they operate in the best possible way, whoever is responsible positively and actively mobilises his or her knowledge of the relevant set of rules and norms, as well as her capacities for action and anticipation. There is also a meaning of responsibility related to **authority**. In their professional activities, individuals are responsible for ensuring that definite tasks are performed, and expected outcomes avoided or favoured. Their responsibility, here, covers a widened set of activities, compared to that included in the definition of a task, as it also implies other individuals' actions and decisions.

Second, assuming **moral agency** implies that we have the ability **to reflect on the consequences** of our actions and that we can engage in a **foresight** exercise by which we increase our knowledge about the world and how our actions might interact with and alter it. This positive capacity also implies the ability to form **intentions**, to act deliberately, and to act in accordance with certain norms and moral or legal rules. Blurring the forward-

DRAFT - DO NOT QUOTE

looking/backward-looking distinction introduced by negative understanding of responsibility, these understandings of responsibility offer a step further compared to legal interpretations in promoting a positive capacity **to commit oneself** to actions and decisions. .

The third route towards a positive account of responsibility sought to move from a purely consequentialist framework, in focusing on the **virtue dimension of responsibility**. This approach rests on a more realistic conception of individual rationality since decisions do not result from a pure rational calculus but are justified by routines and by a constant adaptation to the requirements of the situation assessed by the individuals.

These three interpretations of positive responsibility in RRI form a first group of propositions that have some weaknesses: Are norms of responsibility established by expert discussion between ethicists, philosophers, sociologists, etc.? Do they result from a collective process? And in this case, how can we hope for an agreement on comprehensive doctrines of the good in a context of moral pluralism of values and theories? Are they imposed by a benevolent dictator? Once again the issue of the construction of norms of responsibility is eluded.

We have presented a second group of positive approaches of responsibility, that escape purely retrospective and negative conception of responsibility in focusing on the dynamic of responsibility, on the ability to **adapt** and change one's own action. Two dimensions are highlighted: **accountability** and **responsiveness**.

Accountability is linked with the possibility of providing a justification for one's action as when we have the moral obligation to account for what we did or for what happened. This would be a first passive way of conceiving accountability as a mechanism that focuses on the relationship between a forum and an agent. In such a conception the emphasis is made on political and social control and the task of accountability studies will be to explore whether there are such relations at all, whether these can be called accountability mechanisms, how these mechanisms function, and what their effects are. There is another active conception insisting on the process of learning opposed to the "mechanism of control" by which

DRAFT - DO NOT QUOTE

individuals learn to be responsive to each other and to adapt their behaviour in order to achieve substantive standards of good governance.

With care and responsiveness a step has been made towards a positive understanding of the notion of responsibility. Indeed, both conceptions introduce the possibility for actors to adapt their behaviour and decisions to the situation and to **revise** their judgments according to norms. Social actors of innovation and research are responsive in a way that ensure the efficiency of their practices (for instance economic success or scientific praise) but they can also be responsive in a sense that they adapt their behaviour to certain ethical norms (including the avoidance of bad effects, or the wish to better answer ethical needs, as in the example of biomedicine).

However, the limit we pointed out for negative definitions of responsibility remains. How the norms of “responsible” acts or intentions are settled? By whom? How is it ensured that individuals will follow them? Current definitions of care and responsiveness let these issues pending. Moreover, virtue ethics suffer from its essentialist overtone, as it does not manage to face the issue of the diversity of the conceptions of the good showed by individuals differing in their culture, in their political and religious beliefs, etc.

To sum up, the positive definitions of responsibility add something to the juridical-inspired understanding of the word as they all insist on an ability of individuals or systems to respond to the values and moral conceptions of those who are concerned by innovation and technology. Moreover, they also imply a prospective concern for the future and the possibility to adapt the pathways of technological development according to this (normative) horizon. However, their answers are limited as they all fail to address the crucial issues of the way in which the norms of responsibility are settled. They all promote different solutions in order to favour responsible actions but the way in which what is collectively considered as responsible is elaborated remains completely obscure.

Both current negative and positive definitions of responsibility neglect the central issue of the construction of norms as the latter are supposed to be given from outside. In the model of the

DRAFT - DO NOT QUOTE

sanction, they are given by law, or by routines that establish the amount of the financial or legal sanction. In the case of care and accountability, they rely upon some sort of virtue ethics that determine what the goods practices or activities are, yet in an abstract and *a priori* way, disconnected from actors' values and norms. Finally, in the case of responsiveness, the norms of what is responsible (and what is not) have to be defined, the question of how they are determined, by whom, etc., being left open.

In conclusion, all these approaches rest on a conception where individuals do not participate to the construction of norms regulating their decisions, although this would favour the possibility of their application. Individuals are either supposed to act instrumentally, under the threat of sanction or to have a sense of what is the “good” - sense - which we never know where it comes from. The precise mechanism by which individuals follow a norm imposed from above or happen to know the substantive definition of the good promoted by a specific approach of responsibility is completely left in the dark. Finally, both responsibility governed by sanction or ruled out by positive abilities rely on top-down approaches where, on the one side, individuals are compelled to act in a certain way, or, on the other side are told what value to follow by an essentialist framework.

Chapter 5 analyses RRI governance approaches and studies the different conditions that innovation and research practices should satisfy in order to ensure a responsible pathway of economic, industrial or social development. Despite the issue is rarely developed, this chapter aims at scrutinizing the conceptions of **governance** underlying recent RRI approaches. We critically presented several elements identified in RRI literature along with the frameworks proposed to implement them. Beyond their differences, these perspectives on RRI agree on several elements that we identified as the **five ingredients for RRI**. 1) **Anticipation**. Researchers, policy makers and other members of society have to conduct anticipatory research to think through various possibilities to be able to design socially robust agendas for risk research and risk management. 2) **Transparency** favours a richer dialogue where the various interests and visions of the world of the community members can be expressed and are taken into account. 3) The third component of RRI is **responsiveness** understood as the coupling of reflection and deliberation to action that has a material influence on the direction

DRAFT - DO NOT QUOTE

and trajectory of innovation itself. 4) The capacity of a system to adapt and to change during its course of development can be identified as its **reflexive** stance. Reflexivity asks researchers and innovators to think about their own ethical, political or social assumptions underlying and shaping their roles and responsibilities in research and innovation as well as in public dialogue. Reflexivity should raise awareness about what we call the issue of “framing” and its possible solutions. 5) The **inclusion** of stakeholders into the making of the norms regulating scientific practices and innovation processes should achieve several goals. It should help a) defining and revealing what are the actors’ values and the ends and purposes they assign to science and technology, b) co-establishing norms from these values, c) shaping the design of innovation and research processes and outputs.

Now, does RRI bring something new regarding these conditions which have been already assigned to PTA? The answer is yes, because of their early combination in the process of development of technology and research. RRI promotes the institutionalized coupling of such integrated processes of anticipation, reflection and inclusive deliberation to policy - and decision-making processes – therefore favouring responsiveness.

Chapter 6 is devoted to the analysis of the processes by which responsible innovation is conceived, assessed and implemented. It aims at challenging the RRI approaches proposed so far in light with the relation they build between norms and their contexts. For instance, we already mentioned that participation is one of the most important pillar of responsible innovation and research approaches. However, beyond its pure theoretical weight, it is necessary to consider how it can be implemented, practically. Those questions cannot be disentangled. For instance, issues of transparency and anticipation cannot be dealt in an abstract and decontextualized way, since the schedule according to which information is disclosed, and the content of the information must be subjected to a form of social consensus. In a way, this contributes to make a priori approaches of responsible innovation and research processes irrelevant because transparency and anticipation are required in an abstract way that does not reflect on the practical constraints of the context. Analysing RRI definitions implies examining their governance characteristics simultaneously. And this will mean that even the positive understandings of RRI rely on norms that are disconnected from individuals’

DRAFT - DO NOT QUOTE

normative horizon. In this sense, they do not provide with any answer to the practical issue of the implementation of norm, understood, in a first approximation as the “efficiency” of norms².

In addition of the question of the implementation of norms, another issue raised by most approaches of RRI concerns the way in which the norms framing and regulating innovation and research practices are collectively decided, in order to achieve “ethical acceptability” and “societal desirability”. Interestingly, some of the studied conceptions of RRI, although claiming for collective processes in research and innovation, defend substantive (as opposed to proceduralist, for instance) ways of defining the norms shaping individuals actions and decisions. These conceptions have the practical advantage to rely on norms that result from a consensus, at least at the European level of decision. However, they suppose that the agreement on a general **normative horizon** (embedded in the European ideal of sustainability for instance) has already been achieved.

For these reasons, we sought to analyse how norms can be elaborated in their context, in a way that does not already presuppose the boundaries of the problem. At an epistemological level, it means that the problem underlying the construction of a norm should be opened up (and not be already settled), leaving room to modify the question it raises, the data at stake, its warrant, and the possible exceptions to this norm as well as their modalities of application (to borrow from Stephen Toulmin’s model of argumentation). In order to sketch the main line of a governance approach that will help us to critically review RRI theories, we began to reflect on the process of construction of the context and on the role of reflexivity. We found that governance, in the context of uncertain technology and research development, can be seen as an attempt to answer a “trilemma” between scientific accuracy, policy effectiveness and political legitimacy, i.e. between the rules of scientific knowledge, the efficiency of political norms and rules and their social acceptability.

² The issue of the efficiency of norms is very complex and could be a research question on its own. Del 2.3 aims at investigating some important elements of this issue. A special session will be devoted to discuss this point in the Oxford meeting (January 15-16 2014).

DRAFT - DO NOT QUOTE

Conciliating these different elements requires that we know in a much more concrete way how the process of setting the norms of responsibility is thought in relation with the context. Indeed, understanding the relation between norms and their context ensures that individuals reach an agreement on their interpretation of norms (and not only on norms themselves) and that they possibly will follow them. Here, we emphasized that any process of construction of norms, a fortiori the norms of responsible research and innovation, will have to deal with moral pluralism, where individuals can have conflicting values on definite subjects but also ground their normative horizons on different ethical theories (i.e. consequentialist frameworks, deontological theories, virtue ethics, or some forms of intuitionism, to mention the main ones).

Chapter 6 also pointed out some of the limits of proceduralism, which is often invoked as a sound way to define collective norms. The problem mainly lies in that proceduralism presupposes that the discursive and rational process of norm's construction that could be considered to be relevant is capable by itself of taking into account all the possibilities that are available for a social context to be regulated. But this presupposition is highly problematic because it ignores the fact that the concrete choice of a norm, even when judged relevant following a discursive process of reason, necessarily results from an operation of selection of the possibilities at stake. Moreover the context is not a pure exteriority. It is never simply given but rather constructed by individuals.

Indeed, context, as it is conceived in GREAT project, is not just what we see outside, i.e. the environment in which the decision is to be made. It is not just the problem of what one can perceive as the environment, since **we address the environment from our own perspective and framing**. This will mean that, to elaborate norms in taking the context into account, we will have to study the cognitive framing of individuals. The cognitive framing sets boundaries on the parameters of discussions among stakeholders: it partly determines the ways in which dialogical engagements progress. Individuals think about norms from their particular point of view. They construct their context, and this will affect their positions on norms. Thus, we need to understand the ways in which agents conceive of their own possibilities from which they will elaborate the norms of responsibility. Finally, to be complete, we have to distinguish

DRAFT - DO NOT QUOTE

the **descriptive and the normative part of the context**, the normative dimension helping to cast a critical look on the descriptive one.

To sum up, the idea of **participation and deliberation** is not new and is anchored in a recent tradition of political philosophy and political theory. It has often been presented as a way of dealing with the issues of moral pluralism characterized by a double rejection of monism and relativism i.e. by the acknowledgement of the positive role of conflicting values and value systems among individuals, on the one side, and by the need to answer normative issues inside the realm of ethics, on the other side, not delegating it to group loyalties, cognitive bias, interests, religious or national particularities (c.f. glossary Pluralism and Deliberation). But the question of the real efficiency of deliberation and participation processes is not raised within contemporary conceptions of RRI. Rather, the involvement of stakeholders is presented as the solution that *in fine* will warrant the responsible side of the innovation and research processes. Therefore, RRI approaches do not fully address how participative and deliberative process will be efficient and will effectively shape the design of technology in a way that is ethically and socially acceptable, because they all presuppose their own required conditions and as such do not necessarily involve reflexivity.

These approaches are problematic for they **presuppose the capacity of reflexivity** of the actors to be already existing due to a formal method, such as argumentation, deliberation, debate or discussion. They never challenge the concrete steps by which deliberation will lead to the elaboration of norms and eventually to their following. However, in order to conceive in a more appropriate way our relation to the context, we need to introduce the possibility for the agents to be reflexive and to revise not only their judgments, but also the way in which they size and understand the problem (its epistemic and normative dimensions). The possibility of revision is an important bet in deliberative theory of democracy. Indeed if we don't agree to change our mind in front of better arguments it is useless to enter in such process. Better continue on other ways using bargaining, or force reports. Here, we see that an appropriate conception of reflexivity will rely on a **theory of learning**. We refer to the capacity of actors to identify the various effective possibilities on which the operation of the selection of the norm will be carried out. **Actors not only reflect on the adequacy of their**

DRAFT - DO NOT QUOTE

norms and values, but also on the way in which they construct these norms and values.

These norms and values can be focused on what is right – or false- (epistemic norms) or what is good, just or evil, unjust.

In this sense, governance will not only have to manage or articulate different spheres (politics, civil society, research, industries) but it has also to articulate different spheres of knowledge (ethics politics, economics, science) with their proper ways of relevance, methods or argumentation types. Reflexivity does not only concern the inter-individual or inter-institutional interactions, but also the inter-epistemic one.

And if some conceptions of reflexivity that is put forward in some RRI approaches are not blind to the issue of the “framing”, they do not go as far as required by the definition of the second order reflexivity we will develop in DEL 2.3. The role of the context (including the way in which we conceive the issues at stake) is not fully taken into account, and the possibility of revising not only our judgments in front of a problem but the very manner in which we conceive the problem is not seriously investigated.

In conclusion, with the definition of second order reflexivity (chapter 6), it appears that however rich and innovative, compared to traditional models of technology management, the different approaches of RRI leave some important issues in the shadow. First they do not question how the problem is defined. Secondly, they do not deal with the legitimacy and implementation of norms as they do not ensure that the participative and deliberative process provides with norms that the members of society will find acceptable, choose and follow. There is no “opening” of the framing (descriptive and normative), i.e. no interrogation on the way in which the precise context of RRI is constructed.

In our conclusion we have used part of our results to analyse the limits of the INDECT European project. Indeed, the INDECT project could be considered as “responsible” for it meets some ethical norms considered as desirable among the European territory. To extend its ethical relevance, our enquiry has explored the impacts of a new technology of detection being generalized within the EU. It appears that involving experts of different disciplines does

DRAFT - DO NOT QUOTE

not amount a real inclusive enquiry entailing a comprehensive participation of stakeholders. Moreover, the project rests on a top down model of governance that supposes guidelines to be decided by expert and then being applied by members of the project. There is no reflexivity³ concerning the process in which the ethical norms are obtained, and no involvement of potentially relevant stakeholders for a future wide implementation of the technology. Then, as the level of deliberation is poor, as there are only rare attempts to justify and produce arguments grounding norms, and as there are no inference between reasons and decisions, we are far from a reflexive inter-actor governance of inter-institutional governance.

More generally, we proposed, in deliverable 2.2, a broad and open conception of RRI which offer the theoretical basis to implement a reflexive empirical analysis. **In taking seriously governance with reflexivity in context, we aim at departing from *governance of RRI* to *responsible governance of RRI* and to move from *sciences for society* to *science within society*.**

We conclude this deliverable by mentioning three layers underlying our conception of RRI, whose further analysis is the scope of D 2.3.

Participation (or commitment) would be the first layer towards a more positive definition of responsibility. The innovator, as an author asks for an intellectual property right. In the same way, responsibilities are attached to him as an author.

Secondly, when we act and therefore participate, we are not passive. Our freedom and our ability to revise judgments are engaged. This capacity of revising our judgments, which is very broadly conceived as our **reflexivity**, and which concerns both individuals and institutions⁴, makes our action the fruit of a conscious decision-making process and opens the

³ According to our analysis of the INDECT documents that are available.

⁴ Here, we can consider stakeholders that are often called 'vulnerable', i.e. people who are injured in their capacity of revising their judgment and making conscious decisions, such as, elder people with many ailments. This particular stakeholder's group is at the core of one of the 'Societal challenges' addressed by EU FP7 funding policies: the *aging of society*. To unable vulnerable people to enter in a process of reflexivity, it is, for instance, possible to involve people who are very close to them and who can defend their interests in a decision-making process.

DRAFT - DO NOT QUOTE

way for us to understand the weight and load of the potential consequences and outcomes of our actions.

Finally, our actions do not only affect us but also others, whether human beings, future generations or the environment at large. Therefore, we cannot avoid the **interactive (intersubjective)** aspect of responsibility that comes from the normative horizon in which the subject is embedded.

Of course, this proposition – drawing a more adequate responsibility concept on three elements: participation, reflexivity and intersubjectivity – is only very schematically sketched in this deliverable. It requires further development and rationale, which will be the aim of next deliverable. To this end, d.2.3. addresses in depth the issue of the limits of proceduralism, using Maesschalck and Lenoble's approach, which developed a rich reflection about the relation between norms and their context.

DRAFT - DO NOT QUOTE

Introduction.

For over a century, innovation has been claimed to fuel economic growth and human progress and to guard economies against the threats of the steady state envisioned by classical economists (David Ricardo, for instance). In parallel, the development of technology and research and the emergence of new issues in the twentieth century such as climate change, nuclear power, the precedents of Genetically Modified Organisms (GMO's), not to mention the scandal of bovine spongiform encephalopathy (BSE) that led to a strong public reluctance to, and outlawing of, new commercialisation in the EU⁵, have increased significantly the need to find new ways of monitoring, controlling, organizing and shaping innovation in science and technology. In the seventies, technology assessment approaches were designed to help governments to better anticipate the social consequences of science and technology and increase the public understanding of science through discussion.

In this process, the task of raising ethical issues and the attempts to address them has been left to human and social scientists and expert committees, among others, as illustrated by the emergence of bioethics. In this field, ethical committees (often gathering lawyers, theologians, religious experts of recognized religions, and actors from the medical field) were entrusted to offer some serious consideration on complex subjects mostly related to the boundaries between life and death⁶. In parallel, applied ethics, moral philosophy or moral sociology have seized upon these issues offering different theoretical perspectives to address the issues raised by innovation in biology or, in recent times, information and communication technologies (ICT). In addition, institutions such as the US Office of Technology Assessment or the European Parliamentary Technology Assessment were created to advise parliamentary bodies confronted with the new scientific and technical challenges in the second half of the 20th century. Such practical and theoretical tools, designed on the basis of the critical reflection of

⁵ After the MON810 maize had been approved by the European Commission in 1998, the European Union enacted a *de facto* moratorium (which was not a regulatory decision) freezing all new additional authorization of commercial cultivation of GMO's between October 1998 and May 2004. The GMO regulation in EU is therefore sometimes considered as stringent as it limits the importing and planting of maize seeds, and as it adopts a strong politics of labeling to ensure the freedom of choice of farmers and consumers.

⁶ Interestingly, the question of the definition of what is life is very rarely tackled by ethical committees. Moreover, they mainly tend to focus on the two frontiers between beginning of life and death.

DRAFT - DO NOT QUOTE

selected experts, were aimed at circumscribing the relevant fields of ethics that were previously neglected and at providing some possible answers and objectives likely to support the decision-making process.

However, by the 1990s, technology, science and innovation management added a new concern for wider public inclusion and engagement leading to the rejection of expert's hegemonic assessments based on the concrete decision of ethical boards or committees, or on the reflection of social scientists (Sykes and Macnaghten, 2013). Although the moral depths of their endeavours were hardly ever disputed, ethical boards and committees were said to reflect a view from above, impervious to the way ordinary citizens assessed technology or research. Ethical boards and committees became insufficient resources to adequately deal with the social acceptability and ethical desirability of technology. Because ethical committees were unable to represent adequately the values, beliefs and intuitions of the members of society as a whole, they could no longer be the sole basis of fundamental normative decisions.⁷ In short, the legitimacy of the actors involved in technology and scientific issues challenged the moral power of ethical experts.

Subsequently, Technology assessment approaches⁸, evolved to integrate participatory approaches in an attempt to develop technology in a more democratic way thereby allowing actors of society to discuss their values, or other moral elements (such as moral intuitions, principles, norms), and even influence the shaping of technology. Participation was said to be a way to broaden the set of normative elements required to make technology-related decisions, in stark contrast with the limited ability of expert committee to stand for the values of other society members.

Eventually, participation was promoted as a way to deal with the conflicting values (or moral elements) or normative systems of social actors, even in cases where the response of individuals to technological change was mere indifference. The need to involve citizens, end-

⁷ The normative decisions that are made in the medical field, where technical tools are used to address specific health problems, appear to be more convincing than the decisions made in the field of technology.

⁸ A distinction must be made between the general tools available for assessing the impact of technology, which we refer to as "assessment of technology (TA)", and the specific Participative and organised approach of Technology Assessment (PTA). See (Reber, 2006a).

DRAFT - DO NOT QUOTE

users, and, more generally, various stakeholders having an interest in technology gained steam and served to lay the foundation for a new ‘social contract’ between science, society and technology (Owen, *et al.*, 2013). In this way, the upstream stakeholders’ involvement in a new technology, before marketization, helped to prevent commercial backlash, public reluctance and major environmental, ethical or social damages. Moreover, bringing explicitly the values (or other normative elements) of individuals into light to shape research and economic policies as well as scientific and engineering activities led to raise public awareness on science and technology, and therefore – as it was claimed – contributed to the legitimacy of findings and policies while increasing the social acceptability and ethical desirability for innovation.

