

## 7

# What Is “Responsible” about Responsible Innovation? Understanding the Ethical Issues

Alexei Grinbaum<sup>1</sup> and Christopher Groves<sup>2</sup>

<sup>1</sup>CEA-Saclay/LARSIM, 91191 Gif-sur-Yvette Cedex, France

<sup>2</sup>Cardiff University, ESRC Centre for the Economic and Social Aspects of Genomics (Cesagen), 10 Museum Place, Cardiff, CF10 3BG, UK

## 7.1 Introduction

“Innovation” is the process of bringing something new into the world, through a combination of intellectual and practical ingenuity. In this chapter, we lay out a set of concepts that can help us to understand the meaning of “responsible innovation,” by reflecting on the ethical significance of technological innovation.

What are the key characteristics of innovation in the contemporary world? Technological or otherwise, it typically emerges from the efforts of a community of individuals working together within a complex set of social relationships, through which financial, administrative, and other resources are secured and allocated. What ~~emerges from innovation~~ is not simply a new set of techniques for doing things, but also ultimately new social practices and even institutions that transform the ways in which human beings interact with the world around them. It is, therefore, inherently a *future-creating* activity: by bringing something new into the world, it changes the world itself – perhaps incrementally, perhaps more radically. This means that, to reflect on the conceptual basis for responsible innovation, we have to acknowledge that the responsibility associated with innovation is ~~itself a~~ *responsibility for the future it helps to create*. Consequently, the “responsibility” in responsible innovation presents conceptual and practical difficulties that are related to the future-oriented character of innovation itself.

120 *Responsible Innovation*

In this chapter, we proceed via three steps. First, we clear the ground by setting out why some of the ways in which modern societies have become used to thinking about responsibility are inadequate in relation to **responsible** technological innovation. This section ends by noting that, while in one sense the nature of advanced technologies brings with it new kinds of uncertainty and moral demands, what makes us, as innovators, responsible for innovation is not in itself new, but is rather a condition of mutual vulnerability that we share because of the kind of creatures we are – animals with the capacity to manage their communal affairs through the use of reason; Aristotle’s “political animals.” This leads us into our second step, where we examine how the meaning of responsible innovation might be better understood by analogy with two existing, future-oriented varieties of responsibility, namely *individual parental* and *collective political* responsibility. Finally, we affirm that responsible innovation cannot simply be a matter of following a set of fixed professional rules. We examine the individual and collective “virtues” necessary to exercise the quasi-parental and political forms of responsibility contained in responsible innovation, along with some of the ways in which cultural narratives can help us understand both the complexities involved in moral choice about technologies, and the limits of our capacity to take responsibility.

## 7.2 The Changing Meaning of Responsibility

### 7.2.1 From the Divine Corporation to the Sovereign Individual

On one reading, the whole intellectual and cultural history of the West represents a gradual shift away from what J. B. Schneewind (1984) calls the “divine corporation” view of the world. In this worldview, the basis of morality is a system of offices to which individuals are allotted by divine (or royal, monarchs being **divinely appointed**) edict. Being moral and acting rightly is the same as playing one’s role in this system properly. Whether what one does is right or wrong depends on how one’s actions conform to immutable decrees relevant to one’s position in the hierarchy: the criteria for right and wrong here are *deontological* in nature, that is, related only to a pre-existing rule (Davis, 2001). It is during this period, from Greek antiquity to scholasticism (McKeon, 1990), that we find the origins of two key elements of an understanding of responsibility: *imputability* and *accountability*.