Such is the overall picture that emerges from the contemporary articulation between science, technology and society which has recently brought to the front the principle of “Responsible Research and Innovation” (Owen et al, 2012, 2013; Hellstrom, 2003; Guston, 2006; Barben et al, 2008; Owen et al 2009a; Owen and Goldberg, 2010, von Schomberg, 2011 a, b; Lee, 2012; Armstrong et al, 2012). RRI furthers the tradition we mentioned above, which originated in technology assessment practices, but also continued the ethical reflection on technology and research, as framed by Ethical Legal, and Social Impacts or Assessment (ELSI or ELSA) initiatives which emerged in the early 2000s during the controversy over the development of genomics. In addition, the reflections on corporate social responsibility (CSR) and Sustainable Development (SD) in the last three decades has also paved the way for an embodiment of ethics into the shaping of technology. The latest from a dynamic reflection on the ways to frame adequately technological and scientific evolutions, RRI found in nanotechnology a fertile ground to rise and prosper (see, for instance the US National Nanotechnology Initiative or the National Research Council) although other “hot topics” addressed by RRI, such as geoengineering and synthetic biology, have also been added to the list.

What are the consequences of this historical shift from technology assessment and ELSI approaches to the RRI concept? First, the attention has been drawn from technology to innovation and research, on the one hand, and from ethical reflection to responsibility, on the other hand. Schematically presented, from technology to innovation and research we move

DRAFT - DO NOT QUOTE

towards a concept that widens the nature of the actors involved and gives corporates and organizations (firms or research institutes) a specific role. From ethical reflection to responsibility we move from assessment to action, where the ethical dimensions of innovation and technology are not only thought out, but possibly managed and dealt with. Beyond that, many of the fundamental issues of ethics (such as the definition of the good, or the issue of universality of judgments) are left out by the concept of RRI, since the realm of ethics is no longer explicitly invoked.

Second, the concept of RRI is sometimes considered as an unreachable target because the economic and technological forces ruling the dynamic of innovation would prevent it from the possibility of being a place of a rich discussion about responsibility. In this context, one of the aims of this deliverable is to investigate whether it is possible to build a conception of RRI that would solve this paradox and articulate responsibility with innovation and research processes to think both notions in a complementary way.

Thirdly, there has been no agreement until now on the way in which responsibility is conceived. Indeed, responsibility has been organized according to various meanings that either insist on its negative and external dimension, or on its positive side, highlighting the power of sanctions, or the strength of particular capacities to act in an ethically desirable way. Moreover, different forms of governance exist and have been promoted in the context of research and innovation (delegated, educational, using participation as validation, implemented with the help of different institutional tools (i.e., Participatory Technological Assessment, ethical committees, forums, observatories). Participative tools or systems organize in various ways *accountability and responsiveness*, opening up different conceptions of what *responsible* agents are (focusing on their role, their capacity, their moral obligation, etc.). What does it mean to appeal to responsibility while considering innovation and research? And how can responsibility in innovation be defined (responsibility of whom, to what, etc.)? To answer these questions, this deliverable will analyse the different perspectives of responsibility that have been put forward and will set some foundations towards a more adequate conception of RRI.

DRAFT - DO NOT QUOTE

But this will not exhaust the issues raised by the concept of RRI. Indeed, if participation is a key step in the process of responsible research and innovation, what is the goal targeted by such participation? Who should the participants be (balance between experts, stake-holders and lay-people)? What should the structures of participative devices be? What can participation offer in terms of quality and added-value? Considering the shift from participation to deliberation, promoted as the prominent theory of democracy or in RRI theories, how is it possible to articulate both processes, participation and deliberation, while avoiding their contradictions? To what extent can the involvement of different stakeholders ensure the “ethicity” of the process? Moreover, if participation is supposed to help actors so that they can express different values and moral perspectives, how to deal with the current context of moral pluralism stemming from the fragmentation of social authorities in modern societies and the heterogeneity of normative sources (moral theories) or “comprehensive doctrine” (that includes, among others, the actor’s conceptions of the good, their visions of the world, religious beliefs, etc.)? And finally, is it possible to agree – but we will have to ask whether agreement can still be a desirable achievement – on common RRI norms from participation processes? These issues are relatively complex as it is illustrated, for instance, by the limitations of Participatory Technology Assessment approaches, which do not challenge the relationship between the values brought up by the participants to this procedure and the values of other society members. In taking for granted the link between these two types of values, PTA fail to see that selected participants, who do not comply with all the characteristics of a representative panel of society, are submitted to a specific dynamic of reflection and training. They will not come up with the same judgments than members of society in their everyday life where the questions of technology and science might be less important. Therefore, we need to investigate the “efficiency” of the participation process in meeting citizens’ values (and value systems) and in shaping technological development in a more responsible way.

Addressing these issues will require us to give a closer look at the way in which actors are involved in RRI processes, or at the ways in which participation is conceived and assessed and the ways governance can promote, foster or assess scientific, economic or industrial activities in terms of RRI. In other words, in this deliverable, we will not limit ourselves to a

DRAFT - DO NOT QUOTE

reflection about the definitions of responsible innovation and practices. Rather we have to ask: what can make RRI effective? This will imply an investigation from the viewpoint of the concrete institutional frameworks of RRI that will shape scientists', industrials', innovators' or policy makers' activities.

For grounding innovation and research into responsibility can be interpreted as an attempt to deal with the attendant uncertainty of these practices. It seeks to ensure the dynamic by which innovation would be shaped all along its process of development, according to society members' values that would be embedded in the conception and development of technology. Social acceptability and ethical desirability become key elements in the process of responsible innovation and research, with interplays between both. For both targets are often brought together although they differ fundamentally. Social acceptability is related with the political legitimacy of decisions while ethical desirability is intrinsically normative and relates with the ethical correctness of decisions. And one of the key issues RRI theories have to face – this will be one of the parameters that will help us to assess current approaches of RRI – is precisely to coordinate both.

The last problem that is raised by RRI approaches and by their strong focus on participation is related with the application and implementation of norms. How to ensure that norms constructed out of actors' values will be implemented, and followed? Here, it will be necessary to devote a specific attention to the link between norms and their context and to consider the criteria by which acceptability and desirability are assessed. If the social and moral assessment of innovation gains its legitimacy from they being the result of a collective process, our research has to integrate the practical conditions that are necessary to ensure the ethical efficiency of this process. This will be done in DEL 2.3 while the present deliverable will investigate how RRI theories and their governance approaches that have been put forward by the literature conceive the relation between norms and their context, and how they deal with the issue of the implementation of norms. To put in another way, we will examine the relationship between the construction of norms and their application that are established by the different governance approaches of RRI. However, following a critical methodology, we will not only set some examples of RRI or solve one specific case (related to specific topics

DRAFT - DO NOT QUOTE

such as ICT, bioethics, privacy, etc.). Rather, we seek to provide a theoretical framework common to all of these cases as well as the structure from which every attempt of responsible innovation in research could be understood.

DRAFT - DO NOT QUOTE

Chapter 1: The forerunners of RRI.

The GREAT project brings together research and innovation. The relationship between both notions often insist on the possible contributions to innovation research has to show. Indeed, in many discourse of decision-makers, research is supposed to create new possibilities for industries and job opportunities. However, a closer examination shows that research and innovation differ on two levels (alongside the differences between innovation and invention that are analysed next in this chapter). First, they work on different timescales. Research time is often slower than innovation time. Second, research may focus on the assessment of the effects of innovation, and, more precisely, on its adverse and dangerous effects. This view may be illustrated by the biological controversies surrounding the debate on GMOs. Responsibility will not be the same if research is simply behind innovation processes or if research is broadly opening the scope of medium and long term consequences of innovation. Moreover, there might be different sensitivities to risks of damages (risk assessment) and different perceptions of the precautionary principle.

Focusing both on research and innovation, this chapter aims to introduce some of the issues raised by the concept of RRI. The first section briefly presents the notion of innovation and the features that will be useful in building an appropriate conception of RRI. Section 2 then considers three perspectives that prefigure RRI: Corporate Social Responsibility (CSR), Sustainable Development (SD) and Technological Assessment (TA). Pointing out the limits of these perspectives will help us to sketch some of the main lines of RRI theories that will be developed in the further chapters.

1.1. Innovation

A common feature of many approaches defining the concept of innovation is to focus on what is new and of value for end-users. For example, O'Sullivan and Dooley (2009, p. 5) write that:

DRAFT - DO NOT QUOTE

“Innovation is the process of making changes, large and small, radical and incremental, to products, processes, and services that results in the introduction of something new for the organization that adds value to customers and contributes to the knowledge store of the organization.”

We will come back later to the distinction between product, process and service innovation (to which a forth category, “innovation of paradigm” can be added). For now, let us only emphasise that innovation is related to bringing “something” **new** for the organisation that is of **value for “customers”**. In this perspective, innovation is intrinsically linked to organisations (this include firms but also research institutes selling services linked with their research, for instance) and also related with the possibility to satisfy a “customer” and to make benefits. This anchoring into market activities is the heart of the traditional distinction made between **invention** and **innovation** where invention is the creation of something novel while innovation includes “the exploitation for benefit by adding value to customers” (*ibid.*). Innovation implies the possibility of **bringing something new into the market** in a way invention does not. Indeed many inventions have not given any outcomes in terms of benefits or value for end-users or customers. Moreover, such a perspective of innovation gives a prominent role to the **organization** (as opposed to the scientist) at the crossroad of technological changes, economic constraints, and social “needs”, even if those needs do not always pre-exist to innovation. Analysing innovation, because of its very definition, will imply considering different actors and layers of decisions composed of technological, scientific, economic and social constraints. Of course, more generic conceptions of innovation do not explicitly refer to the economic exploitation of novelty. For instance, Smits *et al.* (2010, p. 1) broadly define innovation as “the development and adoption of new and improved ways of addressing social economic needs and wants.” However inexplicit, the close link between innovation and market activities remains since the latter are the most common way of satisfying “social and economic needs”.

In addition to these different dimensions of innovation, which we have to take into account in a framework of RRI, another important distinction, which is mentioned in the definition above concerns the difference between **radical** and **incremental** innovation. Radical innovation

DRAFT - DO NOT QUOTE

implies that substantial changes are made within the organisation. It is competence destroying, from the point of view of the organization, as the technological knowledge that is required to exploit it is very different from existing knowledge, which could become obsolete. It is a risky activity through which the benefits of organisations can be highly increased but also deeply threatened. Opposed to that, incremental innovation designates smaller changes associated with limited risks but also limited increase of benefits. The knowledge required to offer a new product or a new process of production builds on existing knowledge and is therefore, competence enhancing for the firm. Both cases will yield different consequences if responsibility is introduced for radical innovation entails a disruptive potential that can relate not only with the internal functioning of an organization but more widely to society as a whole, as it is shown by the internet revolution for instance (Christensen, 1997). This will have to be taken into consideration while analysing responsible innovation.

But before presenting the various conceptions that have been developed under the banner of RRI, let us mention different approaches in which RRI is rooted. This includes various perspectives of technology assessment, but also the contemporary concerns for corporate social responsibility and for sustainable development, which are considered as dealing with the ethical issues of science and technology. We will show that even if RRI shares several concerns with technology assessment approaches, sustainable development or CSR, it moves a step forward in insisting on the participative, collective and deliberative dimension of norms' assessment.

1.2. Corporate Social Responsibility (CSR) and Sustainable Development (SD).

Corporate social responsibility has been called on since the 1980s as a self-regulating process integrated in the management of firms to face the challenges of economic and technological changes. It rests on the idea that firms should not only take the interest of shareholders but should also include those of its stakeholders, i.e. employees, customers, suppliers, communities where the firm's plants are located, potential pollutes (Tirole, 2001), but also regulators, local communities, non-governmental organizations (NGOs), civil society

DRAFT - DO NOT QUOTE

organizations (CSOs) or ‘the public’ at large (Groves *et al.*, 2011). In getting involved in a CSR process, firms undertake to comply with national or international legal norms (Human rights, laws relating to child labour) but also ethical norms (ISO norms or the Global Reporting Initiative, for instance). In the larger field of “business ethics” it has been a way to drive corporates to add to their traditional profit maximization behaviour a social concern for the consequences of their activities.

In our analysis of RRI governance approaches, it seems unnecessary to provide with a full normative assessment of CSR (see for instance, Carrol, 2009, Crane *et al.*, 2009; Margolis and Walsh, 2001; Porter and van Linden, 1995; Vogel, 2005.). Here, we will only highlight some elements that can help us to analyse the relevance of RRI approaches compare to that of CSR. The first limit of this framework is that social responsibility is understood in a pure consequentialist way: according to the very definition of the concept, corporates, in their strategic management decisions, have to take into account the consequences of their activities on various stakeholders. This way of approaching responsibility is not satisfactory as it rests on a too optimistic vision of knowledge and rationality. Responsibility cannot only rely on anticipatory capacities or knowledge production efficiency. An adequate conception of responsibility also morally engage individuals or organizations by virtue of their actions, and, somehow, whatever the consequences. This would be true even if we don’t need to fully renounce to consequentialism. And this would render CSR insufficient in dealing with all the aspects that are crucial in RRI. Moreover, CSR, in the case of innovation (such as nanotechnology), is often reduced to the compliance with existing norms. Firms tend to conform to agreed standards but fail to adapt to the new situations created by innovation. They do not get involved into a proactive attitude, where they would tend to anticipate future developments and elaborate norms related to them (Groves *et al.*, 2011; Kuzma and Kuzhabekova, 2011, a,b). In this respect, CSR would only be a first step towards responsible innovation and research but by no means covers its whole scope.

The other notion sometimes invoked when thinking the entanglement of economic and scientific practices, on the one side, and normative frameworks on the other side, is Sustainable Development (SD). SD, as it has been referred to for the first time in The World

DRAFT - DO NOT QUOTE

Commission on Environment and Development Report known as Brundtland Report (1987) implies “a development which meets the needs of the present without compromising the ability of future generations to meet their own needs.” It promotes a process of economic, social and human development by which our use of resources has to ensure the sustainability of natural systems and the environment.

Here again, it is not the place to go through a deep analysis of sustainability, as our focus is on RRI. What has to be mentioned though, is the tension existing between SD and innovation. On the one hand, SD aims at ensuring that the development (or progress) of society is implemented in such a way that it does not compromise the needs of future generations, whilst meeting the needs of the current generation. It is based on a perspective of an equilibrium between the development and needs of individuals and the ability of the nature to produce and renew the resources humans need and take from it. In this sense, all the processes developed for carbon reduction, energy efficiency, *etc.*, which are in the making for many decades, implied efforts towards achieving sustainable innovation (even earlier than the rather recent RRI agenda). On the other hand, SD and innovation can also be seen as somehow antinomic. Indeed, innovation – as we will see later – creates disruption and discontinuity and can lead to the extinction of some social actors, where sustainability seeks to preserve and maintain living entities. If the ‘green economy’, and the objectives of climate change *etc.*, held under the banner of SD attempt to modify and shape existing traditions – in the building of houses or running of factories – SD rest on the normative (and in a way conservative) idea that the stock of natural resources has to be preserved and maintained.

In this context, expressing the issue in terms of RRI offers a much broader framework to reflect on the relation between ethics and technological or economic development. The scope of RRI is, in a way, much larger than SD: it aims at shaping the way we innovate and create new things and new ideas, according to norms whose construction process has to be reflected on and assessed when sustainability relies on existing norms of preservation. This is not to say that sustainability is not an ethically desirable goal. It could even be a norm on which some agreement is reached (as it has been the case in the European Treaty, for instance) in order to

DRAFT - DO NOT QUOTE

build a responsible development of our economies. But in this case, sustainability would be *included* in a RRI process, and would by no means, exhaust it fully.

1.3. Technological and Participatory Assessment.

Aiming at democratically assessing and shaping the development of technology and with the purpose of better including ethical issues, several tools or devices have been provided. They include different approaches of technology assessment (such as Technology Assessment (TA), Constructive Technology assessment (CTA), Participatory Technology assessment (PTA), Real Time technology assessment (RTA)), but also value sensitive design, risk assessment, the precautionary principle or even the more recent New and Emerging Science and Technology approaches (NEST). An extensive analysis and a critical review of these different modes of governance has been proposed within the EU funded FP7 ETICA project (cf. Del 4.1). It is not our purpose here to repeat this work. We will only briefly tackle some of its main conclusive elements (Del 4.1, pp. 28-48).

First, even if these approaches offer different ways of assessing and monitoring science and technology, they all face a common issue of *framing*, due to their process of selection of the actors involved (Goven, 2002; Ladikas, 2009). Technology assessment approaches based on impact assessment, forecasting, scenario analysis or consensus conferences, all give an important role to expert opinion. Consensus conference, for instance, or other consensus-based assessment procedures, have particular difficulties in that the idea of the “consensus” is conceived exclusively by the members of the panel at stake. The formulation of the questions to be asked are already directed, which in turn directs the agenda for the conference. This means that the problem has been already constrained and has been already given its own boundaries. However, already selecting the issues on which individuals will have to reflect predetermines the normative results that will be brought up. And, as we will see in the INDECT case study, this would be problematic, as it prevents the process from any possibility of reflexivity, understood as the capacity to reflect on the way in which the problem is conceived. Yet it is true that the citizen involved in more structured process like citizen or

DRAFT - DO NOT QUOTE

consensus conferences can sometimes have space for initiatives⁹. Moreover, their contribution is more visible in the course of the debate than in the final report that is too often the only piece considered by decision-makers and researchers as well.

Secondly, there is a problem of “capacitation of actors” due to the way in which the various actors and stakeholders are selected. Indeed, it is impossible to include *all* stakeholders, especially in technological projects where they can be very different in terms of culture, socio-economic background, age, gender, and location, to only mention a few variables. Here come three different problems that are rarely explicitly distinguished: a) to find the appropriate learning process to face the diversity of the public, b) to be skilled enough as experts to translate sophisticated knowledge in interdisciplinary arenas, c) to compare the different assets behind the choice of neutrality or plurality in the selection of citizen and experts. This last point is crucial. Do we want to have only ordinary neutral citizen or, on the contrary, committed stakeholders? For the experts in bioethical citizen debates, for instance, experts had to defend different normative options (pluralism) in front of citizens (Reber, 2010a). These different choices carried out different social ontologies. Some were very individualistic (neutrality), some were attached to interest and opinion groups. In general, individuals might know more things on an issue, but they might experience smaller margins of cognitive revisions. We will have to return on these questions with the requirement of responsibility as role and capacity.

For now, let us emphasize that the various devices of TA often entail a situation in which actors are subject to the decisions made by organizers, and by the experts running the assessment. This would tend to give to organizers selecting the participants and to experts leading the discussion the major role, to the detriment of actors’ ethical reflexivity as regards their own framing. This would apply even for participatory technology assessment. Indeed, involving lay people or other stakeholders does not put an end to the predominant role of experts even if it tends to diminish it. In adding the voice of lay people to experts’ analysis, PTA certainly opens the possibility of more multidimensional assessment of technology. This

⁹ Addressing unexpected questions like the existence of independent experts (Reber, 2011a) in the first French citizen conference on GM food and agriculture in 1998.

DRAFT - DO NOT QUOTE

is even truer if participatory devices give more initiative to citizens while formulating round tables and questions and while selecting experts. In this perspective, organizers play a crucial role to warrant the respect of the specific tasks that can be attributed to citizens.

However, these nuances in technological assessment protocols do not solve the issue of how conflicting values and ethical conceptions will be dealt with. Also, these processes finally often contribute to reinforce the dominant expertise and the existing restrictive framing of the debate (Goven, 2002). This is highly problematic as organizers, and sometimes experts too, often reduce the normative horizon to something of their own construction. Moreover, they tend to use stakeholders' input to justify their decisions or to ground the social acceptance of technology (*op.cit.*), instead of questioning the social *acceptability* of technology as a whole.

The final issue raised by technology assessment approaches relates with the efficiency of the norms settled during the process and with the conditions for enabling the required changes. It is not because the expression of a particular norm is sound and justified in an ethical way that it will be adopted and implemented. As we will see with the INDECT case (7.1), the identification of a norm does not include the methods or mechanisms required for its expression within a project. Here again, technology assessment approaches – but this problem will also be faced by RRI approaches – never bring into light the concrete issue of how individuals are driven to follow (or not) a specific norm.

After having presented some of the limits of three approaches – CSR, SD and TA – that can be considered as forerunners of RRI, let us turn to the analysis of the “responsibility turn” into the conceptions of innovation.

DRAFT - DO NOT QUOTE

Chapter 2: Responsibility Turn in Innovation.

What does the idea of “responsible innovation” mean and how is it understood? The first common interpretation that comes into mind tends to wonder about the possibility of associating both terms. As an oxymoron, responsibility and innovation would rely on fully conflicting imperatives and constraints. For example, a requirement of transparency that could appear as necessary in defining responsibility seems to be difficult to associate with innovation (von Schomberg, 2013). Indeed, innovators are often driven by competition and by first-mover advantages that call for secrecy. Moreover, the economic incentives of growth and profit maximisation underlying innovation or the quest of praise in research appear, at first glance, to be incompatible with some of the constraints that responsibility towards individuals, environment or institutions could impose. Such conflicts between divergent values (or other ethical elements of moral theories, such as principles or duties) could happen, for instance, if a long term project investment was required (by a government, by civil society, or by an ethical committee) to be stopped on ground of ethical issues (c.f. the debate on GMO’s or stem cells in Europe).

Therefore, the main challenge a conception of RRI has to face relates with the possibility of seeing responsibility and innovation in rather complementary ways. Before turning to this point, we first present some of the elements that make responsibility difficult to implement while innovating.

Firstly, innovation flourishes out of new technological or scientific possibilities and must have the power to satisfy a need in order to generate some benefits. In this respect, it is related with the bringing of something new into the market and supposes that a profit can be generated from the exploitation of new ideas, new conceptions, new tools, etc. Innovation is thus supposed to feed growth in reducing costs of production, in rising the quality of services or in creating a new demand for product and services. In such an understanding of innovation, responsibility plays a limited role: it can be invoked in cases of **noncompliance** with contracts and more generally in relation with **legal constraints**. However, as we will see (chapter 3), this understanding of responsibility as **liability** is limited and only covers a

DRAFT - DO NOT QUOTE

restricted dimension of the problem. In this context, mainly defined in an economic way and ontologically associated with the idea of financial benefits, innovation, in practice and in theory, is rather alien to the idea of responsibility (if the concept is to be conceived beyond the realm of liability).

Secondly, innovation is often depicted as a factor of uncertainty that raises many issues of knowledge. One important challenge of a conception of RRI will be to deal with uncertainty, i.e. to conceive responsibility without the possibility of knowing all the future outcomes of an innovation. Indeed, with the recent developments of science and technology, innovation is increasingly conceived as a factor of newness that goes along with an attendant uncertainty (Hans Jonas (1979), Ulrich Beck (1992), Michel Callon *et al.* 2001). Following a common distinction, risks (where the set of possible events are known as well as their probability) have to be distinguished from uncertainty (where the probabilities associated to a phenomenon cannot be known) or ambiguity (where there is disagreement about probabilities assigned to the occurrence of an event). The two last meanings are part of the precautionary principle, while risk is connected with prevention. Innovation increase the uncertainties of everyday research highlighted by the precautionary principles, (in EU COM 2000, for instance), when it implies phenomena that are only poorly known.

These different levels of knowledge of an event, (his outcomes, the probabilities related to it and more generally, the state of art associated with its scientific and technological elements) will yield different tools of technology management or risk assessment. For instance, in the case of radical uncertainty, the mere possibility of knowing all the possible outcomes vanishes. With the complexity of our world and the many interrelations between humans and technology, knowing all the consequences of their decisions, including their unexpected or side effects, is impossible. This is known as the paradox of knowledge whose roots can be found in the finitude of human beings outlined by Hannah Arendt – human being whose knowledge is limited, whose life is bounded in time and space, whose capacity of action is circumscribed and whose ability to decide is hindered by conflicting values (pleasure, justice, efficiency, truth, etc.). In a consequentialist perspective, it results that:

DRAFT - DO NOT QUOTE

“We have come to rely on scientific knowledge to create the innovations that help us to transform the world, but we cannot expect it to also enable us to calculate the ethically relevant consequences of using it.” (Grinbaum and Groves, 2013, p. 125)

These limitations of knowledge have already been highlighted in the sociologist, economist and philosopher Otto Neurath’s analogy of the reconstructed boat. Language and scientific verification are like a boat that sailors on the open sea must reconstruct, but are never able to start afresh from the bottom. By using old beams and driftwood the ship can be shaped entirely anew, but only by gradual reconstruction. Our lack of knowledge is even widened by the process of scientific discovery itself, worsening the paradoxical situation to which we are led by our increasing craving for knowledge. The recent great discoveries in the area of molecular biology and neurosciences, for example, and the possibilities opened by new devices such as magnetic resonance imaging have greatly improved our actual knowledge of the human body and brain. However, this step in the understanding of how human beings function has also highlighted how many unknown areas remain and how far we are to explain and understand such a complex structure as the human brain. Then, we are condemned at “gambling that what we know and control is enough for taking effective decisions and what we do not know and do not control is irrelevant.” (Pellizzoni, 2004, p. 545).

We will return later to the issue of the ethical assessment of the relevant consequences of our actions. What is worth outlining now, is that innovation, again in its essential dimension, challenges the limits of human rationality and paradoxically reduces the extent of our knowledge. If on one side, innovation is grounded in scientific discovery enhancing our knowledge of the world, the attendant uncertainty it contributes to create increase the areas of darkness that our understanding has to investigate.

Moreover, processes of creation have to be thought while rationality is bounded (to borrow from Herbert Simon’s criticism of rational choice theory; 1955, 1980), where individuals relies on routines and where full foresight of scenarios and pathways of development are impossible. This process of undermining rationality is sometimes claimed to reduce the scope of individuals’ responsibility in wriggling human decision and actions out of the field of their

DRAFT - DO NOT QUOTE

rationality, hence, of their control and of what they can be held as accountable or responsible¹⁰.

Again, responsibility will be rather alien to the notion of innovation in research because of our limited capacity of foresight. To qualify this alleged incompatibility between both concepts, Jack Stilgoe (Paris Workshop, 2013), evokes the idea of “pathologies of innovation”, echoing Hans Jonas lack of futurology in new technologies assessment. Furthering the paradox of knowledge we just mentioned, Stilgoe recalls what David Collingridge (1980) evidenced as the dilemma of control, according to which managing and “controlling” (to use Collingridge’s terms) technology should take place at early stages of technological development but, at the same time, is hindered by the low knowledge that prevails at this stage. In our decision process, we would be caught up between the need to take decisions too early when knowledge is poorly available and too late when it is impossible to alter the course of things in a significant way. As we have mentioned above, the temporality of processes will gain a crucial importance. More precisely we have to face conflicts of temporalities.

The tension between the belief in economic forces feeding human development and the fear for irreversible damages would lead to what Ulrich Beck named ‘organized irresponsibility’, i.e. the routinely short-circuit of the attempt to circumscribe the negative effects of technological progress through reflexivity. On the one side, there would be an increasing public awareness of risks and uncertainties of technology that is supported by socio-institutional frameworks. On the other side, faith in progress, dependence on rationality, the hegemony of expert opinion and economic constraints would lead to a deleterious denial of collective and individual responsibility. According to Stilgoe, these characteristics of our modern management of science and technology would make a plea for responsible innovation, even if the precautionary principle arising in the European scene has contributed to reduce factors of irresponsibility.

¹⁰ Chapter 3 will show that this negative conclusion strongly depends on the conception of responsibility that is been undertaken, and in this sense remains disputable.

DRAFT - DO NOT QUOTE

A perspective, which highlights the pathologies of innovation is certainly right in explaining the background of RRI and in outlining why RRI is needed. But from our analytical viewpoint, this analysis and most of the current conceptions of innovation ontologically disconnect the processes of innovation from their ethical dimension. This would explain why the GREAT project aims at providing with an approach of RRI governance that allow for this disconnection to be overcome (cf. del D.2.3). In this task, we are not left empty handed, as some recent evolutions the ways of innovating could set the basis for an integration of responsibility into the dynamic of innovation.

First, other areas of interest have appeared out of the traditional distinction between product/process innovations. Let us recall that this distinction discriminates between product innovation, i.e. what a firm offers to the world such as a new devices of communication (smartphones), on the one hand, and, on the other hand, process innovation, i.e., the way in which product and services are created and delivered e.g. new modes of communication (free calling between computers via *Skype* or new modes of trade through *Ebay* or *Amazon*). Innovation in products aims at satisfying an old need in a new way or at filling a new need while process innovation relates with the way in which production is organised and managed. Both types of innovation can feed each other since changing the process of production can lead to product innovation, as, for example, in the automobile market, Toyota's focus on improving quality of process led them to improve the quality of their services and products, contributing to their leading position on this market.

Beyond this first distinction, innovation can take other forms (Bessant, 2013, p. 9; Lhuillier, 2007). First, there is a possibility of innovating in positions. A firm can make an innovation in the position it holds or in the segment of the market it targets (such as *Tata's Nano* car or the tactile tablet conceived by *DataWind* offering to poor Indians the possibility of buying these goods at low price). Secondly, a more radical form of innovation is involved when the "mental model", in Bessant's words, framing what the organisation does change (such as when new platforms like *Amazon* or *Skype* redefine how advertising is conceived, or when

DRAFT - DO NOT QUOTE

new practices such as crowd funding¹¹ or crowdsourcing¹² appear). This last dimension would refer to “paradigm” innovation¹³.

In such a perspective, the emphasis is made on the possibility that the “mental models” of the organisation are changed to propose not only new products, but a whole new conception of the function that products have to serve. This conception opens the possibility of reframing the relation between individuals and products (how does a product satisfies the function of communication, for instance). Moreover, it shows that actors of innovation have a crucial capacity to invent new processes by which a specific function is satisfied. And this will have specific consequences for responsible innovation (*cf.* D.2.3)¹⁴.

In the same vein, innovation has been recently conceived as a collective process – a “multiplayer game” (Bessant, 2013) – involving a whole network of actors, a complex system of interactions between them and an institutional, social and political environment. Innovation is the results of the work of scientists, engineers, companies, policy-makers and end-users that interact and shape a specific set of legal rules, norms, and public policies. This collective dimension of innovation does not necessary imply collegiality. One of the aim of the governance approaches of responsible innovation precisely seeks to organise the possibility of a collegiate design of innovation. But some recent evolutions in practices already appear to be a move toward a more collective way of innovating. Such a tendency is illustrated by what is now called “open innovation” favoured by open source platforms that individuals can freely contribute to improve, modify and change according to their values and preferences (e.g. *Linux* software, open dictionary Wikipedia or new Research and Development practices (R&D) such as crowd sourcing). Of course, the degree of openness may vary greatly from one case to another (e.g. administrators choosing particular contributions and steering the

¹¹ Crowd funding designates the practice of funding a project or venture by raising many small amounts of money from a large number of people, typically via the Internet. Each individual becomes then a micro-investor in the project he/she is supporting.

¹² Crowd sourcing designates new ways for individuals or companies to obtain needed services, ideas, or content by soliciting contributions from a large group of people, usually an online community, rather than from traditional employees or suppliers.

¹³ We will return to this issue in del 2.3.

¹⁴ We do not mean, here, that innovation in product, in process or in position cannot be responsible. We only point out the specific role that paradigm innovation will play for RRI, alongside the other types of innovation.

DRAFT - DO NOT QUOTE

participation process or standard templates allowing only for particular data entries). However, the idea of implying other stakeholders in the design of innovation is a step toward a co-construction of technology. Finally open innovation raises new organisational issues related with the need to coordinate an important number of contributors more or less active, and to organise how knowledge is shared (and rewarded) among them.

These recent evolutions of innovation in paradigms have had the interesting consequences of promoting an early involvement of end-users in the design and conception of some products. Participatory design practices, where end-users co-conceive the product with the engineers or designers of the company have emerged¹⁵. End-users become co-creators of the product as in the case of *Adidas*, for instance, which offered the possibility to design one's own model of shoe in combining several elements (colour, prints, shape, etc.) that could then be produced and delivered. Another example emphasised by Bessant is the *Smart* design contest launched by *Daimler-Benz*, which gathered participants all over the world, allowing them to propose their design and to vote for other's propositions. Also worth is to mention the emergence of living laboratories (Living labs), developed from the 1990s at the MIT by William Mitchell (with further help of Kent Larson and Alex Pentland) where end-users test a product in their living conditions. For instance, in the case of a new phone or a new application, the test phase will add to traditional panel methodology a survey into the way in which end-users concretely use the technology in their everyday life. As an illustration, this methodology has been implemented by energy companies (such as *EDF* in France or *Oxxio* in the Netherlands), to test "smart" meters, i.e. devices installed in a number of volunteer (or remunerated) households to control their use of electricity. The aim of the experience was to determine, *in situ*, if individuals were keen to use these devices and also their behaviour towards them. Beyond the pure fad, these new ways of creating products and services ensure the company that their output will meet its target. Also, it is a step toward a better incorporation of end-users' tastes, practices, habits and values into the making of a product. If these values might conflict with budget constraints, technical standards or existing infrastructure and might be therefore hard to integrate, the early involvement of some of the stakeholders concerned by an

¹⁵ Co-design, living labs and "smart" cities are now hot topics of innovation theory, as it is illustrated by the program of the recent ICE IEEE International Technology Management Conference on "responsible innovation and entrepreneurship" that happened June (24-26) 2013 in The Hague.

DRAFT - DO NOT QUOTE

innovation, contribute to strengthen the linkage that has to be made between innovation and responsibility. These different attempts to step out traditional ways of innovation (such as living labs and other processes of technology co-shaping) offer new possibilities towards responsible innovation in opening the making of products to actors outside the firm. However, responsibility cannot be reduced to individual short-term desires. Rather, individual and collective desires should be challenged and shaped by responsibility. Therefore, an accurate process of co-construction will require to include other stakeholders (scientists, the public at large, NGO's etc.), and should not be limited to the involvement of end-users.

Finally, as highlighted by Xavie Pavie (2012), innovation introduces gaps and ruptures in the lifecycle of products. Sometimes, it is possible to forecast the evolution of the product and to integrate these evolutions as it has been promoted by sustainable development practices. However, some ruptures are unexpected when they result from unforeseen market successes or failure, for instance. In that case, they can have deep consequences on responsibility. As an illustration, Pavie mentions what Mark Zuckerberg had to face after the creation of his social network *Facebook*, which brought him to hold an unanticipated responsibility – at least, at the very early steps of the process –related to privacy, individual data storing and misuses of data. Here, the economic success of innovation (i.e. the large number of users) led to the question of responsibility (bringing forward the issue of members' privacy and right to control their data), reversing the conclusion we mentioned above. In this view, innovation would, at least in some case, intrinsically lead to responsibility.

This would be even truer if we recall Joseph Schumpeter's understanding of growth as rooted in disequilibrium. In his view, growth is characterised by a process of 'creative destruction' which refers to the disappearance of the least competitive actors in favour of those who manage to adapt, to create new ways of production, new products, or new practices. This disruptive power of innovation again raises the issue of responsibility as it can be at the root of individual and collective tragedies and inequalities (see for instance, the difficulties of the coal era's end in post-industrial societies).

DRAFT - DO NOT QUOTE

Now, while reflecting on a possibility of responsible innovation, these different examples are of particular interest as they open a way of linking innovation to responsibility almost in a logical way. In other words, some of the elements that intrinsically constitute innovation would already imply a form of responsibility. First, innovation is a dynamic process. And intrinsically, it is about change. Successful innovators show a high capacity to adapt to their changing environment and to social, institutional, technical or ethical constraints. This dynamic aspect of innovation involves **responsiveness**. Indeed, successful innovation implies a capacity to react to changes, to ameliorate, modify or even abandon a project that seems to be not profitable. If being responsible cannot be limited to the capacity of responding, the latter certainly plays an important role in the former. In itself, innovation contains the possibility of integrating adaptive processes of conception and production, which is one of the pillars of RRI. However, it is not because the possibility of thinking innovation in an ethical way exists that it will be necessarily exploited. Therefore we will need a specific enquiry into the normative conditions needed to ensure a responsible way of innovating (next chapters).

To sum up, innovation is conceived as a complex process that results on the one side, from forces that favour a limited conception of responsibility (when it is **reduced to liability** or when innovation is only driven by the quest for economic benefits without taking other parameters into account). On the other side, innovation practices evolve rapidly, and recently they relied on the need for a co-shaping of technology and products including the persons to whom they are designed¹⁶. In addition, current understanding of innovation emphasised the role of **responsiveness** in successful organisations. These elements, as we will see afterwards, can be taken as building blocks toward an understanding of responsible innovation governance. Beyond that, they show that a possibility of intertwining responsibility and innovation exists, as opposed to traditional conceptions. However, to ensure the transition from a potential link between innovation and responsibility to an intrinsically normative concept of innovation in research we need now to conduct an inquiry into the meanings (and their potential relationships) of responsibility when related with innovation. This is the aim of the two next chapters.

¹⁶ Of course, as we already mentioned it, this is only a first step towards a responsible way of involving stakeholders as the latter do not only include the end-users of a product but many other actors that might not be directly affected.

DRAFT - DO NOT QUOTE

Chapter 3: What responsibility?

To address the issue of responsibility in research and innovation, we need now to better understand this complex and central notion in philosophy, and more precisely for the GREAT project, in moral philosophy. Indeed, it will be important to clarify the conflict of interpretations surrounding the notion of responsibility, the implicit blind spots in various RRI approaches and the possible ways to articulate different understandings of responsibility (see in Annex 2, General problematic of the Paris workshop on RRI, 2013). Responsibility is a rather loose word that refers to different meanings (Hart, 1968; van de Poel, 2011; or Vincent, 2011). For instance, the **causal** (logical) dimension of responsibility (as when we say that a hurricane is responsible for the death of 1000 people) has to be distinguished from dimensions of **blameworthiness, liability or accountability**. Causal determinism and possible freedom is largely discussed in moral philosophy (Fischer, 1999). For some authors causality contradicts moral responsibility, when others defend compatibilist positions.

These various interpretations and approaches of responsibility - more precisely moral responsibility - that have been developed for decades by moral philosophers are of different relevance to understand the problem of RRI. Indeed, next to definitions that insist on **sanctions**, other understandings claim for a focus on **positive capacities** such as **care** or **responsiveness**. However, as we will see, these approaches either rely on an external, negative and retrospective conception of responsibility that empty the notion from its content (3.1), or promote a positive, internal and prospective conception that rests on a substantive definition of the good (chapter 4). In both cases, the way in which the norms lying being responsibility are constructed and the way in which individuals apply them is not questioned. This two next paragraphs aim at showing these limits. Our goal is to examine how the issue of RRI has been seized and fathomed by the literature on responsibility. In the chapter 6, we will show how they have neglected the relationship between norms and their context.

DRAFT - DO NOT QUOTE

3.1. Negative understanding of responsibility.

Etymologically, responsibility comes from the Latin root of *respondere*, i.e. to respond. However, in French or in German, for instance, “répondre” and “antworten” not only refer to giving an answer but also **to be held responsible for one’s own actions**, to own up what you have done, in the expression “*répondre de quelque chose*” or “*verantworten*”. The second important root of responsibility, as outlined by Paul Ricoeur (1995), comes from the idea of **imputation**, i.e. attributing or ascribing something to somebody. While the “**responsive**” dimension rather focus on the **intention** of the doer, imputation rests on the causality that bounds an individual and a course of action (Pellizzoni, 2004). Indeed, in one case, individuals have to **account** for their actions and explain the reasons of them, while in the second case we attribute an action to someone who is acknowledged to be the author.

Now, if we consider the meaning mostly given to responsibility, it appears as closely related to the first historical root we mentioned above, linking responsibility to someone’s wrongdoing and to sanction (Ricoeur, 1995; Grinbaum and Groves, 2013). Responsibility is conceived as imputed to someone for his or her actions, whose negative outcomes or harms have to be compensated or **repaired**. In this **legal-oriented** interpretation of responsibility, it is possible to distinguish between **blameworthiness** (when A can be blamed for an outcome X, for instance, a car accident) and **liability** (A is liable to pay for the damages caused by outcome X). In both cases, someone is *held* responsible for her actions or decisions who happened to break the law or to infringe a social or a moral norm.

This conception of responsibility encounters several limits that come from its general **neglect of the internal capacities** of individuals to mobilise their will to act in a responsible way. Here, responsibility is external to individuals. It is backward looking and relies on a norm coming from outside that has the potential of influencing someone’s action through the threat of sanctions. This leads to a misconception of responsibility in at least three ways. Responsibility is (a) bypassed, (b) diluted, and (c) amalgamated with accountability.

DRAFT - DO NOT QUOTE

In order to make our point, let us give a more precise account of responsibility understood in its negative sense (as entailing penalty). Several conditions have been put forward in the literature (Vincent, 2011; van de Poel, 2011) – conditions that are five in number in van de Poel’s presentation that we will follow here.

First, there must be a **causal** link between the doer and her action: person A must be causally involved in the course of action X¹⁷. Secondly, in the specific cases of blameworthiness and liability, the action X or its outcome has to be considered as something wrong (condition of **wrongdoing**). Individuals are held responsible because they engaged in practices that are socially or ethically considered as harmful or wrong. The third condition that is often imposed to assess liability or guilt requires individual A to have the **moral agency** or the moral capacity to act responsively. Inspired by the criminal model, it means that an individual acting in an undesired way but who would lack the capacity of acting morally (because of a cognitive impairment, for instance) could not be held responsible of his or her acts. Fourth, to invoke someone’s responsibility, the agent involved must know about the consequences of her action and about “scenarios, mere speculative concerns, expectations, etc.” (Grunwald, 2011, p.11). Not having access to available **knowledge** or the stock of knowledge being very low because of uncertainties and a lack of information would prevent individuals to which undesirable outcomes are imputed from liability or blameworthiness. Indeed, our actions are constrained and shaped by the available information we have on their consequences. But not only that. Here, information has to be understood in a very broad way that includes, in addition of the knowledge on outcomes, the set of legal rules and moral norms that circumscribe individuals’ field of action. Finally, to assess responsibility – and this would hold beyond the interpretations in terms of liability and blameworthiness – individuals must enjoy **freedom** in the sense that they are not forced to bring about action or outcome X. Responsibility can only emerge from the possibility to choose to act (or not to act) in a free way.

¹⁷ Ibo van de Poel does not enter in the complex debate of causal determinism, which cancel the possibility of moral responsibility (i.e. Fischer, 1999). Indeed, to hold someone morally responsible, he or she must have the intention to act and the capacity to realise his or her action. Human actions and *a fortiori* moral actions are not only produced by causal determinism. The complexity of actions’ network (in a company, for instance), the probabilities of events and above all, phenomenon fallen under the precautionary principle make problematic the simple understanding of causality. These three dimensions - social interactions, probability and uncertainty - ask for a more nuanced conception of causality.

DRAFT - DO NOT QUOTE

3.2. Bypass of responsibility.

Let us first consider the condition of knowledge. In chapter 1, we outlined the different degrees in which innovation can affect our stock of knowledge, creating in the most extreme cases, radical uncertainty (where we know neither the probabilities associated with definite events, nor the whole set of possible events). In this case, but even in less radical cases where the knowledge on the occurrence of outcomes is partial, the condition of knowledge leads to a paradox. On the one hand, we are supposed to exercise our moral capacity to act in a responsible way that push towards the quest of the most extended set of information, while, on the other hand, individuals might try to consciously limit their knowledge, in order to avoid to be held responsible for some undesirable outcomes. For instance, corporates can encounter limits in their incentive to finance toxicity studies (in bio- or nanotechnology, among others) if the latter would imply stopping promising applications. If firms cannot avoid to take into account the current state of art and to include all the already available knowledge, they can limit their quest for new knowledge since being aware of possible unwanted outcomes will engage their responsibility in case of damages in a way ignoring these consequences will not.

From a policy viewpoint, the possibility that firms, and more generally innovators, can be forced to pay financial compensations about their activity, can work as an incentive to reduce environmental accidents or sanitary damages in increasing the companies' incentives to prevent future costs (Jacob, 2013, Paris workshop). But it can also play the opposite role, in favouring opportunistic behaviour, where actors tend to limit their access to knowledge to avoid future liability (*ibid.*). In other words, **reducing responsibility to liability creates a dilemma**, which comes from the fact the conditions for establishing one's responsibility rely too strongly on the burden of proof:

“This problem is quite serious. Since one cannot be held liable for events that happened at a time when there was insufficient evidence of harm caused by one's actions or omissions, scientific ignorance acts as practical exoneration. However, deeming

DRAFT - DO NOT QUOTE

companies liable for any future consequence of their present choices would result in an unsustainable financial hazard.” (Pellizzoni, 2004, p. 553)

Focusing on the possibility to impute future damages on the basis of the available knowledge contributes to build a perspective of responsibility that is purely instrumental. There is no normative involvement of actors as the only driver of their behaviour will rely on the **fear of financial or legal penalties**. Moreover, it has to face the issue of the limits beyond which responsibility cannot be invoked anymore.