To be responsible for a transgression is, in this tradition, to (i) be recognized as having performed a transgressive act and (ii) be required to give an account of oneself to an authority (or its representative) whose decree one has transgressed against. Responsibility therefore implies being accused of something, that is, being *held* responsible after the fact – a common-sense conception of responsibility called by the British philosopher F. H. Bradley the “vulgar” sense of the word (Bradley, 1927, pp. 1–8). This denotes the fact of authorship of an act: to hold someone responsible for having done something, it must be possible to *impute* an act to him/her after it has occurred: “she was driving when she hit him!” The legal scholar H. L. A. Hart, in his analysis of the different senses of responsibility, splits this *ex post facto* mode of responsibility into causal responsibility (“she caused the collision!”), and blame or liability – which attract punishment (Hart, 1968, Chapter 9, Part 1). **Being** held responsible also means someone having to *account*

*What Is "Responsible" about Responsible Innovation? Understanding the Ethical Issues* 121

for what transgressions against the law are imputed to him or her. This is preserved in the etymology of the various terms that, since the fourteenth century, have emerged in Western European languages (responsibility, *responsabilité*, *Verantwortlichkeit*) as means of translating the Latin legal terms for accountability derived from *respondere*, "to respond," as in a court of law to one's accuser.

As we shall see, historical developments in more recent times have undermined the "divine corporation" worldview, bringing with them changes in how and why people are held accountable, by whom or what they are held accountable, and in what wrongs can legitimately be imputed to them. With these changes comes, in turn, an overall transformation in the meaning of "being responsible." Perhaps the most important element of this transformation affects what we might call the *tense* of responsibility. Under the "divine corporation" view of morality and law, responsibility tends to be a "backward-facing" condition. It exists because of what one has done, and how one's actions relate to the duties decreed by the divine ~~or royal~~ "CEO" of the corporation. Under a new emerging dispensation, from early modernity onward, the temporal focus of responsibility increasingly falls on a secular future (even if the ultimate significance of this future is still shaped by religious belief). Rather than duties being pre-ordained, it is up to the individual moral subject to *take* responsibility for deciding what s/he should do, and to prepare to be accountable later for the consequences. As we shall see, it is this individualist and *consequentialist* concept of forward-looking responsibility – which we accept as being of everyday as well as of legal validity – that initially appears relevant to the idea of responsible innovation. We shall also discover, however, that problems arise when we try to apply it in this context.

The "modern period" (for our purposes here, since the seventeenth century) is one characterized by emerging social complexity. Mercantile societies, as they gradually became industrialized ones, developed highly differentiated divisions of labor, in which individuals typically came to occupy several private and public roles – in direct contrast to the "offices" of the divine corporation. In his classification of varieties of responsibility, Hart describes a prospective, future-oriented meaning of the word that is relevant to such societies. *Role responsibility* is a sociological concept that implies the duties one needs to accomplish by dint of occupying a certain role in society. This is, however, not sufficient to spell out the *ethical* meaning of responsibility. One telling indication that this is so comes from Hannah Arendt's analysis of Adolf Eichmann's role as an administrator in the Nazi hierarchy, in which the gap between role responsibility and a wider responsibility to consider the moral significance of one's actions was egregiously evident (Arendt, 1963). This and other exemplary cases of the gap between one's social role and its ethical content notwithstanding, in modern societies one is often expected to judge one's own actions based on shared understandings of what the responsible "occupational" behavior appropriate to the role of doctor, mother, engineer, politician, and so on, should be (Hart, 1968, Chapter 9). One uses these ideals of responsible conduct for guidance, affirming in the process social and moral boundaries. At the same time, ideals of conduct do not tell us exactly what to do. The lack of fit between ideal and situation creates uncertainty and a need to exercise individual judgment. That role responsibility is unthinkable without some uncertainty or indeterminacy is most clearly reflected in the idea that some occupations are not just roles one *plays* but professions to which one *belongs*, and within which one is expected to exercise professional judgment,

judgment which cannot ultimately be reduced to firm rules. Role responsibility implies specific responsibilities for which one is accountable to particular social institutions, such as professional bodies and the like.