3.3. Diluted and unlimited responsibility.

The second type of problems resulting from a conception of responsibility limited to its negative and backward looking interpretation derives from its **individualistic** overtone based on a strong linkage between individual and outcomes. The 2013 EU report on the “Options for Strengthening Responsible Research and Innovation” (p. 57), for instance, mentions a first meaning of responsibility that is supposed to be “predicated primarily of persons and only derivatively of their actions”, while for Grunwald (2011, p. 11), responsibility implies that “*someone* (an actor, e.g. a synthetic biologist) assumes responsibility for *something* (such as the results of actions or decisions, e.g. for avoiding bio-safety or bio-security problems) [...]”¹⁸ In this view the agent and her or his actions are the locus where responsibility can be given any significance.

However, this simple relation between an action (or a set of actions) and an individual is highly problematic. While considering innovation or research, it is often difficult to isolate who is cause of what. This problem, sometimes labelled as the “**many hands**” issue (Thompson, 1980; Doorm and van de Poel, 2012; Bovens, 1998) also results from the future being uncertain or ambiguous, the consequences of emerging technologies being often impossible to forecast. When it is difficult to unravel the tangles of the causal chains that led to a particular set of unacceptable outcomes, and when responsibility can be ascribed to too

¹⁸ Grunwald add to his definition two other items “... according to a *body of rules* (laws, norms, principles, values and customs) and relative to the *quality of available knowledge* about the consequences of the actions (deterministic, probabilistic or possibilistic knowledge (Betz 2010)). (*ibid.*)

DRAFT - DO NOT QUOTE

many individuals, there is a risk that *in fine*, no one can be held responsible at all. Ricoeur (1995, p. 66) goes a step further in mentioning ‘adjacent effects’ – something also grasped by Ian Hacking (1986)’s ‘interference effects’ – that includes the unattended consequences of our acts. This raises the question of ‘how far’ can individuals be held responsible. Should their commitment involve secondary effects, and if yes, wouldn’t this move create a sort of “unlimited” responsibility, which would deprive the concept of its meaning? As stated by Ricoeur:

“taking *all* the consequences into account, including the most contrary to the original intention, results in holding the human agent responsible for all, indiscriminately, let’s just say responsible for anything he can take charge.” (*op.cit.*, p. 66).

Again, the purely **consequentialist** approach of responsibility collides with its own frame: in seeing responsibility as the result of a calculus (the assessment of the outcomes), one is confronted with the time, space, and interactions limits that seem reasonable to assess guilt.

And this would even hold for definitions of responsibility that are not only ascribed to individuals and their action. Indeed, as proposed by the EC report (p. 57) the adjective ‘responsible’ can also be “attached to events and processes which are quite separate from any identifiable individual agent”. It is under this interpretation that we can talk about ‘responsible ways of proceeding’, ‘responsible investments’, or ‘responsible procedures’. Also, the mere terms of ‘Responsible Research’ and ‘Responsible Innovation’ relating to the field of technology, applied science and engineering illustrate the shift that has occurred in our use of them. It now designates **the complex network** of actors, institutions, public policies that is entailed in an innovation process. And it can apply to processes, to single acts, as well as this complex aggregate of practices, of individual decisions and of the environment. But this would render the task of attaching responsibility even more difficult since defining who and when is responsible would become even more problematic.

DRAFT - DO NOT QUOTE

3.4. Reduction to accountability.

This will lead us to the third “sin” of **sanction-oriented interpretation of responsibility**. Furthering his analysis of the different interpretations of the notion, Ricoeur (1995) shows that there is a conceptual displacement from imputation to risk by which responsibility ends up as conflated to accountability. Indeed, the idea of solidarity against risk that led to the advent of insurance systems in the 19th century and to 20th century’s welfare state contributed to alter the pure understanding of responsibility as implying **obligation and repair in the case of fault**. The institutionalisation of the management and prevention of social risks (by means of insurance and social-security systems) replaced the reparation of an individual fault. And not only emerged a conception of responsibility without fault (as when the state compensates its citizens from the effects of a natural disaster). This process also added to the relationship between the doer and his action, a relationship between the doer and the one who suffers, between the actor and the victim. More generally, responsibility now also imply a concern for the “condition of vulnerability” itself. This is particularly evident in the environmental and ecological movement, and in Hans Jonas’ conception of the precautionary approach (which shall be distinguished from the precautionary principle) that has the task to fill the gap between our increasing actual knowledge, our limited knowledge of the future and our capacity for “futurology” as Jonas writes it. All these conceptions promote a strong concern for the future victims of our damageable course of action, (i.e. for the survival of future generations in conditions that can be assessed as acceptable).

Interestingly, this displacement entails another change in the interpretation of responsibility. Ricoeur identifies it as an upstream movement, taking place before the action and its negative outcomes and calling for preventive actions aimed at avoiding potential damages. From the repair of damages we move to a conception according to which responsibility also implies that risks should be prevented (the passive form illustrates **the disappearing of the author of the injury**). If this move opens the possibility of thinking responsibility in a more prospective way, it raises the same kind of difficulty that the purely legal interpretation. To what extent can public institution prevent us from risks? Taking future generations into account, what is

DRAFT - DO NOT QUOTE

the time horizon of the analysis? And in this prevention framework, who holds the responsibility of complex and unexpected effects?

As for the dilution and avoidance of responsibility, the conceptual **reduction of the analysis into a paradigm of accountability and risk prevention, implies a consequentialist framework which conflicts with our limited possibility of evaluating outcomes.** Indeed:

“The rationale for strict liability is that such subjects as entrepreneurs, employers or owners are able to sustain the cost of repairing the damage produced by their property or subordinates and independent from their own negligence.”(Pellizzoni, 2004, p. 551)

But in order to bring a company, an entrepreneur, a policy maker or a scientist¹⁹ to legal sanctions or financial compensation due to her or his irresponsible behaviour or choice, we must be able to assess the value of the damage. In the case of identified accident, such as a polluted river, it might be possible, but when it comes to the possible damages done to future generation, or even unknown risks associated with new devices (as in the case of nanotechnology), human rationality is once again challenged (Pellizzoni, 2004). This would call for a **more “flexible” approach of responsibility focused on adaptive processes (“responsiveness”)** as suggested by the Latin word *respondere*.

But beyond the practical issue of the calculus lies the ethical acceptability and the epistemological validity of a conception of responsibility as a purely external process by which norms of conduct are imposed to individuals by means of instrumental tools (sanction). At no time, the positive ability of individuals to act in a responsible way is called on, reducing the **epistemological** relevance of such a conception of responsibility but also its **normative** validity and finally its **practical** power to influence individuals’ action. Indeed, neglecting the

¹⁹ In October 2012, there has been an interesting debate about the lawsuit verdict sentencing six Italian scientists to six years in prison because of multiple manslaughterers. The scientists were accused of having provided "inaccurate, incomplete and contradictory" information about the danger of the tremors felt ahead of 6 April 2009 quake. As opposed to many interpretation of the verdict, their scientific responsibility in the forecasting of a devastating earthquake was not challenged, as, by now, there is no reliable way of predicting earthquakes (in the short term). What motivated the decision of the court related with the way the scientists accessed and communicated risk, undermining it and leading the public to a fatal overconfidence.

DRAFT - DO NOT QUOTE

possibility that individuals engage their responsibility in a prospective way would raise an epistemological issue since it rests upon a misconception of the whole range of attitudes and behaviours actors can adopt towards their action.

Secondly, the normative or ethical validity of such a conception can be questioned as it excludes the ontological link existing between our actions and our responsibility. Yet, engaging freely and with a relevant set of knowledge in a course of action makes me ontologically responsible for this action.

Finally, the practical relevance of a purely negative conception of responsibility is also challenged since the threat of sanction is by no mean the only driving force of human action. The hope and want to act in a responsible way can be strong incentives that have to be analysed and favoured²⁰.

If we consider responsibility in a broader meaning, not only reduced to liability and blameworthiness, the condition of wrongdoing, for instance can be extended to other understandings of responsibility as we can also be held responsible for positive outcomes, as we will see in the next chapter. The same kind of reasoning applies for the condition of moral capacity or freedom. In negative understandings of responsibility, individuals are supposed to exert **a moral capacity and to be free to act**. However, these capacities are only implied from a logical point of view. The negative perspective never draws out of them all the potentialities they contain to conceive individuals' decision and actions in a closer connection with ethics. In this respect, **liability, blameworthiness or risk prevention, by no means, offer a comprehensive approach of the practices, activities and capacities that lies behind the concept of responsibility**.

²⁰ Cf. DEL2.3. for a further analysis of the threefold distinction by which a negative conception of responsibility fails to adequately understand this concept.

DRAFT - DO NOT QUOTE

Chapter 4: Towards a positive Conception of Responsibility.

In this context, will more “positive” conceptions of responsibility fare better? As we will see, the current conceptions of responsibility that focus on **responsiveness, care, or moral capacities**, offer a way to overcome some of the difficulties we pointed out with negative meanings of responsibility.

More positive and prospective understandings of responsibility assume that individual not only pay for the (possibly wrong) things they did but engage in a process through which they take care of others (other human beings, future generations, non-human beings or the environment). In this perspective, what counts is no longer *being held* responsible for a passed wrongdoing but *taking responsibility* for present and future acts and decisions.

In these more forward-looking interpretations of responsibility, several “degrees” can be distinguished that organise differently the meanings of responsibility.

4.1. Task, role and authority.

First, there is a kind of responsibility involved when somebody is given a **specific task** or a **specific role**, as when the bus driver is responsible for driving the bus (Hart, 1968; van de Poel, 2011). Individuals are assigned to these activities and have to ensure that they operate in the best possible way: whoever is responsible positively and actively mobilises his knowledge of the relevant set of rules and norms, as well as his capacity for action and anticipation. In a similar vein, there is also a meaning of responsibility related to **authority** (van de Poel, 2011), as when we say, for instance, that a project manager is in charge of a project. In their professional activities, individuals are responsible for ensuring that definite tasks are performed, and expected outcomes avoided or favoured. Their responsibility, here, covers a widened set of activities, compared to that included in the definition of a task, as it also implies other individuals’ actions and decisions.

DRAFT - DO NOT QUOTE

4.2. Capacity.

There is a fourth level of this broad set of “positive” meanings of responsibility that draws on the capacity of being responsible - capacity we already mentioned. Individuals not only have to be able to act in a desirable way in order to be held responsible for their actions (necessary condition), they also can show a capacity to act in a responsible way. Compared to the mere *condition* that should be satisfied to assess retrospective responsibility, responsibility as a *meaning*, relies on a richer conception of the individual as the latter can show a positive ability to change his decisions and the course of actions and events. Assuming **moral agency** implies that we have the ability **to reflect on the consequences** of our actions and that we can engage in a **foresight** exercise by which we “increase [our] knowledge about the world and how [our] actions might interact with and alter it” (Grinbaum and Groves, 2013, p. 122). This positive capacity also implies the ability to form **intentions**, to act deliberately, and to act in accordance with certain norms and moral or legal rules (Hart, 1968; van de Poel, 2011; Grinbaum and Groves, 2013). It will have to be distinguished from the moral obligation to act responsibly, for in the latter case, individuals’ action are subject to a moral injunction (van de Poel, 2011).

When knowledge is not too low, when parts of the consequences of our actions can be forecasted and when the social, legal and political environment is well-known and does not evolve too quickly, conceiving responsibility as a task or a capacity can help to size responsible innovation practices and governance. Assigning the responsibility for a project or for a task to somebody involves not only that she positively has to engage to cope with the objectives she was assigned for but also that she will be accountable for the results, implying repair or sanctions in the case of failures (for instance, being fired). These conceptions of responsibility blur the forward-looking/backward-looking distinction that negative understanding of responsibility introduced, as they both involve an ability to anticipate and to give account for one’s own actions. In this sense, compared to the negative-legal interpretation, these understandings of responsibility offer a step further as they imply a positive capacity **to commit oneself** to actions and decisions.

DRAFT - DO NOT QUOTE

However, in the specific context of innovation we highlighted, which is characterised by a high level of uncertainty, these ways of grasping responsibility might be limited, because of their anchoring in a pure consequentialist approach. Innovation, with its power to alter in some radical ways our environment, other human beings and the future generations, cannot be circumscribed in a purely consequentialist framework (Grinbaum and Groves, 2013, p. 124). We mentioned it already: our capacity of foresight is limited as well as our knowledge of future consequences. And through innovation, the world of tomorrow might not be the same as the one of today. In this context, how is it still possible to invoke a conception of responsibility based on to the possibility of foresight and on knowledge of the outcomes?

4.3. Care.

To move from a purely consequentialist framework, one route that has been taken is to focus on the **virtue dimension of responsibility**. Kermish (2012, p. p. 93), for instance, drawing on John Ladd (1991), recalls that responsibility can refer “to the absence of care or concern for the welfare of others” (and not only to fault)”. In focusing on the concern (or the lack of care or concern) that individuals have for a particular situation or succession of events, the accent is made on a positive faculty of individual to favour or prevent some course of actions, as opposed to the backward looking position that imply compensation once the damage has already happened. This perspective is therefore both prospective and retrospective and both descriptive and normative (as taking care of somebody or of a situation supposes that we act within a set of definite norms).

As a response to the deficiency of consequentialist frameworks in adequately dealing with responsibility, this relation of care has been represented by means of a metaphor borrowed to the family field²¹: innovators (including entrepreneurs, scientists or a network of actors) should act towards their innovation as parents take care of their children (Grinbaum and Groves 2013, following Jonas, 1984):

“What characterizes the role of parent is, first, caring for a child’s capabilities and ensuring that they are given the opportunities to develop them, which requires a thickly

²¹ Hans Jonas also proposed the metaphor of the paternalistic relation between the king and its people.

DRAFT - DO NOT QUOTE

detailed understanding of what makes for a valuable set of such capacities (or “good” character, if you prefer). The parent is not expected to have the capacity of superior foresight regarding the future consequences, which as we have seen, is an expectation that technological societies both promote and undermine.”(Grinbaum and Groves, 2013, p. 131)

Obviously, the analogy will encounter some limits as the autonomy of technological artefacts, once they evade their creator’s control, cannot be compared to the one of a child becoming adult. But the idea underlying the metaphor is that “during early elaboration stages, technologies are dependents with sensitive and malleable potentialities, in which it is hoped the ingenuity of innovators will produce [...] particular virtues” i.e. “the efficient production of ‘right impacts’” (*ibid.*) As in value-sensitive design approaches of technology (Kelty, 2009; van den Hoven, 2013), innovators are supposed to shape the design of technology and to accompany its development in a “good” way, without being tied to the obligation of forecasting all possible consequences.

The perspective that is depicted here is a good illustration of moral pluralism as it brings together elements of moral theories that it is usual to oppose, namely consequentialism and virtue ethics (Reber, 2006b). Indeed, the relation of care suppose that the actors have a concern for the outcomes of their actions but also that they rely on their capacities of adaptation more than on a purely mathematical ability to forecast all the possible consequences. The framework remains consequentialist but not necessarily monist (as it is the case in a utilitarian perspective, for instance). On the other side, the capacity of acting in a virtuous way is also called on, since parents/actors of innovation have to take care of their “creation”, i.e. to ensure their living in the best possible conditions. This conception, then, also borrows from virtue ethics to qualify the behaviour and decisions of the actors involved.

Such a way of dealing with technology and innovation supposes a redefinition of the relationships towards individuals. As contented by Adams and Groves (2011, p. 22), taking care of somebody implies that we **commit ourselves** to provide her with what she needs. There, the “value of the relationship is the key factor in motivating responsible action, and it

DRAFT - DO NOT QUOTE

is also the object of acting responsibly.” For “we act not because of a sense that the other person is of equal value to ourselves, but because they are of special and unique value to us.” (*Ibid.*) In taking care of its innovation, a creator will not only shape it and design it according to some acceptable values, but he will also take care of the individuals whose life are altered (directly or indirectly) by means of the innovation, including future generations. This is the core of the virtue perspective, to which Grinbaum and Groves (2013) add a concern for the political and collective responsibility innovators have to undertake as they are the “unacknowledged legislators and co-creators of the world”.

This perspective escapes some of the difficulties of the pure imputation framework. Indeed, it helps to promote a vision of responsibility that now relies on positive abilities of individuals. It rests on a more realistic conception of individual rationality (as decisions are not the result of a pure rational calculus, but are justified by routines and by a constant adaptation to the requirements of the situation as it is assessed by individuals). However, it assumes a substantive definition of responsibility in terms of what is good and right. It supposes that good practices can be defined without giving any information about the way in which these good practices are determined. Are they established by expert discussion between ethicists, philosophers, sociologists, etc.? Are they the result from a collective process? And in this case, how can we hope for an agreement on comprehensive doctrines of the good (to borrow this Rawls’s broad concept) in a context of moral pluralism of values and theories? Are there imposed by a benevolent dictator? Once again the issue of the construction of norms of responsibility is eluded.

4.4. Process.

The second way of escaping purely retrospective and negative conception of responsibility more particularly focuses on the dynamic of responsibility, on the ability to **adapt** and change one’s own action. Two dimensions have been highlighted, **accountability** and **responsiveness**.

As we have seen previously, accountability is linked with the possibility of providing a justification for one’s action as when we have the moral obligation to account for what we did

DRAFT - DO NOT QUOTE

or for what happened (Blagescu *et al.*, 2005; Bovens, 2010; van de Poel, 2011). This would be a first passive way of conceiving accountability as a mechanism that focuses on the relationship between a forum and an agent, i.e. on the obligation of the agent “to explain and to justify his or her conduct” and the possibility of the forum to “pose questions and pass judgment.” (Mark Bovens, 2010). In such a conception the emphasis is made on political and social control and the task of accountability studies will be to explore “whether there are such relations at all, whether these can be called accountability mechanisms, how these mechanisms function, and what their effects are. (*Ibid.*)

However, as we showed previously, to be held accountable for our acts or decisions does not fully cover the whole scope of responsibility. Accounting for our actions not only implies a retrospective attitude or a “moral ledger” as ironically suggested by Ricoeur (1995)²². It also entails an active involvement of individuals as it associates the justifiability of decisions with the possibility of modifying one’s action according to a commitment to accountability (Grunwald, 2010). In Boven’s words, this second view refers to **accountability as a “virtue”** and insists on the active performance of the agents, who take other stakeholders’ needs into consideration in engaging them in a learning dialogue. In explaining the reasons behind decisions to those who are concerned with and in favouring interaction between different actors, virtue accountability is supposed to enhance the sustainability of activities and to open the possibility of better performance. This approach insist on the “process of learning” (opposed to the “mechanism of control”) by which individuals learn to be responsive to each other and to adapt their behaviour in order to achieve “substantive standards of good governance.” Then, the role of accountability studies is to formulate these substantive standards of good public or corporate governance and to assess whether officials or organisations comply with them (Bovens, 2010, p. 9).

²² Ricoeur argues that the “metaphor of counting” is present in the judgment of imputation as the Latin verb *putare* implies a calculus that “suggests the idea of a strange moral accountability of merits and failures, as in a ledger with two inputs: revenue and expenditure, credit and debit with a view to a kind of positive or negative balance. [...] This metaphor of the balance sheet seems underlying the apparently banal idea of accounting, in the meaning of telling tales, relate, at the end of a kind of reading of this strange file”.

DRAFT - DO NOT QUOTE

Here the possibility of modifying our actions, which is logically impossible in negative definitions or responsibility, plays a major role. And it brings **responsiveness** – a dimensions that would have been neglected compared to accountability or liability – at the forefront, as suggested by Pellizzoni (2004). Pellizzoni (2004) distinguishes two different meanings of responsiveness. The first one is related to the individuals' capacity to survive, when individuals react almost mechanically to a situation not excluding opportunistic behaviour (as when we say that a company has adapted to market constraints). The other one implies a real commitment to respond that “entails *previous listening* to a question” as well as “openness, a willingness to understand and confront the other's commitments and concerns with ours, to look for a possible terrain of sharing.” (*op.cit.* p. 19) In addition of what has been highlighted by Boven's framework, Pellizzoni (2004) sees accountability as the possibility, not only to give an account of actions, but more broadly and more actively as an ability to respond to other social actors' needs, values and interest, i.e. to adapt one's behaviour to them.

In conclusion, with care and responsiveness a step has been made towards a positive understanding of the notion of responsibility. Indeed, both conceptions introduce the possibility for actors to adapt their behaviour and decisions to the situation and to **revise** their judgments according to norms. As we mentioned in chapter 1, innovation also entailed a form of responsiveness. We already can see the theoretical pathway opened by the idea of responsiveness. Social actors of innovation and research are responsive in a way that ensure the efficiency of their practices (for instance economic success or scientific praise) but they can also be responsive in a sense that they adapt their behaviour to certain ethical norms (including the avoidance of bad effects, but also the want to better answer ethical needs, as in the example of biomedicine). And, nowadays, ethical requirements can be comparative economic advantages. Consumers agree to pay more for products that comply with ethical norms. They can be sensitive to environmental ethics or to the equitability of the wages of the first producers as illustrated, for instance, by the success of organic food and faire trade. Only narrow utilitarian calculus would have missed these promising possibilities.

However, the limit we pointed out for negative definitions of responsibility remains. How the norms of a “responsible” acts or intentions are settled? By whom? How is it ensured that

DRAFT - DO NOT QUOTE

individuals will follow them? Current definitions of care and responsiveness let these issues pending. In considering the relation of parents towards their child, the care perspective of Grinbaum and Groves (2013) for instance, rely on a kind of virtue ethics that settle the good norms of education. But the specific content of these norms and their justification is not explored. Moreover, virtue ethics suffer from its essentialist overtone, as it does not manage to face the issue of the diversity of the conceptions of the good showed by individuals differing in their culture, in their political and religious beliefs, etc. Finally, virtue ethics can also fall down into the trap of paternalism (and asymmetry) in knowing better than individuals what is right and good for them. The example of education is highlighting since actual debates on this topic show how far from an agreement individuals are.

To sum up, the positive definitions of responsibility add something to the juridical-inspired understanding of the word as they all insist on an ability of individuals or systems to respond to the values and moral conceptions of those who are concerned by innovation and technology. Moreover, they also imply a prospective concern for the future and the possibility to adapt the pathways of technological development according to this (normative) horizon. However, their answers are limited as they all fail to address the crucial issues of the way in which the norms of responsibility are settled. They all promote different solutions in order to favour responsible actions but the way in which what is collectively considered as responsible is elaborated remains completely obscure.

Both current negative and positive definitions of responsibility neglect the central issue of the construction of norms as the latter are supposed to be given from outside. In the model of the sanction, they are given by law, or by routines that settle the amount of the financial or legal sanction. In the case of care and accountability, they rely upon some sort of virtue ethics that settle what the goods practices or activities are in an abstract and *a priori* way, disconnected from actors' values and norms²³. Finally, in the case of responsiveness, the norms establishing

²³ Values, norms and rules have to be distinguished. A *value* is the content of a statement that *can* function as a standard and that entails an *evaluation* that expresses the taste or the preference of an individual or a community (e.g.: 'I prefer walking', 'I enjoy drinking alcohol'). A value is attractive when a norm is prescriptive. A *norm* is the content of an 'Ought' statement that *must* function as a moral, legal or social standard and that entails a *prescription* (from obligations to recommendations or suggestions) for the conduct of an individual or a community (e.g.: 'You ought to take care of your ill mother', 'Don't drink too

DRAFT - DO NOT QUOTE

what is responsible (and what is not) have to be defined, the question of how they are determined, by whom, etc., being left open. For instance, Boven's framework aims at favouring the "substantive standards of good governance", but the way in which the latter is defined remains unclear.

In all these approaches, individuals do not participate to the construction of norms regulating their decisions, although this would threaten the possibility of their application. Individuals are either supposed to act instrumentally, under the threat of sanction or to have a sense of what is the "good" – sense which we don't know where it comes from. The precise mechanism by which individuals follow a norm imposed from above or happen to know the substantive definition of the good promoted by a specific approach of responsibility is completely left in the dark. Finally, both responsibility governed by sanction or ruled out by positive abilities rely on top-down approaches where, on the one side, individuals are compelled to act in a certain way, or, on the other side are told what value to follow by an essentialist framework.

We will return on these problems in chapter 6. Let us now consider the ingredients common to the main existing approaches of RRI.

much alcohol before driving'). » A *rule* is the content of a Is/Ought-statement and that can be universal (e.g.: a scientific law, or a moral law) or particular (e.g.: a rule of life) according to the scope assigned to this rule by an individual or a community (e.g.: 'As for me, I never drink alcohol'). Cf. EGAIS, DEL 4.3, p. 32. In moral philosophy, norms and values are defined and connected according different ways.

DRAFT - DO NOT QUOTE

Chapter 5: Normative elements of actual RRI approaches.

This chapter analyses some of the RRI governance approaches that have been recently developed and studies the different conditions that they require for innovation and research practices in order to ensure a responsible pathway of development. Because this issue has been rather neglected, this chapter aims at scrutinizing the different conceptions of **governance** that underlies most of these RRI approaches, (i.e. the institutional and organizational arrangements that favour or hinder the conditions recognized as necessary for RRI).

From chapter 1 it appears that what is missing to the approaches that paved the way of RRI (such as technology and participatory assessment approaches, corporate social responsibility or sustainable development), is a real collective construction of norms whose efficiency and quality is challenged and evaluated – this evaluation being an important part of the process. RRI approaches are partly constructed precisely to overcome these limitations, as they are deeply grounded in a collective assessment of innovation and research practices.

Indeed, most of RRI approaches focus on the *conditions* that research and innovation should satisfy in order to be considered as “responsible”, conditions among which the **collegiality of the process** plays a major role. This is illustrated by one of the most often referred definitions, provided by René von Schomberg, according to which RRI would be:

“a transparent, interactive process by which societal actors and innovators become mutually responsive to each other with a view on the (ethical) acceptability, sustainability and societal desirability of the innovation process and its marketable products (in order to allow a proper embedding of scientific and technological advances in our society).” Von Schomberg (2011, a, b)

Several elements can be identified in this definition and more generally in the RRI literature. Indeed, in addition of von Schomberg’s definition, Grunwald (2011), Owen *et al.*, (2012), Stilgoe *et al.* (2011), the EC report on innovation (2013) or Sutcliffe (2011) provided with

DRAFT - DO NOT QUOTE

definitions of RRI and frameworks to implement them. Beyond their differences, these perspectives on RRI agree on several elements that we chose to present synthetically instead of recalling each definition. We identified **five ingredients for RRI** (but there are sometimes only four, such as in Stilgoe *et al.*, 2011) that we will present and describe in this chapter. Their critical assessment will be held in chapter 6, once the main elements of our theoretical framework would have been sketched.

5.1. Anticipation.

The first condition that has been put forward in RRI approaches focus on anticipation (EC report on innovation, 2013, Sutcliffe, 2011, Stilgoe *et al.*, 2011, Lee and Petts, 2013). Researchers, policy makers and other members of society have to conduct anticipatory research “to think through various possibilities to be able to design socially robust agendas for risk research and risk management.” (EC, 2013, p. 59). As we previously insisted on the uncertainty and lack of knowledge attendant to current innovation and research, it could seem contradictory to promote anticipatory activities. However, it is not because full anticipation is impossible that the whole task of forecasting should be abandoned. Intending to get the more precise picture of the future, given the available knowledge, can still help to build governance scenarios of responsible research and innovation. Moreover, anticipating is not only useful as a rational activity. It also helps individuals to reflect on ethical issues (Nordmann and Macnaghten, 2010) and more generally on the ends they want to achieve through technology (Robinson *et al.*, 2013). Drawing on the imaginary of individuals (to reveal their vision of the world through narratives), and building different scenarios of development allow for a better understanding of how individuals incorporate technology in their lives and what they expect from it. It helps knowing the values of individuals that will be an essential input in the co-construction of research and innovation’s pathways of development.

DRAFT - DO NOT QUOTE

5.2. Transparency.

The second condition of RRI relates with transparency (René von Schomberg, 2011, a, b). A responsible development of research and technology implies that the available knowledge about technology, its consequences and forecasted uses, the results of deliberation (if the latter is involved), the information stemming from technology assessment or any other tool to engage an ethical and critical reflection on technology, must be accessible and distributed among stakeholders. Transparency, it is argued, allows members of a community to form their own opinion and to reflect on the desirable ends of technology and research. Hence, it favours a richer dialogue (in fact a “multi-logue”), where the various interests and visions of the world of the community members can be expressed and are *taken into account*. However, we already pointed out how this condition might conflict with other constraints. It has been argued, that, ontologically, creating something new could require secrecy. Especially when innovators have to cope with economic imperatives and seek to preserve a reflection area away from the public sphere in order to grasp first mover incomes, in the case of innovation, and the praise of discovery, in the case of research. And, as we will argue in the next chapter, solving this issue requires that we move from a “definition” of RRI based on some “good ingredients” to a governance approach of responsibility in research and innovation that takes into account how norms related to transparency are collectively constructed.

5.3. Responsiveness.

The third essential component of RRI would be responsiveness (von Schomberg, 2011, a,b; EC report, 2013), understood as “the coupling of reflection and deliberation to action that has a material influence on the direction and trajectory of innovation itself” (Owen *et al*, 2013, p. 2). Most RRI approaches insist on the fact that pathways of research and innovation have to be flexible and have to show the capacity of being changed and adapted according to *public values*. We already evoked this dimensions while analysing the different meanings of accountability. Responsiveness for RRI implies that socially or ethically desirable values have to shape the evolution and use of technology and research. Not only

DRAFT - DO NOT QUOTE

upstream once the technology is already widely distributed (as it was the case for GMO's, for instance), but continuously during the life cycle of technology from the very early first concepts to its wide application and commercialization. The idea of “safety by design” (Kelty, 2009), of value sensitive design (van den Hooven, 2013), or the adaptive process promoted in the case of nanotechnology, for instance (Kjølberg *et al.*, 2008; Fogelberg and Sanden, 2008) are good illustration of this focus on “an iterative, inclusive and open process of adaptive learning, with dynamic capability” (Owen et al 2013, p. 9).

5.4. Reflexivity.

The capacity of a system to adapt and to change during its course of development can be identified as its *reflexive* stance (Beck, 1992). And this would be the fourth essential elements of RRI. Researchers, and innovators – at least them – should be able to identify the ethical, societal and political issues raised by the technology they contribute to develop, while also assuming a responsibility in their development. As expressed by the EC report (2013, p. 60):

“Reflexivity asks researchers and innovators to think about their own ethical, political or social assumptions to enable them to consider their own roles and responsibilities in research and innovation as well as in public dialogue. Reflexivity should raise awareness for the importance of framing issues, problems and the suggested solutions.”

Ascribed to individuals and systems, RRI approaches claim for an ability to acknowledge the ways in which problems at stake are sized and answered.

DRAFT - DO NOT QUOTE

5.5. Collective/Inclusive/interactive.

Closely related to this issue, the last dimension that is outlined by all RRI approaches focuses on the interactive process by which innovation is scrutinized. Whether it should be interactive as in von Schomberg's definition or 'inclusive' (or 'collective') as in Mitcham (2003), Owen *et al.* (2012, 2013), Sutcliffe (2011) or the EC report (2013), those terms, if not equivalent, outline the need for involving various (when not all) actors affected²⁴ by an innovation, at an early stage. This would include end-users, researchers (including social scientists), civil society, NGO's, industry, policy makers, etc. As expressed by Callon's idea of a distributive and participative innovation: "One innovation is responsible as soon as it is attentive to all concerns and suggestions that shall be expressed in relation to it." (Callon, 2011, p. 22).

Against a top-down approach of technological drive, RRI conceptions emphasize the importance of multidimensionality "allowing innovation to be developed in a *co-building* mode that 'ensures *co-responsibility*'" (Owen *et al.*, 2013, p. 9 emphasis added).

Since in a perspective of science and technology *within* society, the adaptive process required by responsiveness and reflexivity can no longer rely on approaches where ethical norms are imposed from legal frameworks, already existing and *a priori* norms or from the fear of sanction, deliberation and participation are claimed to have many advantages. First, the inclusion of stakeholders into the making of the norms regulating scientific practices and innovation processes has a heuristic value. When it integrates individual narratives, it prevents conceiving ethical issues as a "rote" process where terms are listed without any relation with their context. Instead it helps closely linking ethical issues to the motives, interest and the desirable ends of individuals (Ferry 1991; Nordmann and Macnaghten 2010). According to Owen *et al.*, 2012) this process should lead to think the purposes of technology (science for society) focusing on what we want rather than what we do not want. Secondly, deliberation and participation allow for individuals to better accept and respect norms, increasing their efficiency. Constructivist approaches all make a plea for participation and deliberation as a

²⁴ We should add "concerned".

DRAFT - DO NOT QUOTE

way of ensuring the political legitimacy of innovation in research, its economic and scientific accuracy and finally its ethical desirability.

To sum up the inclusion of stakeholders into the making of the norms regulating scientific practices and innovation processes should achieve several goals: it helps 1) defining and revealing what are the actors' values and the ends and purposes they assign to science and technology, 2) co-establishing norms from these values, 3) shaping the design of innovation and research processes and outputs.

To conclude this brief overview of the conditions traditionally assigned to RRI, it can be worth to mention that if all these elements are not new, their combination early in the process of development of technology and research is: "It is the institutionalized coupling of such integrated processes of anticipation, reflection and inclusive deliberation to policy- and decision-making processes—i.e. the dimension of responsiveness—that is an important, if evolutionary, contribution that RRI makes" (Owen *et al.*, 2012, p. 755). This would be the most important contribution of RRI.

DRAFT - DO NOT QUOTE

Chapter 6: Norms in Contexts.

As we have seen in the general introduction, one of the main issue of responsible innovation and research is related with the way in which norms of responsibility²⁵ are (collectively) constructed and how they are implemented. Before, we highlighted the relevance of various dimensions of responsibility and of innovation that have been put forward recently in the context of RRI. Now we need to add a reflection on the governance of RRI, that is, on the processes by which responsible innovation is conceived, assessed and implemented. For instance, as participation is one of the most important pillar of responsible innovation and research approaches, it is necessary to both consider the definition of what RRI is (chapters 3 and 5 mainly) and the governance issues it raises (how to implement participation, for instance). Those questions cannot be disentangled. Analysing RRI definitions implies to examine its governance characteristics simultaneously. The second section aims at challenging the RRI approaches proposed so far in light with the relation they build between norms and their contexts. Even positive understandings of RRI rely on norms that are disconnected from individuals' normative horizon and, in this sense, do not provide with any answer to the practical issue of the implementation and efficiency of norms. For this reason, we seek to analyse how norms can be elaborated in their context, in a way that does not already presuppose the boundaries of the problem. The problem underlying the construction of a norm should be opened up, leaving room to modify the question it raises, the data at stake, its warrant, the possible exceptions to this norm as well as their modalities of application (to borrow from Stephen Toulmin's model of argumentation, 1958). This chapter reflects on the process of construction of the context and on the role of reflexivity in order to sketch the main line of a governance approach that will help us to critically review RRI theories.

²⁵ We let open the problem of the links and distinctions between norms and responsibility.

DRAFT - DO NOT QUOTE

6.1. RRI Governance.

A first clarification of what we understand in the GREAT project as ‘governance’ might be useful, especially because the word – like responsibility – is polysemic and widely used in many different contexts (as illustrated by expressions such as corporate governance, green governance, international governance, etc.).

The term has been generalized during the 1990s expressing among other, a change of focus, from government to governance, i.e. from a perspective where regulation is grounded in control, centralized and top-down relationships between the State, the citizens and all types of institutions (private or civil) to a view where dynamic modes of organization, interaction and participation play an important role. Understood as the political practices and modes of organization of private corporate or public administration, it stems from the idea that political power is not the only way of coordinating actors in the political spheres, and that other actors can contribute to economic and social regulation. At the meta-level, the coordination of individuals addresses the problem of governance of different spheres: political, economic, research, civil, industrial.

As recalled by ETICA project (Del 4.3.), governance can be defined as “the reflexive self-organisation of independent actors involved in complex relations of reciprocal interdependence” (Jessop, 2003)²⁶. More recent EU developments further qualify this mode of co-ordination as democratic, participative, and pragmatic, and as recommending support for collective action (Maesschalck, 2007). Also highlighting its dynamic and participative aspects, Lyall, Papaioannou, & Smith (2009) see governance as a way to describe “the changing nature and role of the state in advanced societies and the changing boundary between state and civil society”.

²⁶ Here, we face the issue of the distinctions and relationships between dependence and interdependence. Dependence and interdependence can be seen as contradictory. But it seems preferable to hold a conception of bounded actors who are constantly entangled in (inter)dependencies and subject to explicit or subtle forms of powers.

DRAFT - DO NOT QUOTE

These different approaches all have in common to emphasise the increasing importance of the involvement of stakeholders in policy research, as opposed to the traditional “top-down imposition” forms of policy formation and implementation. Moreover, the idea of governance integrates the active role of actors in the making of the norms and rules they will have to follow. For this reason, in the context of uncertain technology and research development, governance can be seen as an attempt to answer a “trilemma”²⁷ between “scientific accuracy, policy effectiveness and political legitimacy” (Pellizzoni, 2004), i.e. between the rules of scientific knowledge, the efficiency of political norms and rules, and their social acceptability.

The focus on the governance aspect of RRI theories will allow us to highlight what are the different “solutions” brought up to answer this trilemma. In doing so, we move beyond the initial question of what is responsibility for innovation, to the issue of the governance of responsible innovation or how to determine the conditions for a responsible governance of responsible innovation. These questions are central and original in the GREAT project.

In the previous chapters (4 and 5) we concluded that most approaches of responsibility in research and innovation were eluding the question of the construction of norms in their context, and therefore were not (or only partially) dealing with the practical aspects of RRI. What is needed now, is to know in a much more concrete way how the process by which norms of responsibility are established, is thought in relation with the context. This would ensure that individuals will reach an agreement on *their* interpretation of norms (and not only norms themselves) and that they possibly will follow them. As we already emphasized, any process of construction of norms, *a fortiori* the norms of responsible research and innovation, will have to deal with moral pluralism (Reber and Sève, 2006), where individuals can have conflicting values on definite subjects but also ground their normative horizons on different ethical theories (mostly consequentialist frameworks, deontological theories, virtue ethics, or some forms of intuitionism).

²⁷ The word, here, is broadly conceived and designates three layers that can oppose to each other. It is not a real moral trilemma.

DRAFT - DO NOT QUOTE

Among the different propositions brought up to deal with our pluralist modernity, procedural theories, such as Habermas' discourse ethics, have played an important role in reintroducing the need of communication and in partially taking into account the context²⁸. In Habermas's approach, for instance, due to their communicational capacities, social actors engage in a discourse process and achieve agreement through a process of revision that is submitted to logical rationality. Discourse ethics links up the individual and the community's will, without supporting any substantial particular statement on the content of moral rules or ethical ends. It therefore avoids relying on comprehensive doctrines, to use Rawlsian terminology that will fail to face the challenges of moral pluralism (as it is the case for virtue ethics)²⁹.

However, proceduralism encounters several limits. Deliverable D.2.3 will address this issue in depth, using Maesschalck and Lenoble's approach, which developed a rich reflection about the relation between norms and their context. For now, let us only mention that one of the difficulties faced by proceduralism lies in that it rests upon the logic of the best argument according to which, once the individuals agreed on a norm on ground of its ethical relevance, there is no particular issue concerning the implementation of the norm. Moreover, Habermas, and with him the tenets of deliberative democracy, do not define the requisites for argumentation: i.e. how to define a suitable argument, and more generally what are the rules of a good argumentation. With the RRI problematic we need to deepen the possibilities of co-argumentation in interdisciplinary contexts (Reber, 2010b).

Another issue, lies in the fact that the procedure of discussion within a community, be it an informed community like a Parliament, for instance, is not sufficient to warrant the relevance of norms to a wider community who is supposed to apply them. There must be something more than a mere procedural discussion to elaborate a rational justification of a norm and almost simultaneously adapt it to the specificity of a social and cultural context. Here, we face a first problem with two possibilities: the translation of a norm from one context to another, or its translation by extending one context to a broader one. A second problem arises: whatever rational a norm might be, its relevance does not presuppose that it will be followed by social

²⁸ Habermas is not a sociologist, but a philosopher of social theories. For this reason he does not take seriously the context and its different potentialities.

²⁹ See Reber, 2012a.

DRAFT - DO NOT QUOTE

actors and that its moral legitimacy will be acknowledged. Reason doesn't contain in itself its conditions of application. The rationale and arguments given to support norms are insufficient to ensure compliance.

“The treatment of rational acceptability of norms ignores their practical acceptance. But this is precisely the point where the emphasis of the political value of the norm takes part. One thing, for a norm is to be acceptable in principle, another is to be valid in practice.” (Maesschalck, 2001, p. 83).

All in all, dialogical procedures cannot avoid the objection of the limited relevance of norms to one context, even though their supporters claim for the moral norms to be regarded as universal rules, or at least, universalisable ones. As it will be extensively analysed in Del 2.3, the semantic content of statements is not important in these approaches, to which, only matters the approval process by which concerned parties come to validate the norm.

6.2. Determining the context.

This conclusion brings us to the fact that one of the main elements that has been neglected by procedural theories, whether Rawls' or Habermas', lies in the role they give to the context. The ethical reflection is supposed to take place within a given and identifiable context. It is presupposed that context will supervene through the properties of dialogue and discussion.

“The epistemological insufficiency of every theory that supposes the context as given or identifiable is important because such presuppositions, even in the form of conventions that are adaptable or revisable by an individual, don't take into account the reversible or reflexive character by which *one gives oneself* this preference, this convention or whatever it is that makes this ability to adapt or revise possible.” (Lenoble and Maesschalck, 2003, p. 90-91).

DRAFT - DO NOT QUOTE

Within proceduralism, discourse ethics presuppose that the discursive and rational construction of a norm, which could be considered as relevant, is capable by itself of taking into account all the possibilities that are available for a social context to be regulated. But this presupposition is highly problematic. It ignores the fact that the concrete choice of a norm, even when judged relevant following a discursive process of reason, necessarily results from an operation of selection of the possibilities at stake. This operation of selection is prior to the simple discursive operation of reason and therefore, depends on something different than it. Individuals agree (or disagree) on norms according to their values (or other normative foundations) and their visions of the world. This set of values and preferences is not immutable and given once for all. It evolves constantly. And the way in which we reflect on our values and on norms cannot be overhanging. Moreover, as expressed by Mark Hunyadi (2012), with the help of a cinematographic metaphor the context is not a pure exteriority. It is never simply given but rather constructed by individuals.

“Context is not a mere landscape in front of which we move in a detached way, as in old movies where actors play at the foreground of an image that is projected behind them to create the illusion of a decor. The context, quite the opposite, we are woven to it, and it weaves us back; we never can fully cut ourselves off from it neither tear us from it” (2012, p. 26).

To sum up, context, as it conceived in GREAT project, is not just what we see outside, i.e. the environment in which the decision is to be made. It is not just the problem of what one can perceive *as* the environment, since **we address the environment from our own perspective and framing**. Behind the second metaphor of this quote - the woven canvas - lies the idea that the context is also part of our framing, our background, and our stock of knowledge.

This will mean that, to elaborate norms in taking the context into account, we will have to study the cognitive framing of individuals, i.e. the way in which people perceive the contexts in which they operate, keeping in mind that context has a **descriptive and a normative part**, the latter helping to cast a critical look at the former.

DRAFT - DO NOT QUOTE

The cognitive framing of stakeholders sets boundaries on the parameters of discussion among them as it partly determines the ways in which dialogical engagements progress. Individuals think about norms from their particular point of view. They construct their context, and this will affect their positions on norms. So we need to understand the ways in which agents conceive of their own possibilities from which they will elaborate the norms.

This interrelationship between **the things we understand and the way in which we understand them** can be illustrated by both our historically situated conception of risk and the way in which we assess responsibility, which are social constructs. Drawing on Kermish (2012), Doorms and van de Poel (2012, p. 6) outline that “since each culture has a propensity to select those dangers that contribute to the stabilization of the corresponding social organization, thereby “translating” these into risks, while neglecting other dangers, risks are the result of an interaction between social processes and the external world.” Risks are not given objectively, they depend on the ways in which social actors conceive themselves in an environment and on their interpretation of this environment. This shows how crucial is to understand and reflect on the way in which we settle a problem such as risk.