Although the meaning of responsibility in the modern period is heavily influenced by the emergence of role responsibility and the need for individual judgment that accompanies it, the indeterminacy of responsibility here ultimately goes much deeper, as the Eichmann example and others like it suggest. Increasingly, the rightness and wrongness of acts is decided according to individual or collective purposes that are subject to change, innovation, and improvisation – processes often shaped by the application of scientific knowledge. As well as being an age of proliferating roles, this is the age of Max Weber's *Zweckrationalität* and consequentialism, that is, justifying actions in terms of what they can accomplish, rather than in terms of whether or not they conform to pre-existing decrees of right and wrong. A world that can be improved, a world of technology and innovation, is essentially a world in which right and wrong *depend on outcomes*. The individuals who inhabit this world have to balance the competing demands of their roles as parents or children, workers or managers, users or developers of technology, and citizens. They are expected to engage in internal dialog with themselves over the outcomes of the diverse activities they perform (Davis, 2001, pp. 10–11). Responsibility in this sense increasingly becomes a forward-pointing quality or virtue that one's actions can possess or lack. Insofar as it requires that moral agents anticipate the consequences of their actions, taking responsibility means exercising foresight and increasing one's knowledge about the world and how one's actions might interact with and alter it.

The ideal consequentialist moral agent is, therefore, Jeremy Bentham's utilitarian subject, who strives to assess, with precision, the contribution each action makes to the aggregate happiness of society. Without the assurance of occupying one of the fixed "offices" (with their set duties) within the divine corporation, the individual must, as Henry S. Richardson points out (Richardson, 1999, p. 231), judge what to do solely on the basis of the social benefits an act is expected to bring about, over and above the requirements of any particular role she plays. Indeed, this ultimate criterion is meant to help decide any conflicts between the duties associated with different roles. Consequently, the individual is directly accountable to society itself, rather than to some authority outside it, whether deity or monarch.

As Richardson points out, however, this age of innovation and consequentialism is also, by its very nature, an age of surprises. With the focus on consequences, whether or not one is actually *being* responsible becomes subject to increasing uncertainty, particularly because changing situations may require a moral agent to revise or even ignore pre-existing rules and expectations.

Not only does the individual have to juggle the demands of the roles s/he inhabits, s/he also has to question the limits of these roles, and where it is appropriate to go beyond them. Moreover, when it is consequences that matter, the individual may be accountable, despite not possessing any intention to harm; it might be the case that unpredictable future events will have a retroactive impact on the way his or her conduct is judged (Grinbaum and Dupuy, 2004). This phenomenon, referred to as "moral luck" (Williams, 1981), means that the perspective of an ideal utilitarian agent is ultimately impossible to adopt, as to overcome uncertainty would require perfect foresight covering the entire span of consequences of any act. Consequently, the individual is responsible for

deciding when enough foresight is enough, and when s/he must simply stop deliberating, and make up his/her mind to act. Sometimes, as in situations that are relatively routine, this is undemanding. With respect to other activities, however, the opposite is true – and technological innovation is one such.

### 7.2.2 Knowledge, Uncertainty, and Human Finitude

The growing reliance of societies in the modern period on the extensive use of ever more complex technologies changes the nature of uncertainty faced by the individual who takes responsibility for his/her actions. With the growing ubiquity of complex technologies and the global and intergenerational (Jonas, 1984, p. 40) reach of their effects come significant social changes that expose a tension at the heart of consequentialist approaches to responsibility – a tension between the need to *take* responsibility and the evidentiary standards by which moral agents are *held* accountable for the consequences of what they do. The future on which one has an eye when one strives to act responsibly is no longer the same kind of future as in previous historical periods, even in the early part of modernity (Adam and Groves, 2007). The future faced by humans has always been uncertain, but the quality of this uncertainty now takes on a different character.