From an epistemological viewpoint, this linkage between social actors and their context has been widely neglected by procedural (but also pragmatist) political philosophy. And finally, a sound epistemological theory of the relationship between actors, the formation of norms and their context it still missing in a way that threatens the possibility of implementing norms in an efficient way.

6.3. The role of reflexivity.

At a cognitive level, in order to conceive in a more appropriate way our relation to the context, we need to introduce the possibility for the agents to be reflexive and to revise not only their judgments as in Rawlsian reflective equilibrium, but also the way in which they

DRAFT - DO NOT QUOTE

size and understand the problem (its epistemic and normative dimension). The possibility of revision is an important bet in deliberative theory of democracy. Indeed if we don't agree to change our mind in front of better arguments it is useless to enter in such process. Better continue on other ways using bargaining, or force reports. Supposing reflexivity is essential to ensure participation in RRI.

To illustrate the role of reflexivity, Mark Hunyadi (2012) highlight how our trust in a situation – trust given by a set of implicit knowledge we share, which makes us walk in a pavement, for instance, relying on the fact that the pavement will not disappear or break – can be broken apart by some disruptive moments. Similarly, our trust in the context is sometimes jeopardized by an unexpected event, a “dissonance” that substitutes to our automatic experience of action, a reflexive stance that allows us to take some distance with the situation. In this process, we suspend the automaticity of our action and elaborate on this experience and on the issues we had to handle (*op.cit.* p. 119). Here, we see that an appropriate conception of reflexivity will rely on a **theory of learning**, i.e. on a theoretical background describing how individuals learn or how they learn to learn (cf. Del 2.3). Moreover, we should associate descriptive knowledge (what it is, and will be) and normative knowledge (what ought to be).

By introducing the possibility of a reflexive stance in actors' reflection about norms, we are in quest of the conditions by which individuals can really carry out a reframing operation. Now what do we precisely mean with reflexivity? To Johnson (1977, p. 172): “To be reflexive” in its most elementary meaning is the capacity to turn or bend back on oneself. Reflexivity refers, at least in a methodological sense, to “the mutual interdependence of observer or knower to what is seen or known”. But reflexivity is not only attached to individuals. It can also concern larger social phenomena, such as in Beck's analysis of our late modernity where the premises, structures and institutions taken for granted in first modernity are questioned and reconsidered, opening, among other, the possibility of controlling the development path of economies or technologies. Integrating the idea of an adjustment between perceived needs and the situation, Lenoble and Maeschalck define reflexivity as the operation by which a

DRAFT - DO NOT QUOTE

social group seeks to respond to its perception of a need in order to adjust its capacities for action. These adjustments comprise descriptive and normative perceptions.

But let us give an account of what is being revised and adapted through reflexivity, expressed in another way than Hunyadi. According to Voß and Kemp (2005), there would be a first-order reflexivity, a kind of ‘reflex like’ reflexivity that would capture the unconscious and unintended consequences of industrial modernization – what Beck labels the ‘self-confrontation’ aspect of reflexive modernization (see Beck, 1994). In contrast, second-order reflexivity would refer to the self-critical and self-conscious reflection on processes of modernity, in particular our instrumental rationality. This deeper process of reflexivity assumes a sense of agency, intention and change, since actors reflect on and confront not only the self-induced problems of modernity, but also the approaches, structures and systems that reproduce them.

In other word, first order reflexivity only concerns the collective definition of the problem at stake (involving interaction between the governing organizations and their citizens) and, possibly, the solutions that can be applied. It is a first step towards political legitimacy as it implies a collective process of defining what the problems are, who the social actors experiencing them are, and what the solutions could be. But second order reflexivity assumes a more radical process of critical revision as it introduces the possibility of reflecting on the structures (the framing) that produce our **appreciation** of a problem. Here we refer to the capacity of actors to identify the various effective possibilities on which the operation of the selection of the norm will be carried out. **Actors not only reflect on the adequacy of their norms and values, but also on the way in which they construct these norms and values.** Second order reflexivity supposes that we can “open up” the framing of the context, i.e. not only the way in which actors relate to their context (their identity), but their means for justification, when they are elaborating norms. These norms can be focused on what is right – or false- (epistemic norms) or what is good, just or evil, unjust (Reber, 2011b).

In this sense, governance will not only have to manage or articulate different spheres of society (politics, civil society, research, industries), it also has to articulate different spheres of

DRAFT - DO NOT QUOTE

knowledge (ethics politics, economics, science) each of which has its own relevance and its own methods and types of argumentation. In other words, reflexivity does not only concern inter-individual or inter-institutional interactions, but also inter-epistemic ones (Reber, 2012b).

Then, what is needed not only includes a reflection about our own actions (as individual or as a society), but a reflection on how the presupposition, the governance principles and the values determine our way of acting. This is why the GREAT project aims at promoting a **reflexive governance of RRI**, through an institutional and organizational structure that will help individual to be reflexive, self-reflexive or even co-reflexive.

To sketch some of the elements of a reflexive governance of RRI, let us turn back to the different approaches of RRI presented in chapter 5 that we can now assess in light with the perspective we just presented, i.e., analysing the type of relation to the context they are relying on and evaluating their level of reflexivity.

6.4. Governance regarding normative reflexivity and context in existing RRI approaches.

As we have seen, taking the context into consideration is a key step toward the construction of norms, if we want them to show some efficiency. How the different approaches of RRI presented in chapters 4 and 5 do conceive the relation between norms of RRI that will emerge and the context? This is what we will examine now.

At a general level, RRI *definitions* face the same kind of issue than definitions of responsibility (chapter 4). Determining the different conditions that a process (RRI) has to satisfy is a first interesting way of reconsidering the link between society and innovation, since it helps to identify some of the necessary elements without which no responsibility in innovation and research could emerge. But even being as precise and comprehensive as possible, the form itself of the definition when conceiving RRI leaves several issues aside.

DRAFT - DO NOT QUOTE

This approach does not take into account the way in which values and norms are embedded into a collective process to shape innovation.

If we reconsider the five ingredients of RRI, the condition of anticipation and transparency come up first against the Collingridge dilemma of knowledge we already mentioned. Indeed, if it seems undisputable that knowledge about new technologies or new product is needed if early assessment is demanded, the boundaries between what is imposed by our liberal and capitalist way of producing science and technology and the social need to ensure co-participation in including other stakeholders will have to be settled according to the context. The issues of transparency and anticipation cannot be dealt in an abstract and decontextualized way, since the schedule according to which information is disclosed, and the content of information must be subjected to a form of social consensus. For instance, in bio- and nanotechnology, the precise line between what should be hidden for economic purposes (to protect innovation) and what should be disclosed for health and safety purposes (to protect individuals and their environment) cannot be decided *a priori*, and should result from an agreement between the different stakeholders. In a way, this contributes to make *a priori* approaches of responsible innovation and research processes irrelevant because transparency and anticipation are required in an abstract way that does not reflect on the practical constraints of the context.

Second, but both issues are related, a responsive dynamic of innovation systems supposes that technology can be submitted to incremental changes with no irreversible and disruptive upheaval. This would require that technology has a “soft” design that allow for reversible decisions and deep modifications, which might not always be the case (as illustrated by the internet revolution or by the control of atomic energy for example). When innovation and research are not reversible, their responsive capacity is reduced, also calling for a closer look at the context (the reversible characteristics of a specific innovation process). Again, settling issues such as the “degree” of responsiveness or transparency *a priori* and in a decontextualized way can by no means ensure that an appropriate solution to the situation will be found.

DRAFT - DO NOT QUOTE

The third issue raised by most approaches of RRI concerns the way in which the norms framing and regulating innovation and research practices are collectively decided, in order to achieve “ethical acceptability” and “societal desirability”. Interestingly, some of the previous conceptions of RRI, although claiming for collective processes in research and innovation, defend substantive (as opposed to proceduralist, for instance) ways of defining norms of responsibility. Von Schomberg’s approach, for instance, excludes a way back to Aristotle’s abstract “good life” and avoids to be grounded on essentialism (such as in Martha Nussbaum’s approach, for instance). However, and this is also true for EC report, it rests on a substantial definition of “good” and “right” practices of innovation and research, which come from the democratically agreed values of the Universal Declaration of Human Rights or of the Treaty on the European Union. In the case of the European Treaty, these values promote, among others, “a sustainable development of economic activities” in “a competitive market economy”, and advocate for achieving quality of life and quality of the environment, a high level of employment, social progress, high level of protection, *etc.*

These conceptions have the practical advantage to rely on norms resulting from a consensus at least at the European level of decision³⁰. However, they suppose that the agreement on a general *normative horizon* (embedded in the European ideal of sustainability for instance) has already been achieved. And here, we could wonder what the purpose of promoting the inclusion of many stakeholders is, if the normative horizon is already presupposed. In other words, European citizens might have agreed on the general framework promoted by the European Treaty. But does the latter fully cover the conscious or unconscious background that shape social actors’ decisions, i.e., their ideal conception of how the world should be? And if not, is it not the case that a RRI governance approach should reflect on the way in which the individuals holding different normative horizons will come to a co-construction of norms? Finally, this ideal of sustainability might be difficult to implement in practice, and leaves us with no answer in front of the possible conflicts it might provoke with economic imperatives.

³⁰ This statement should be nuanced when we see the distrust of some public national opinion majorities – like in France- that have rejected the Project of European constitution, precisely laid on these Rights.

DRAFT - DO NOT QUOTE

A fourth issue emerges with the idea of **participation and deliberation**, which is anchored in a recent tradition of political philosophy and political theory. Participation and deliberation have often been presented as ways of dealing with the issue of moral pluralism entailing a double rejection of monism and relativism (Reber and Sève, 2006). Both processes help to acknowledge the positive role of conflicting values and value systems among individuals, on the one side, and, on the other side, the need to answer normative issues inside the realm of ethics, i.e. not delegating their answer to group loyalties, cognitive bias, interests, religious or national particularities (c.f. glossary Pluralism and Deliberation). However, the question of the real efficiency of deliberation and participation processes is not raised within contemporary conceptions of RRI. Rather, the involvement of stakeholders is presented as the solution that *in fine* will warrant the responsible dimensions of innovation and research (Stilgoe et al., 2013, for instance). Once again, the practical means by which deliberation and participation are implemented (how questions, stakeholders, experts are selected, what is the strength of the norms resulting from deliberation, what are the different protocols chosen, etc.) are not investigated by current RRI research.

Finally, most of contemporary RRI approaches are problematic for they **presuppose the capacity of reflexivity** of the actors to be already existing due to a formal method, such as argumentation, deliberation, debate or discussion. They never challenge the concrete steps by which deliberation will lead to the elaboration of norms and eventually to their following.

Yet, the conceptions of reflexivity that is put forward in RRI approaches is not blind to the issue of the “framing”, as it is illustrated by the definition of the EC report (2013) we already mentioned according to which “researchers and innovators [have] to think about their own ethical, political or social assumptions” (*op.cit.*). However such a conception does not go as far as required by the definition of the second order reflexivity we developed in 6.3. The role of the context (including the way in which we conceive the issues at stake) is not fully taken into account, and the possibility of revising not only our judgments in front of a problem but the very manner in which we conceive the problem is not seriously investigated.

DRAFT - DO NOT QUOTE

Chapter 7. Conclusion. Defending responsibility as a polysemic concept embedded in a contextual and reflexive governance.

With the definition of second order reflexivity (presented in the precedent chapter) in mind, it appears that however rich and innovative, compared to traditional models of technology management, the different approaches of RRI leave some important issues in the shadow.

First they do not question how the problem is defined. Secondly, they do not deal with the legitimacy and implementation of norms as they do not ensure that the participative and deliberative process provides with norms that the members of society will find acceptable, choose and follow. There is no “opening” of the framing (descriptive and normative), i.e. no interrogation on the way in which the precise context of RRI is constructed. Yet, the issue of context is not totally absent of RRI approaches as illustrated by Owen *et al.* (2013, p. 13).

“There are in fact numerous ways of implementing the dimensions of responsible innovation, and [...] as an approach it should not be strongly prescriptive, or rules based in its implementation (Richardson, 1999). Beneath the general framework researchers, innovators and those who fund them should have flexibility in the details of how its dimensions are taken forward, in creative and imaginative ways that suit its context of application best and that they themselves value.”

However, the way in which the context is understood and constructed is absent of the reflection. Therefore, RRI approaches do not fully address how participative and deliberative process will be efficient and will effectively shape the design of technology in a way that is ethically and socially acceptable, because they all presuppose their own required conditions and as such do not necessarily involve reflexivity. Therefore, it will be important to make sure that every application of a norm presupposes a formal moment of choice of its acceptable normative constraints, and a selection operation by which the possibilities that are the basis

DRAFT - DO NOT QUOTE

for the construction of the norm are chosen according to an acceptable way of life within the community concerned.

7.1. Case Study: the limitations of INDECT project.

To illustrate our aims and give some of the limitations that can occur with top-down approaches where the experts frame all the normative questions that can be raised in a context of innovation, let us briefly study a specific case: the EU INDECT project³¹. The goal of this security project “is to develop advanced and innovative algorithms aiming at *human decision support* in combating terrorism, and other criminal activities, such as human trafficking, child pornography, detection of dangerous situations (e.g., robberies) and the use of dangerous objects (e.g., knives, guns) in *public* spaces” (DEL O8, p. 6). The project seeks to develop detection tools to increase European security while, at the same time, preserving privacy of the citizens as well as their freedom. It tries to combine two different elements: a) satisfying a social need for safety b) within the limits of an ethically acceptable framework (respect of privacy). On the one side it strives for improving the fight against terrorism and criminalities through innovation in the field of detection tools (video monitoring, plate’s analysis, web monitoring, etc.). On the other side, since this innovation implies an intrusion into the privacy of individuals, the challenge is to develop innovations limiting the level of intrusion to what is considered as just necessary, and also to prevent from unauthorized access attempts.

To achieve this goal, the main tool promoted by INDECT project is to address ethical issues through an ethical board (EB). The board is interdisciplinary and gathers policy officers, human rights layers, one ethicist, and representatives from NGOs or data protection. The role of the EB can be summed up as following: a) ensuring compliance with existing (national and international) legal constraints, b) providing norms to frame testing on human subjects, and c) raising the relevant ethical issues of the project (which entails and assessment of the ethical content of the project and the bringing about of recommendations).

³¹ Cf. Annexe 1 for a full analysis of this project.

DRAFT - DO NOT QUOTE

Through several meetings, members of the ethical board identified and discussed some of the ethical issues raised by the project and provided with some recommendations that had the power to shape the design of the experiences conducted on the behalf of the project (for instance in limiting the disclosure of data when it was not absolutely needed). Indeed, a close analysis of the project (c.f. Annex 1) shows that an attempt to cope with an ethical norm of protection of individual data from unnecessary disclosure has helped to shape the process for collecting and diffusing information and the way in which technology is designed and used in the project.

However, some limitations can be pointed out. First, the EB had to assess the ethical content of the project, but not the ethical relevance of a technology of detection that would be widely and daily applied within the European Union. This is an important limitation of the work. Indeed, the INDECT project could be considered as “responsible” for it meets some ethical norms considered as desirable while the technologies that would be extracted from the project and distributed among the European territory would not. To extend its ethical relevance the enquiry should have also explored the impacts of a new technology of detection being generalized within the EU.

Second, if the various ethical issues (such as personal data protection or the procedures needed to involve individual subjects to testing) are analyzed in close relation with the specificities of the project, implying a contextualized relation to norms, no mention is being made on how the ethical issues are raised. They come from the input of the different members of the EB. However the process by which these ethical issues are put forward is not challenged. Also not challenged is the way in which EB members provide researchers with recommendation and ethical guidelines. This could have been done in a more collective approach (involving the different work packages members). But it is not the case, as the EB is the only institution in charge of enouncing ethical norms. And even if the EB shows a capacity of revision of its conclusions (among other, in changing its composition), the way in which it defines what is socially and ethically acceptable is not enquired.

DRAFT - DO NOT QUOTE

In conclusion, the INDECT project appears to make a first move towards responsible research in allowing for responsiveness (as the project managed to be shaped by norms) and interdisciplinarity. However, involving experts of different disciplines does not amount a real inclusive enquiry entailing a comprehensive participation of stakeholders. Moreover, the project rests on a top down model of governance that supposes guidelines to be decided by expert and then being applied by members of the project. According to our analysis of the available INDECT documents – since, in practice, there may have been instances of reflexivity that these texts do not capture - there is no reflexivity concerning the process in which the ethical norms are obtained and no involvement of potentially relevant stakeholders for a future wide implementation of the technology. Then, as the level of deliberation is poor, as there are only rare attempts to justify and produce arguments grounding norms, and as there are no inference between reasons and decisions, we are far from a reflexive inter-actor governance of inter-institutional governance.

The INDECT project is therefore a good illustration of the limit of ethical issues in science and technology being only sized with the help of an expert committee, even if the latter is responsive and interdisciplinary. In INDECT, no mention has been made of responsible innovation. The focus was rather on ethical issues. But our conclusion will be even truer for RRI and shows again the urgent necessity to conceive RRI as a process of norms' construction whose efficiency in implementing these ethically desirable norms has to be assessed.

7.2. Proposal for a positive participative, interactive and reflective conception of responsibility in context.

In interpreting the difficulties that are encountered by the perspectives of responsibility we have analysed, it can be worth to emphasise that they were all in quest of a definition, i.e. of a theoretical approach that outlines the different *dimensions* of responsibility. However, the issues raised by responsible innovation and research cannot be solved by following only a theoretical definition. Favoring RRI can only result from a *practical* methodology made possible by an appropriate governance approach. Therefore,

DRAFT - DO NOT QUOTE

moving from the theories of responsibility, we have analysed and assessed the (mostly implicit) governance models in RRI existing approaches that have been put forward in order to determine the way in which responsible innovation practices are defined and assessed. In other words, we have been focusing on the ways in which RRI approaches think how RRI practices are implemented, facilitated, and favoured through institutional and organisational arrangements. We have discovered and shown that these conceptions carry out blind spots.

Let us add a few concluding remarks related to the possibility of developing a positive and active conception of responsibility. Indeed, the method we conducted to analyse RRI, working as a “negative theology”³², gave us some hints.

First, from the paradox of knowledge, we learned that a binary conception of knowledge (as absent or present) failed to engage responsibility. Indeed, to avoid that ignoring the consequences of our actions prevents us from responsibility, we would need to define a **collective definition of an acceptable level of knowledge** from which responsibility can be engaged. To ground political, scientific, economic and technological decisions, we must define – and this can only be collectively – a *reasonable* or *acceptable* level of knowledge. In other words, from a practical point of view, it is necessary to determine a set of knowledge that actors must acquire, lying in between a threshold that would be too low and that would **erase responsibility** and an **unlimited responsibility** that would dilute it, as we have seen.

Moreover, to prevent us from a further dilution of actors’ involvement and commitment and to bypass the ‘many hands’ obstacle (chapter 3), a conception of “collective”, “distributed” or “**shared**” **responsibility might be useful**. The network of micro-actions and decisions of individuals involved in a project but also the institutional environment as well as the different public policies that stir or break innovation and research have to be considered together. In addition, the idea of a **gradual conception of responsibility** opposed to the traditional binary approach could be explored, as in Coeckelbergh (2012), for instance. We would have to collectively decide *to what extent* somebody can be held responsible or what the degree of

³² Christian theological approach that attempts to describe the nature of God in a negative way, i.e. in focusing on what God is not rather than what God is.

DRAFT - DO NOT QUOTE

his/her responsibility is. Yet, in answering these questions, the crucial issue of the setting of norms and of the process by which they are established remains, calling for an approach of responsible innovation in terms of governance as we have explored it in this deliverable.

Now, the practical perspective of the *degree* by which responsibility is or is not involved or the *acceptable* level of knowledge that could be required cannot make us forget the ontological dimension of the problem. This dimension can be analysed with the help of three layers that contribute to build a more positive understanding of responsibility. These three layers will be only briefly mentioned here as a concluding remark of the document since they will extensively be analysed in D 2.3.

First, at the very beginning, “my” actions make “me” epistemologically and ethically responsible in front of the others. This would make a plea for considering our actions (from which innovation and research are only a subset) as ontologically ethical. In this sense, responsibility cannot be avoided because “I” always *participate* (commitment) when “I” act. **Participation** (or commitment) would be the first layer towards a more positive definition of responsibility. The innovator, as an author asks for an intellectual right. In the same way, responsibilities are attached to him as an author.

Secondly, when we act and therefore participate, we are not passive, like when a stone falls from the mountain peak. Our freedom and our ability to revise judgments are engaged. This capacity of revising our judgments – very broadly conceived as our **reflexivity** – makes our action the fruit of a conscious decision-making process and opens the way for us to understand the weight and load of the potential consequences and outcomes of our actions.

Finally, my actions do not only affect me but also others, whether human beings, future generations or the environment at large. Therefore, we cannot avoid the **interactive (intersubjective)** aspect of responsibility that comes from the normative horizon in which the subject is embedded.

DRAFT - DO NOT QUOTE

Of course, this proposition – drawing a more adequate responsibility concept on three elements: participation, reflexivity and intersubjectivity – is only very schematically sketched here. It requires further development and rationale, which will be the aim of next deliverable.

Del 2.3. will investigate the institutional conditions that will allow for responsible reflexivity in complex context processes. In order to set the basis of a framework where norms can be co-constructed and then implemented and followed by individuals, the next step of the reflection has to investigate how RRI can rely on a participation of stakeholders that is no “flat” or simple, combining descriptive and normative framings. To this end, we will have to reflect on a way of constructing participation as an effective and specified process subject to assessment, a process that would allow the various social actors to revise their judgments and to adapt their actions and decisions according to their normative horizons.

For, in taking seriously governance with reflexivity in context, we aim at departing from *governance of RRI* to *responsible governance of RRI* and to move from sciences *for* society to science *within* society.

DRAFT - DO NOT QUOTE

References

- Adam, B. and Groves, G., (2011) Futures Tended: Care and Future-Oriented Responsibility. *Bulletin of Science, Technology & Society*, 31(1), pp. 17-27.
- Armstrong, M., G., Cornut, S., Delacôte, Lenglet M., Millo Y., Muniesa F. , Pointier A. and Tadjeddine Y. (2012) Towards a Practical Approach to Responsible Innovation in Finance: New Product Committees Revisited. *Journal of Financial Regulation and Compliance*, 20(2), pp. 147-168.
- Barben, D., Fisher, E., Selin, C., &Guston, D. H., (2008) Anticipatory governance of nanotechnology: Foresight, engagement, and integration, in O. A. Edward J. Hackett, M. Lynch and J. Wajcman (eds), *The Handbook of Science and Technology Studies* (Third ed., pp. 979-1000), Cambridge, MA: MIT Press.
- Bessant, J. ,(2013) Innovation in the 21st Century, in R. Owen, M. Heintz, and J. Bessant, (eds), *Responsible Innovation*, Chichester, UK: Wiley, pp. 1-25.
- Bovens, M., (1998) *The Quest for Responsibility: Accountability and Citizenship in complex Organisations*, Cambridge: Cambridge University Press.
- Bovens, M., (2010) Two concepts of accountability: accountability as a virtue and as a mechanism. *West European Politics* 33, pp. 946–967.
- Callon, M., (1998) Des différentes formes de démocratie technique. *Annales des Mines*, January, pp. 63-73.
- Callon M., Lacoste A., (2011) Defending responsible innovation. *Debating Innovation*, 1(1), pp. 19-27.
- Carroll A. B., (2009) A History of Corporate Social Responsibility, Concepts and Practices, in Crane et. al (eds.), 2009, pp. 19-46.
- Christensen, C. M., (1997) *The innovator's dilemma*, Boston: Harvard Business, School Press.
- Coeckelbergh, M., (2012) Moral responsibility, technology, and experiences of the tragic: From Kierkegaard to offshore engineering. *Science and Engineering Ethics*, 18(1), pp. 35–48.
- Crane A., Mc Williams A., Matten D., Moon J., Siegel D., (eds) (2009) *Handbook of Corporate Social Responsibility*, New York: Oxford University Press.
- Davis, M., (2012) “Ain’t no one here but us social forces”: Constructing the professional responsibility of engineers. *Science and Engineering Ethics*, 18(1), 13-34.
- Doorn, N., (2012) Responsibility ascriptions in technology development and engineering: Three perspectives. *Science and Engineering Ethics*, 18(1), pp. 69-90.
- Doorn, N., van de Poel, I., (2012) Editors’ Overview: Moral Responsibility in Technology and Engineering. *Science and Engineering Ethics* 18 (1), pp. 1–11.
- Ferry, J-M., (1991), *Les puissances de l'expérience*, 2 vol., Paris: Cerf.
- Fischer, M, (1999) Recent Work on Moral Responsibility. *Ethics*, 110(1), pp. 93-140.
- Fogelberg, H., Sanden, B., (2008) Understanding reflexive systems of innovation: An analysis of Swedish nanotechnology discourse and organization. *Technology Analysis & Strategic Management*, 20(1), pp. 65-81.

DRAFT - DO NOT QUOTE

- Goven, J., (2002) Citizens and Deficits: Problematic Paths toward Participatory Technology Assessment, in S. Karner and B. Wieser (eds), *Conference Proceedings of the International Summer Academy on Technology Studies: Technology and the Public*, Graz: Austria, pp. 75-84.
- Grinbaum, A., Groves, C. (2013) What is the 'responsible' in responsible innovation, in Owen, R., Heintz, M. and Bessant, J. (eds) *Responsible Innovation*, Chichester, UK: Wiley.
- Groves, C. (2006) Technological Futures and Non-Reciprocal Responsibility. *The International Journal of the Humanities*, 4 (2), pp. 57-61.
- Groves, C., L. Frater, R. Lee, E. Stokes, (2011) Is There Room at the Bottom for CSR? Corporate Social Responsibility and Nanotechnology in the UK. *Journal of Business Ethics*, 101, pp. 525-552.
- Grunwald, A., (2011) Responsible Innovation: Bringing together Technology Assessment, Applied Ethics, and STS research. *Enterprise and Work Innovation Studies*, 7, pp. 9-31
- Guston D.H. (2006) Responsible Innovation in the Commercialised University, in D.G. Stein (ed.) *Buying in or Selling Out: The Commercialisation of the American Research University*, New Brunswick: Rutgers University Press, pp. 161-174.
- Hacking I., (1986) Culpable Ignorance of interference effects, in MacLean D. (ed.), *Values at Risk*, Totowa, NJ: Rowman & Allanheld, pp. 136-154.
- Hart, H., 1968, (2008) *Punishment and responsibility: Essays in the philosophy of law*, New York: Oxford University Press.
- Hellstrom T. (2003) Systemic innovation and risk: technology assessment and the challenge of responsible innovation. *Technology in Society*, 25, pp. 369-384.
- Jessop, B. (2003) Governance and Metagovernance: On Reflexivity, Requisite Variety, and Requisite Irony, in H. P. Bang (ed.), *Governance, as Social and Political Communication*, Manchester, UK: Manchester University Press, pp. 142-172.
- Johnson, J. M. (1977) *Ethnomethodology and existential sociology*, in J. D. Douglas and J. Johnson (eds.), *Existential sociology*, Cambridge: Cambridge University Press.
- Jonas, H., 1979 (1984), *The Imperative of Responsibility: In Search of Ethics for the Technological Age*, Chicago: University of Chicago Press.
- Kelty C. M. (2009) Beyond Implications and Applications: the Story of 'Safety by Design'. *NanoEthics*, 3(2), pp. 79-96.
- Kermisch, C. (2012) Risk and responsibility: A complex and evolving relationship. *Science and Engineering Ethics*, 18(1), 91-102.
- Kjølborg, K., G.C. Delgado-Ramos, F. Wickson, R. Strand (2008) Models of governance for converging technologies. *Technology Analysis & Strategic Management*, 20(1), pp. 83-97.
- Kuzma, J., Kuzhabekova A., (2011a) Corporate social responsibility for nanotechnology oversight. *Medicine, Health Care and Philosophy* 14, pp. 407-419.
- Kuzma, J., Kuzhabekova A., (2011b) Nanotechnology, voluntary oversight, and corporate social performance: Does company size matter? *Journal of Nanoparticle Research* 13, pp. 1499-1512.
- Ladd, J. (1991) Bhopal: An essay on moral responsibility and civic virtue. *Journal of Social Philosophy*, 22(1), pp. 73-91.

DRAFT - DO NOT QUOTE

- Ladikas, M. (2009) Introduction, in Ladikas, M. (ed.), *Embedding society in science & technology policy: European and Chinese perspectives*. Luxembourg: Office for Official Publications of the European Communities, pp. 7-20.
- Lee, R.G. (2012) Look at Mother Nature on the Run in the 21st Century: Responsibility, Research and Innovation. *Transnational Environmental Law*, 1(1), pp. 105–117.
- Lenoble, J., Maesschalck, M. (2006) *Beyond Neo-institutionalist and Pragmatic Approaches to Governance*. REFGOV, FP6.
- Lenoble, J., Maesschalck, M. (2003) *Toward a Theory of Governance: The Action of Norms*, New York: Kluwer Law International.
- Lhuillier, J.-N. (2007) chapitre II; Innover, c'est faire du nouveau, in S. Boutillier, D. Uzunidis (eds.), *La gouvernance de l'innovation. Marché et organisations*, Paris: L'Harmattan.
- Macnaghten, P., Kearnes, M., Wynne, B. (2005) Nanotechnology, Governance, and Public Deliberation: What Role for the Social Sciences? *Science Communication*, 27 (2), pp. 1-24.
- Maesschalck, M. (2001) *Normes et Contextes*, Hildesheim: Georg Olms Verlag.
- Maesschalck, M., Lenoble, J. (2010) *Democracy, Law and Governance*, Aldershot: Ashgate.
- Margolis J. D., Walsh J. P. (2001) *People and profits? The Search for a Link Between A Company's Social and Financial Performance*, Mahwah, NJ: Lawrence Erlbaum Associates
- Mitcham, C. (2003) Co-Responsibility for Research Integrity. *Science and Engineering Ethics* 9, pp. 273-290.
- National Nanotechnology Initiative (2007) 'National Nanotechnology Initiative Strategic Plan', December 2007, <http://www.nano.gov/sites/default/files/pub_resource/nni_strategic_plan_2007.pdf?q=NNI_Strategic_Plan_2007.pdf>
- National Research Council (2006) 'A matter of size', Triennial Review of the National Nanotechnology Initiative, <http://www.nap.edu/catalog.php?record_id=11752>
- Nordmann A., Macnaghten, P. (2010) Engaging Narratives and the Limits of Lay Ethics: Introduction. *Nanoethics* 4, pp. 133–140.
- Owen R., Baxter D., Maynard T., Depledge M. H. (2009) Beyond regulation: Risk pricing and responsible innovation. *Environmental Science and Technology*, 43(14), pp. 5171– 5175.
- Owen R., Crane M., Deanne K., Handy R.D., Linkov .I, Depledge M.H. (2009) Strategic Approaches for the Management of Environmental Risk Uncertainties Posed by Nanomaterials, in I. Linkov, J. Steevens (ed.), *Nanotechnologies: Risks and Benefits*, Springer, pp. 369-384.
- 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, pp. 751–760.
- Pellizzoni, L. (2004) Responsibility and Environmental Governance. *Environmental Politics*, 13(3), pp. 541-565.
- Porter, M., van der Linden C. (1995) Toward a New Conception of the Environment-Competitiveness Relationship. *Journal of Economic Perspective*, 9, pp. 97-118.
- Reber, B., (2006a), Technology Assessment as Policy Analysis: From Expert Advice to Participatory Approaches, in Fischer F., Miller G., et Sidney M. (eds), *Handbook of*

DRAFT - DO NOT QUOTE

- Public Policy Analysis. Theory, Politics and Methods*, New York, Public Administration and Public Policy Series, Rutgers University/CRC Press, 2006, 125, pp. 493-512.
- Reber, B., Sève R., (ed.) (2006b) *Le Pluralisme. Archives de philosophie du droit*, Tome 49.
- Reber, B. (2010a) *La Bioéthique en débat. Archives de philosophie du droit*, Tome 53.
- Reber, B. (2010b) *La délibération des meilleurs des mondes entre précaution et pluralisme*, Paris: Sorbonne.
- Reber, B. (2011a) *La démocratie génétiquement modifiée. Sociologies éthiques de l'évaluation des technologies controversées*. Québec: Presses de l'Université Laval.
- Reber, B. (2011b) L'éthique ne souffre-t-elle aucun compromis? Le pluralisme éthique permet d'en douter, in Nachi M. (eds.), *Actualité du compromis. La construction politique de la différence*, Coll. Recherches (sociologie), Armand Colin, pp. 64-77.
- Reber, B. (2012a) Argumenter et délibérer entre éthique et politique, in Reber B. (dir.), *Vertus et limites de la démocratie délibérative, Archives de Philosophie*, Tome 74, avril-juin, pp. 289-303.
- Reber, B. (2012b) Coexistence des arguments. Entre éthique et querelles des facultés ?, in P-A. Chardel, C. Gossart, B. Reber(eds.), *Conflits des interprétations dans la société de l'information. Ethique et politique de l'environnement*, Hermes Sciences, pp. 253-279.
- Report of the Expert Group on the State of Art in Europe on Responsible Research and Innovation (2013) *Options for Strengthening Responsible Research and Innovation*, Chair: Jeroen van den Hoven; Members: Linda Nielsen, Françoise Roure, LaimaRudze, Jack Stilgoe; Rapporteur: Klaus Jacob.
- Richardson, H. S., (1999) Institutionally Divided Moral Responsibility, in Paul, E. F., Miller, F. D., Paul, J., (eds.) *Responsibility*. Cambridge University Press, Cambridge, pp 218-249.
- Ricoeur, P., 1995 (2002) *The Just*. Chicago: Chicago University Press.
- Simon, H. (1955) A Behavioral Model of Rational Choice. *Quarterly Journal of Economics*, 69, pp.99-118.
- Simon, H. (1980) From substantive to procedural rationality, in Latsis, S. (ed.), *Method and appraisal in economics*, Cambridge: Cambridge University Press, pp. 129-148.
- Stilgoe, J. (2011) A question of intent. *Nature Climate Change* 1,(7), p. 11.
- Thompson, D. F. (1980) Moral responsibility and public officials. *American Political Science Review*, 74, pp. 905-916.
- Tirole, J. (2001) Corporate Governance. *Econometrica*, 69, pp. 1-35.
- Toulmin, S. (1958), *The Uses of Argument*. Cambridge, Cambridge University Press.
- van den Hoven J. (2013) Value Sensitive Design and Responsible Innovation, in Owen *et al.*, 2013, p. 75-84.
- Van de Poel, I., Vincent N., van den Hoven J. (eds.) (2011) *Moral responsibility, beyond free will & determinism*, Dordrecht, NL: Springer
- Van de Poel, I. (2011) the relation between Forward-looking and Backward looking Responsibility, in Van de Poel, *et al.* (eds.), 2011, pp. 37-52.
- Verbeek P.-P. (2011) *Moralizing Technology*, University of Chicago Press.
- Vincent, N. (2011) A structured taxonomy of Responsible concepts", in I. Van de Poel *et al.* (eds.), 2011, pp. 15-35.
- von Schomberg, R. (2011a) The quest for the "right" impacts of science and technology. An outlook towards a framework for responsible research and innovation, in M.

DRAFT - DO NOT QUOTE

- Dusseldorp, R. Beecroft (eds.) *Technikfolgen Abschätzen Lehren. Bildungspotenziale transdisziplinärer Methoden*, Springer Verlag, pp. 394.
- von Schomberg, R. (2011b) Towards Responsible Research and Innovation in the Information and Communication Technologies and Security Technologies Fields, European Commission, Brussels.
- von Schomberg, R. (2013) A vision of responsible innovation, in R. Owen, M. Heintz and J Bessant (eds.) *Responsible Innovation*. London: John Wiley.
- Vogel D. (2005) *The Market for Virtue: The Potential and Limits of Corporate Social Responsibility*, Washington: Brookings Institution Press.

DRAFT - DO NOT QUOTE

Annex 1

INDECT project: An assessment

Sophie Pellé, Bernard Reber

(Post-doctoral Fellow, University Paris Descartes; Director of Research CNRS)

1. The theoretical background and problem hypothesis of the project

Here, we consider two levels of reflection: first, the theoretical background of the project (its aim and the context in which it takes place (which shows a balance between different ethical requirements in tension) and second, the theoretical background related to the way in which the project deals with responsibility.

a) General background of the project

In brief, the INDECT project aims at creating a new algorithm supporting decision to detect unlawful or criminal activities in a way that seeks to preserve privacy and freedom.

“The primary objective is to develop advanced and innovative algorithms aiming at *human decision support* in combating terrorism, and other criminal activities, such as human trafficking, child pornography, detection of dangerous situations (e.g., robberies) and the use of dangerous objects (e.g., knives, guns) in *public* spaces. Efficient tools for dealing with such situations are crucial for ***ensuring safety of citizens***. A significant part of the project is dedicated to the development of tools and methods for data and ***privacy protection***. The processed information is protected before its transmission or storage to prevent *unauthorized access* attempts. Dedicated tools are developed to protect the privacy of citizens in the areas covered by visual monitoring systems.” D08, p. 6³³.

The project aims at combining two different elements: a) satisfying a social need, that includes some ethical dimension (safety) b) within the limits of an ethically acceptable framework (respect of privacy). On the one side it seeks to improve the fight against terrorism and criminalities by innovating in the field of detection tools (video monitoring, plate’s analysis, web monitoring, etc.). But since innovation implies, here, an intrusion into the

³³ *Italic* and **bold** are from us.

DRAFT - DO NOT QUOTE

privacy of individuals, the challenge would be to limit the level of intrusion to what is considered as just necessary. A practical level, it means preventing unauthorized access to data and avoiding mass detection.

To this end, the INDECT project has to answer the issue of how to qualify and define what dangerousness (but also normality) is, as well as the relationships of these two notions with safety. Two other issues of the project relates with the automation of the inquiries (supposed to be less intrusive in avoiding any human reading of the real data) and the separation between the normative findings of a research project and their implementation or application in real life.

Our synthesis aims at establishing if the project is conclusive on these points.

b) Background on responsibility

The project invokes several tools to deal with the ethical issues raised by the introduction of new and innovative detection tools: mainly, the introduction of an Ethical Board (EB), Deliverables reviews, the production of an *Informed Consent Form* and of methodologies of recording. In spite of the fact that the project tackles an ethical problem (protection of the people), responsibility is not addressed directly, but indirectly through the ethical aspects related to the project. We have conflicting responsibilities targeting different ethical entities: security and privacy. Finally, another tension is related to the restrictions imposed by responsibility in research and their extension to possible applications.

The following analysis will be mainly focused on the way in which ethical issues are dealt. The link between ethical reflection and responsibility will be made in the synthesis.

2. Method used for the analysis - parameters of analysis

To assess the INDECT project, we will have to answer several questions.

2.1. Approach to ethics (tools and use of the tools)

What are the tools put forward in the project to deal with responsibility and ethical issues (ethical committee, participatory approach, forum, etc.)? What use of these tools is being made (what is their purpose: advice the research project, shape the design of the experiments, identify ethical issues (raising question that should be answered), ensure the compliance with existing laws, etc.)?

DRAFT - DO NOT QUOTE

2.2. Role of ethical issues

What were (if any) the consequences of invoking ethical issues for the output of the project? If norms were established, did they lead to modifications of the process, the range, the scope or procedures of the project?

2.3 Ethical issues in relation to the context

Inspired by previous FP7 projects (mainly ETICA, EGAIS and Consider) another parameter of analysis is the relation of ethical norms to context. Is the context ignored (meaning that no ethical issue were raised)? Are ethical issues put forward in a “rote” way to borrow from Nordmann and Macnaghten (2010), meaning that they are only reflected on in an abstract way, without any link to the specific context of the project)? Are ethical issues addressed in relation to the context (but only in a limited form, Restricted Contextualised) or are they fully Contextualised (to use EGAIS’ and ETICA’s terminology)?

2.4. Which sort of presupposition?

Using Lenoble and Maesschalk thesis, which kind of relation between norms and their application is presupposed by the project (Mentalist, Schematising, Intentionalist).

2.5. Revision process - reflexivity

Are judgments and norms proposed under the banner of an ethical enquiry, subject to a process of revision? If yes, how is this process conceived? In other words, is the process of production of norms challenged at any stage? If yes, in which way?

2.6. Which patterns of governance?

Which is the model of governance implemented by the project (top-down determination of norms, co-constructive, etc.)?

3. Results of the analysis

3.1. Approach to ethics.

DRAFT - DO NOT QUOTE

Ethical purpose of the project

First, in the mere definition of the project, some constraints are applied to the design of the research, reflecting a concern for ethical issues such as privacy, individual freedom or the dual use of technology. For instance, the project explicitly mentions intending to avoid “mass surveillance” and seeking to reutilize already available information (existing cameras, Web Pages) without relying on new devices. Moreover, the use of an algorithm as a support for decision is planned to reduce human intervention and thereby “opportunities for illegitimate use of the monitoring records”, should help to protect the stored and transferred information from unauthorized use (DEL 05, p. 8). It thus appears that one of the explicit purposes of the technology developed by INDECT is directly related to the achievement of a norm related with the possible dual use of technology: the project seeks to reduce unwanted and illegitimate use of data in the context of monitoring and controlling by police services.

Main tool: the Ethical Board

However, the main way for the project to address ethical issues relies on the formation of an Ethical Board (EB). The members of this EB are chosen by the coordinator of the Consortium. This way to proceed is not collective or expert driven (for instance, there is only expert in ethics).

Description of the board:

The EB goes along with the project from its beginning and originally gathered (DEL 05):

- Two police officers and one retired police officer
- A human rights lawyer, Professor of law and of human computer interaction.
- Professor specializing in Ethics in Scientific Research
- A technical specialist
- Three researchers in the domain of security related technologies

But, in 2006, it evolved to include “a Professor of ethics, Doctor of Philosophy, and Professor of Human Rights” while one police officer and two other members internal to the project left the Board.” December 2009 the EB has included another Professor of ethics (DEL 07, P. 33). If the composition of the project is evolving during the project to integrate ethical expertise, changing its original conception (that rejected such an expertise), the price to pay is some negative withdraws regarding the continuity of the project.

At the beginning, the expertise in ethics is not pluralist including only one ethicist (but we don’t know if this person had the role to propose some cartographies of the existing positions on issues at stake). It also involved a student, with a consultative position. However, it is hard to understand the idea behind this.. In 2007, members of the project claimed to have

DRAFT - DO NOT QUOTE

“representatives from different areas: academia, ethics, law, data protection, NGO, police, and industry.” (DEL 07, p. 16.)

Role of the board

The general purpose of the EB is to “supervises the ethical aspects of the project’s activities.” (DEL 06, p. 23). The EB is supposed to allow for a “permanent monitoring of proposed solutions in the INDECT project with regard to the ethical issues.” (DEL 05, p. 15) This general task is supposed to be ensured by regular meetings specifically devoted to ethical issues.

More precisely the EB role entails three different elements: a) ensure compliance with existing (national and international) legal constraints, b) providing norms to frame testing on human subjects, and c) raising the relevant ethical issues of the project (which entails and assessment of the ethical content of the project and the bringing about of recommendations).

a) Ensuring compliance with existing EU (and national) rules

“The Ethics Board ensures strict fulfillment of the ethical rules set to deal with privacy, data protection, prevent dual use and guarantee informed consent of users in the project.” (ibid.). This also includes, for instance, a compliance with the Charter of Fundamental Rights. As an important part of the process by which technologies developed through the project are tested involves the use of personal data, the project explicitly limits the personal data processing by human rights and guarantee voluntary acts on “informed consent forms signed by persons whose data is being processed.” (ibid.)