Let us examine more closely what is involved here. As we have noted, the modern period is one in which innovation – in the broad sense of transforming the natural and social worlds for purposes which might, in fact, themselves be novel – increasingly comes to depend on scientific knowledge. ~~The consequences and broad benefits of action become of primary social concern.~~ Science transforms, too, how people are held responsible for the consequences of what they have done. This is clear if we consider legal evidence. Before a court, what comes to count in general as a true version of events is a scientifically-validated causal story. Liability in a court of law is settled with reference to scientific standards of evidence (“CCTV cameras prove that she was driving, and breathalyzer tests confirm she had twice the legal limit of alcohol in her blood”). But to be liable for one’s behavior requires that one had to be aware that this sort of chain of events could have happened, and that acting in that way would have been wrong. One needs to have had the *capacity* to have foreseen the consequences of one’s actions (Hart, 1968) – “she should have known that another glass of wine would put her over the limit, yet still she chose to drive.” Generally, morally, and legally speaking, “capacity” here refers to a very general ability to act in accordance with moral rules, which young children and people with severe learning difficulties (for example) may not possess.

But in certain circumstances, *anyone* might be judged to be not culpable for the effects of their actions, due to their ignorance at the time of acting. Now, not being culpable due to ignorance, and not having the capacity to be morally or legally responsible are not the same. Nonetheless, there is a sense in which the nature of innovation as an activity creates ignorance in ways that *undermine* the capacity to act responsibly, as a consequentialist might understand it.

It is important to understand that this effect is not simply a matter of innovation adding new “facts” to the world, of which we just happen to be ignorant. As noted above, there is a difference between lacking “the facts” and lacking the *capacity* to understand the facts, a difference which tracks the distinction between law and morality. Being inescapably ignorant of the facts ~~can have legal consequences;~~ for example, not being in



possession of the full facts regarding the potential harmful outcomes of his/her actions can, on the legal principle of reasonable foreseeability, absolve someone of liability even if s/he is found to be causally responsible for a wrong in due course (Pellizzoni, 2004). This is particularly so if ignorance about these outcomes is of a general kind, that is, no-one could have predicted the risk of causing a particular harm, given the then-current “state of the art” in a relevant body of scientific knowledge. In such circumstances, an innovator may have been a model of probity, determinedly living up to the highest demands of their occupational role and exercising due diligence and careful foresight as to potential harmful consequences. Yet because robust scientific knowledge provides the “gold standard” evidence base for due diligence, limitations on the “state of the art” also place constraints on future liability. Ignorance of “the facts,” in such cases, may be *exculpatory* ignorance, in a legal sense. Legal systems may, therefore, face significant difficulties in principle in dealing with new technologies: this is the legal face of what Ulrich Beck has called “organized irresponsibility” (Beck, 1999, p. 149).

~~On this level, the problem is clearly one of knowledge.~~ Scientific knowledge gives us an evidentiary standard suited to a consequentialist concept of responsibility. Scientific evidence is how we assess the quality of someone’s foresight after the fact, and we expect people taking responsibility (particularly in certain socially-important roles) to use such evidence as the basis for exercising foresight. The problem is that science is often badly suited to understanding the consequences of the actions it enables us to perform, particularly when they introduce novel entities (like nuclear reactors, genetically-modified organisms (GMOs), and so on) to the world, because important aspects of how these function in the world may not be covered by the state of the art. By assembling data and creating explanatory hypotheses about how it works, scientific understanding tracks the world the past has made. But creative action and innovation *point forward*, opening up the world the past has created and adding new entities to it that change the way it works. The world that scientific understanding extrapolates into the future – with its gaze still directed toward yesterday – might tomorrow no longer exist.