To cope with his task, the EB can report any “improper use of research result” to the Commission, if needed. It is also mentioned that the EB should tailor the solutions being developed to the new legal acts. In term of responsibility these norms are mainly coming from law and are not specific as positive ethical one are.

b) Norms of experimentation within the project

Second the EB has to settle the norms and ethical control procedures regulating the performing of tests with human subjects (Appendix A. INDECT Informed Consent Form DO.7).

c) Ethical assessment

DRAFT - DO NOT QUOTE

A third type of task is assigned to the EB since it “does not view its role as ensuring compliance as a minimalist task, solely designed to ensure legal compliance.” (ibid.). As an “independent body”, it should also involve an “overseeing” of “scientific and societal issues related to the research activities conducted within the project” (ibid.). Managing societal and scientific issues also includes an assessment of “ethics related content of project reports” though a review of the project Deliverables and implies providing with recommendations and proposals “for possible implementation of tools elaborated in INDECT.” (ibid.)

In other word, the EB should bring into light the ethical issues raised by the different elements of the project and offer some recommendations to deal with.

In order to achieve this goal, a special focus has been given to the discussion of ethical issues: for instance a special session “Ethical Issues in Police Operations” was scheduled for MCSS 2010 (Multimedia Communications, Services and Security) including invited guests involved in research on ethics and human rights (DEL 05, p. 9). Also, EB meetings were planned all along the project. If initially scheduled at least once a year, EB members, almost from the beginning, agreed on the necessity to raise the frequency of the meetings. In addition, they sought for collaboration with other projects in going at several international conferences on relevant ethical issues (for instance the European Joint Conference of the HIDE and RISE Projects "Ethics and Governance of Biometrics and Identification Technologies" (Brussels, 9-10 December 2010); and in cooperating with the FP7 project DETECTER (Detection Technologies, Terrorism, Ethics and Human Rights).

3.2. Role of ethical issues

a) Ethical issues raised/ not raised

The first output of EB activities can be identified as the highlighting of a series of ethical issues closely related to the technologies developed within the INDECT project.

Among other, the meeting of the EB identifies several areas of enquiry:

- Ensuring that the research conducted by the project impacts citizens' lives and is connected to public concern (for instance in WP4: which aims at identifying criminals, criminal organizations through the analysis of forums, blogs & social networks) (DEL 07, p. 9).
- Setting internal rules of the board, such as independency of the members (DEL 07, p. 11)
- Reflecting on definitions such as the rules for distinguishing what is criminal from what is abnormal (*ibid.*) or what is unlawful from what is suspicious. (DEL 08, p. 14)
- Reflecting on data protection (DEL 08),

DRAFT - DO NOT QUOTE

- Reflecting on the procedures to involve human subjects in testing
- (For other issues see parts 4 and 5 of our document).

b) Example of precise ethical issues: the case of UAV

But the EB also raised more specific issues, for instance, in WP2, which developed tools for identification and observation of mobile objects in urban environment i.e. Unmanned Aerial Vehicle (UAV), to support operational activities of police officers or other public services.

In DEL 07, the EB addresses various ethical questions raised by this device in an utilitarian way by asking if the potential damages of the device can be counterweighted by its benefits:

“First we have to consider artificial intelligence programmed for the UAV. How much ‘free will’ will it have? It should be able to fly on its own, track suspects, patrol desired areas. The UAV should also know when it is supposed to return to base. But can it disobey a direct order? For instance the operator would like the UAV to follow a vehicle even though the UAV cannot return due to limited energy supply. This will result in a crash and could harm innocent people. What is more important? If it is only a car thief it should return and if there are terrorists in a vehicle it should follow orders? Who should make such distinctions? And what is also important who will be responsible if there is a crash and people are injured or even killed? The operator because it ordered such flight or the creators because they did not support the UAV with order rejection system?” (DEL 07, p. 18.)

In all this, more questions are raised than justifications are given.

c) Some areas do not present any difficulties

In other cases, the EB detected no relevant ethical issue. For instance, while analyzing blogs and web sites, the EB concludes that:

“It does not affect citizen's privacy and human rights, for two important reasons:

1. It operates on plain data such as URLs, keywords, blog nicknames that are processed and analyzed for structure and relationships. There is no personal information processing including creating and storing of personal profiles.
2. It operates on publicly available data from the open sources of information. It may happen that a blog user will have a nickname containing his name and surname. By publicly publishing this nickname he/she agrees and accepts that this information can be retrieved by other Internet users.” (DEL 08, p. 17-18).

DRAFT - DO NOT QUOTE

To sum up, the use of the information is being anonymised in avoiding a direct link between personal data and the information gathered on blogs and websites. Here, the mere design of the procedures avoids raising a prominent ethical question about privacy and the right to protect personal data.

Constraints imposed to the design of the experimentation

Another example of constraint imposed to the use of devices in order to comply with ethical norms on individual privacy is given by the use of video content:

“Sensitive data in video content are anonymized by automatic obscuration of faces and car plate numbers, and by sending original information in encrypted form. Similarly, for the sound stream, only the processing results are presented and sensitive recordings are encrypted. Therefore the operator has a possibility to review the event without violating privacy rights, and only after specific access rights were issued he can replay the original material. » (DEL 06, p. 12.)

A third example lies in the treatment of personal data related to internet based information. The project aimed at distinguishing between different levels of disclosure of personal data, one of them, as we saw previously, being the analysis of the content of blog and web sites which was conceived as not favoring any unwanted identification. But the third subsystem for supporting information identified by the project, the CrimNet Tool, is specifically devoted to “supporting the information management about organized crime gangs [that] may contain the names, surnames, and roles of the inspected gang members.” (DEL 08, p. 18). Here the use of personal data is justified by the criminal activities allegedly attributed to individuals and by the restricted use of these information to internal police network with security mechanisms and management.

Conclusion of 3.2.

These different examples show how trying to achieve an ethical norm of protection of individual data from unnecessary disclosure has helped to shape the process for collecting and diffusing information and the way in which technology is designed and used in the project. However, it has to be mentioned that all the ethical issues that were raised are internal to the project. In INDECT, the EB has to assess the ethical content of the project, but not the ethical relevance of a technology of detection that would be widely and daily applied within the European Union. This is an important limitation of the work. Indeed, the INDECT project could be considered as “responsible” for it meets some ethical norms considered as desirable while the technologies that would be extracted from the project and distributed among the European territory would not.

DRAFT - DO NOT QUOTE

Relation of ethical issues to the context

During the project, the various ethical issues mentioned above (such personal data protection or the procedures needed to involve individual subjects to testing) have been analyzed in close relation with the specificities of the project. In that we could say that this project offers a contextualized relation to norms.

Which sort of presuppositions?

As mentioned above, the norms put forward by the project concern the research project and not a wide use of the detection technologies developed by INDECT. The relevant community of actors affected by the norms is therefore limited to the members of the project (researchers). They are “significantly more sensitive to ethical aspect of their work and tools elaborated in the project” (DEL 6, p. 27). Using Lenoble and Maesschalk thesis, it appears that these norms rests on at least, two of the three presupposition the authors highlight: the EB used set of guidelines already existing in relying on the European Legal Framework regulating the use of personal data (Schematizing). In addition, the set of norms they identify for the project is supposed to be followed by the other researchers (Intentionalist).

Revision process

All along the project, the EB showed a capacity to engage in an adaptive process, in particular, in modifying the composition of its member to open it to philosophy and other social sciences, and in raising the frequency of its meeting.

Examples from the above paragraph show a capacity of the project to adapt the design of the experimentation to norms that are considered by the EB as ethically and/or socially desirable. In this way, the project seems to have allowed members to be responsive to each other. For instance

During the meeting of February 2011 (Belfast), “it was agreed that the Leader of Work Package 4 would summarize the research, providing the Ethics Board an opportunity to highlight the associated ethical concerns. This would enable the Work Package Leader to respond as to how the research could take account of such issues.” (Del 07, p. 9).

Finally, INDECT project was submitted to an external ethical review (for instance in Brussels, in March 2011), whose conclusions admitted that “- there is an appropriate supervisory/monitoring structure for the management of the ethical issues. The final rating in the Overall Assessment of INDECT done by the Review Panel is as follows: "Acceptable compliance with the FP7 ethical guidelines provided that the identified Requirements are comprehensively addressed." DEL 06, P. 27.

DRAFT - DO NOT QUOTE

A question to ask here is whether the recommendations provided by this external evaluation of INDECT has been followed by the project or not?

However, several questions remain. First, no mention is being made on how the ethical issues are raised. They come from the input of the different members of the EB and show a close relation with the specificities of INDECT project. However the process by which these ethical issues are put forward is not challenged. Also not challenged is the way in which EB members provide researchers with recommendation and ethical guidelines. This could have been done in a more collective approach (involving the different members of WP). But it is not the case, as the EB is the only institution in charge of enouncing ethical norms. The way in which they define what is socially and ethically acceptable is not challenged.

Which patterns of governance?

The project is based on a top-down conception of governance with an independent committee, setting the rules for other users (here, researchers of the project) without involving them. If the consequences of such a governance approach might not be as problematic as in the case of widely distributed technology (that affect much more actors and individuals), the construction of norm does not proceed from a co-construction. However, the interdisciplinary composition of the EB and its ability to engage in an adaptive process of revision (of its composition) should be outlined.

Thus, one main limitation of the project is that the final users of the type of technology that is been developed in the research project (i.e. European citizens) are not involved at all, at any stage of the project.

4. Synthesis with some interpretation of those results

The management and the role of the coordinator mainly cover the ethical issues. We are far from a reflexive inter-actor governance of inter-institutional governance. The level of deliberation is flat (or poor). We have rarely normative justifications, inference between reasons and decisions or productions of arguments. In this respect the project governance is pre-deliberative. We have to move from full collective responsibility to a reflexive (at a normative level) conception of responsibility.

In an inductive move, we could say that the elements or characteristics of responsible research and innovation that can be identified from this study include.

- Responsiveness: or the propensity of a project to be shaped by norms.

DRAFT - DO NOT QUOTE

- A collective setting up of norms which goes from including inter-disciplinarity in the ethical assessment of the project (as in INDECT) to a real inclusive enquiry (implying a comprehensive participation of stakeholders).

These elements will have to be confronted with richer definition of responsible innovation governance (presented in the theoretical landscape (del 2.2.) forthcoming) including a production of norms that question the link between norms and their application (and focus on the process of participation of stakeholders and the role of deliberation).

In conclusion, INDECT project appears to make a first move towards responsible research in allowing for responsiveness and interdisciplinarity. But it rests on a top down model of governance that supposes guidelines to be decided by expert and then being applied by members of the project. Moreover, there is no reflexivity concerning the process in which the ethical norms are obtained and no involvement of potentially relevant stakeholders for a future wide implementation of the technology.

5. Recommendation for the project

- The issues addressed in our General Background (GB) are not fully taken in charge. Because the EB only investigates the issues raised by new detection tools in the context of research (neglecting the practical application of these tool)s, the trade-off and the possible balance between security and privacy we mentioned in the GB is not specified.
- There is only very few attempts of forecast and foresight (one important pillar of RRI, as we have seen).
- There are some contradiction and confusions in the problems, as, for example when detection in *public* spaces and uses of Internet data that are *private* are considered together (D08, p. 6).
- This fuzziness is correlated with the combination of different contexts, technically linked, without them being clearly specified. Subsumed under the broad distinction between virtual and real context we have more contexts. They configure differently the interpretations and implementations of norms.
- The project implicitly carries a very limited conception of innovation, from which applications have been excluded. Partly aware of this limitation, the project asks for more time to be in the capacity to prevent all the undesirable side effects of the innovations it proposes.
- A first important issue in case of large implementation will be the difficulty to have the unlighted consent of all the European population (see the project regulation of the UE on this issue, 25/01/2012, COM (2012) 11 final). The choice could follow 2

DRAFT - DO NOT QUOTE

different lines: a) security purposes and exceptionality (police and relevant authorities); b) RRI, with inclusiveness and accountability. The transparency requirement reaches here an interesting limitation. As it is transparent, it loses its efficacy towards the potential criminals.

- The project contains no ethical justification about the fact that automatic surveillance is better than human one.
- To deal with these issues, a more participatory approach should be chosen (implying EU citizens, but also the project researchers), to avoid the traditional “division of labor” between “scientists” and “ethicists”. More than that, the project should give voice to all the stakeholders involved, pro and contra (i.e. *Anonymous*, La quadrature du Net) and to different institutions (i.e. European Parliament who entered the controversy and reached divergent conclusions on this project). The inclusion of these types of stakeholders and of their arguments will make a more pluralist assessment, allowing for proportionality between conflicting responsibilities, ethical targets (security, privacy) and ethical foundations to assess them (see pluralism in the GREAT Glossary). It will make possible to move from responsibility as personal participation to intersubjective or full collective responsibility.
- The IDECT project should offer a deeper reflection on the production of norms (purposes or ethical goals) that are conceived as ethically desirable. If the project wants to further the question of what is socially and politically desirable, it will have to deal with (and respect) the different ontologies and epistemologies of the political, social and ethical spheres.
- The introduction of normality (“abnormal behavior”, FAQ Q1.3) by the EC makes the focus of the project more problematic. The qualification of INDECT project as “criminal behavior...related to terrorist acts” is more focused.

6. Considerations for GREAT

This Project opens four important issues to add to our General problematic (Paris Theoretical Workshop). 1. To be responsive towards an ethical problem (security) could be conflicting with other ethical entities (privacy protection). 2. It makes more complex our three folds positive definition of responsibility (Paris General Problematization), because of different responsibilities involved. 3. The need to arbitrate conflicted responsibilities, carried by different responsible actors (role responsibility). 4. The strict division between research responsibility and innovation responsibility (implementation at large) is highly problematic and could be conflicting (i.e. How will it be possible to assess the efficiency of the tools regarding security or a kind of preemptive action without any test? How will it be possible to do that with the consent requirement?)

DRAFT - DO NOT QUOTE

DRAFT - DO NOT QUOTE

Annex 2

Paris Workshop

Methodological Meeting

(22-24.04.2013)

Organized by the Research Center, Meaning, Ethics, Society

(CNRS-University Paris Descartes)

Bernard Reber, Sophie Pellé

<http://cereses.shs.univ-paris5.fr/?lang=en>



Governance of REsponsible innovATion

Organisational issues

Exact timing:

Monday 22.04.2013 - 12.00 to Wednesday 24.04.2013 - 13.00

Venue: University Paris Descartes

PRES Sorbonne Paris Cité

45, rue des saints-pères

75006, Paris, France

<http://www.sorbonne-paris-cite.fr/index.php/en>

General problematization

DRAFT - DO NOT QUOTE

GREAT project aim is neither to find a common definition of RRI to settle interpretative quarrels, nor to make an heterogeneous collections of the stakeholders perceptions on it. It is not even to accumulate all the existing (sometimes conflicting) key responsible activities that could be cover by a kind of meta-responsibility. If all can be useful, our ambition is bigger. We have to take into account three sorts of representations (Aristotle *mimesis*): a) what the things are, b) what people say they are and c) what they have to be. Because responsibility is a strong normative concept, it would have been not enough to depict existing practices (a) or to interview appropriate actors (b). Indeed, cumulative approach will let the users in front of different approaches and conception without any criteria to assess them. Moreover, the so-called axiological neutrality is useless, as moral sociologists have shown this, avoiding on the one side poor descriptivism and, on the other one, massive normative decontextualized judgement.

We have to pass from the analysis and understanding of moral (mores) to the one of ethics focused on responsibility. With a stance that focuses on the question of normativity connected to responsibility. We should be in the capacity to analyse the ways RRI, not only as a norm but with its normativity (**reflexive stance in the condition of norm construction**) is understood and implemented by different actors in their **contexts**, to be effective. This dynamic is an on-going process of adjustment between normative horizon context and between norms and values.

Thus it is too limited and arbitrary to select one definition, trying to impose it, especially in the fluctuant domains of research and innovation in its tension/complementarity with responsibility. Therefore our method will be a procedural-comprehensive (reflexive equilibrium) one, context-adaptive and normative-sensitive regarding the agents embedded. It analytically explores the possible choices to deal with the problem, with a back and forth between empirically informed and theoretical research.

In GREAT project, the question of RRI is closely connected with governance, not reduced to regulation nor democratic rules, because they are not specific enough for responsibility (ethically understood). Different forms of governance exist (delegated, educational, using participation as validation, co-responsible), implemented thanks to different institutional tools. According to us these institutional arrangements, through different kind of devices (i.e. Participatory Technological Assessment, ethical committee, forum, observatory) should be component appropriate for new institutional conditions that allow responsible reflexivity in complex context processes. Governance is focused on the framing of the context, the normative horizon used by the actors to understand their situation and the RRI normativity within it. This horizon is ethically pluralist and not only because of a fragmentation of social authorities in modern societies and the heterogeneity of normative sources or comprehensive doctrines. For this reasons, we have to consider **different levels of contexts**: a) real context (too rich to be depicted) of

DRAFT - DO NOT QUOTE

the actors, b) the conditions of its framing (reflexivity on different ways to frame it), with an intertwining of the epistemic and the normative horizon conditions. Practically the actors find the normative resources via dialectic between their rich real context and an ethical one (counterfactual) that could help them to reframe their understanding of their context and their action to change it. If the context is limited to the practical constraints, it would be inefficient to speak of responsibility or of ethics, or it will stay on a discursive level. Beyond that, different governance tools pass through different de-contextualisations and recontextualisation (i.e. citizen conferences convening experts and citizen). That makes a plea for of a theoretical approach seeking a generality ascent tacking into account the epistemic and normative pluralism of different referential spheres (science, economy, law, ethics, personal values...).

The move of responsibility, with the notion of RRI, changes the configuration of the couple innovators - that creates problems (directly or indirectly, potential and sometimes real) - vs. opponents in society (from Civil Society Organisations or without any affiliation), laying on different conceptions of responsibility. This couple takes part in a quarrel of improvements (technological and ethical).

Therefore responsibility becomes a positive concept (with three sides). "My" actions make "me" epistemologically and ethically responsible from the very beginning in front of the others. "I" *participate* (commitment) when "I" act. *Intersubjectivity* and *reflexivity* let me discover the weight and load of potential consequences of my actions. This is stronger in innovation and research process. Here, the concept of **full collective deliberation** is perhaps a promising hypothesis in its individualistic, intersubjective and systemic (deliberative inter institutional system) dimensions matched with a co-dependant epistemic and normative evaluation on possible futures (forecast and quarrels on possibilities).

If RRI requires participation of stakeholders, this one is not flat (simple) but qualified (effective and specified). **Therefore, our research question will be: what are the conditions of reflexivity while considering responsibility in innovation to be effective?** Implying that we have to analyse the different patterns of governance and what they offer in term of social reflexive outputs. Indeed, participative tools or systems organize in various ways *accountability and responsiveness*, opening up differently *responsible* agents (role, capacity). And they should try to reach a relevant responsibilities sharing, to avoid dilution and poor involvements and contributions. **Taking seriously governance with reflexivity in context permits - as we aim - to depart from governance of RRI to responsible governance of RRI. This move from sciences for society to science within society implicitly plays with the different meanings of responsibility: Responsible actors in responsible governance system.**

This workshop is a third step after of a) the kick-off meeting (Brussels February 4th-6th, 2013), that has generated the awareness of the complexity of the problems related the RRI efficiency in contexts, and b) the different consortium partners problematizations to explicit the research object of the GREAT project. This has helped to shape this

DRAFT - DO NOT QUOTE

workshop and to select the external experts able to reflect on different sides of this argument and not on the limited question: What is RRI?

Agenda

1) Monday 22.04.2013

ROOM J231 (building Jacob)

Day one: WP2 clarification and problematisation discussion

no	time	Item	Explanation	Responsible
0001	12.30-13.30	Lunch		UPD
Chairman: Sophie Pellé				
0002	13.30-13.45	Welcoming introduction (Sophie Pellé)	Quick presentation of the meeting	UPD
0003	13.45-14.45	Feedbacks on the problematisation	Discussion on the synthesis proposed by UPD	UPD
	15.45-16.15	Coffee break		UPD
0004	16.15-17.45	Links between the analytical grid and empirical work	20 minutes presentation and 1.10 discussion	UOXF

2) Tuesday 23.04.2013

Day two: WORKSHOP: RESPONSIBILITY AND GOVERNANCE OF INNOVATION

ROOM: Lavoisier B (main building)

no	time	Lecturer/item	Title	responsible
	Tuesday 08.45-9.00	Welcome		
0006	09.00-09.30	Introduction and general greetings (Sophie Pellé)	Purpose of the workshop and research question	UPD
Chairman: Bernard Reber Responsibility and Innovation				
0006a	09.30-10.10	Julien Jacob (Assistant Professor, University of	"Economic analysis of liability, or liability as an incentive	guest

DRAFT - DO NOT QUOTE

		Lorraine)	policy tool: Application to risk regulation and innovation fostering"	
0006b	10.10-10.50	Armin Grunwald (head of Institute for Technology Assessment and Systems Analysis and head of the office of Technology Assessment at the German Bundestag)	"The EEE-concept of responsibility – ethical, empirical and epistemological constituents"	guest
	10.50-11.05	Coffee break		UPD
0006c	11.05-11.15	Introduction to discussion (1) – Sophie Pellé	Links between invited speaker's thesis and GREAT project	UPD
0006d	11.15-11.25	Introduction to discussion (2) – Robert Gianni	Links between invited speaker's thesis and GREAT project	FUNDP
	11.25-12.30	General discussion		all
	12.30-14.30	Lunch	Restaurant universitaire Mabillon, 3 rue Mabillon, 750006, Paris	UPD
Chairman: Philippe Goujon				
0007a	14.30-15.10	Klaus Jacob (head of Environmental Policy Research Centre, FU, Berlin)	"Governance of Responsibility of Research and Innovation"	guest
0007b	15.10-15.50	General discussion		
	15.50-16.05	Coffee Break		UPD
0007c	16.05-16.45	Ibo van de Poel (Associate Professor in ethics and technology, Delft University of Technology)	"Responsibility problems in the governance of responsible innovation"	guest
0007d	16.45-17.25	Jack Stilgoe (Senior Lecturer, Science and	"Frameworks for responsible innovation"	guest

DRAFT - DO NOT QUOTE

		Technology Studies, UCL)		
0007e	17.25-17.35	Introduction to discussion (1) – Robert Gianni	Links between invited speaker's thesis and GREAT project	UPD
0007f	17.35-17.45	Introduction to discussion (2) – Sophie Pellé	Links between invited speaker's thesis and GREAT project	FUNDP
0007e	17.45-18.45	General discussion		all