This problem of knowledge, which makes law lag behind the reality of innovation, opens up a deeper and properly moral problem of capacity, which has to do with human finitude rather than scientific ignorance, though it is related to it. In societies that have become thoroughly dependent on advanced technologies, the background of human action has changed. No longer coherently imaginable as Nature, eternal, and unchanging, it has become an amalgam of natural processes and technological artifacts that intimately interact with them. In Alfred Nordmann’s phrase (Nordmann, 2005), advanced technologies thus become *naturalized* technologies. The more complex our technologies become, and the more embedded in the social and natural worlds, the more like unfathomable “black boxes” they become – perhaps even to their creators. The more intimately they interact with nature – and with each other – the more “autonomy” they have in creating unpredictable effects. Being ignorant of the future impact of one’s innovations may, therefore, become the norm rather than the exception. As Ian Hacking (Hacking, 1986) notes, new technologies may interact with other bits of the world (including other technologies) to create unpredictable negative outcomes, “interference effects.” What may have been well understood in the lab or on the testbed alters the world (and is perhaps altered itself) in unforeseeable ways when released “into the wild.” We have come to rely on scientific knowledge to create the innovations that help us to transform the world,

*What Is "Responsible" about Responsible Innovation? Understanding the Ethical Issues* 125

but we cannot expect it to also enable us to calculate the ethically relevant consequences of using it, even though that is exactly what consequentialist morality requires of us.

As a result, the legal problem of exculpatory ignorance becomes a moral problem linked to human finitude. Humans are finite beings in several ways: because of limitations on their knowledge and of the boundedness of their lives in space and time; because of the diversity of conflicting values that guide their lives (truth, beauty, pleasure, justice, solidarity, etc.); and because the meaning of their existence is *conditioned*. On the one hand, they are born into a social and natural world they have not themselves made, and on the other, whatever they create or shape by acting on the world itself becomes then an objective condition of their lives and to some degree escapes their control over it (Arendt, 1998/1958, pp. 9–10). This multifaceted sense of finitude is challenged by the promise of technological mastery that comes with scientific knowledge. The prospect of unending progress and infinite creative possibility opened up by scientific investigation creates a sense of separation from Nature that is different from the sense of being “thrown” into an already existing world, because it rests upon the promise that human beings have the power to transform this world in accordance with their desires. But the experience of living with naturalized technologies brings into the foreground again the *conditioned* character of human life. Living with technologies offers up new forms of uncertainty and insecurity, making us *more* human (in the sense of contributing to the evolution of ~~the human condition~~) rather than superhuman.

If scientific knowledge ~~does not~~, therefore, ultimately give us the guide for action that consequentialism requires, then the resulting lack of moral capacity is ~~not~~ a deficiency that is ~~attributable to any particular human being. Rather it is a flaw inherent within~~ the culture of consequentialism itself. This is underlined by the inadequacy, in relation to innovation, of certain tools generally employed by consequentialist moral philosophers to resolve problematic situations where a good intention might produce bad outcomes (~~such as~~ Hacking’s negative interference effects). For example, the “principle of double effect” states that an intention to create beneficial outcomes (e.g., via technological innovation) may be judged permissible, even if it may also produce bad effects so long as (i) these effects are not intended, (ii) the desired outcome does not require these bad effects to happen, and (iii) that the intended good effect should not be grossly outweighed by expected bad effects (Timmons, 2002, pp. 78–80). If technological innovation fills this bill, then we can consider ourselves to have taken a responsible choice in pursuing it. Yet we face here, once again, the problem of foresight. Given that interference effects are highly unpredictable, and may have uncertain consequences that vary enormously in seriousness, the principle of double effect, with its criterion of proportionality, is not applicable (Yara, 2011, pp. 154–155).

If the experience of naturalized technologies reveals to us the inadequacy of one concept of responsible, forward-looking action, however, it also drives home that one can still be *held* morally responsible by society for what happens, even if one is ignorant (through no fault of one’s own) of the *specific* potential harms at the time of acting. Hacking points out that, even if we cannot reliably predict the consequences of new innovations, we *can* know that new technologies by their very nature tend to produce “interference effects.” Further, we can expect that some unforeseen side-effects of technologies may turn out to be positive, but that many will likely be negative. Given that, in technological societies, a huge variety of complex technologies are ubiquitous, and

may be strongly interdependent with natural systems, such interference effects must be expected to be relatively common, and will range from relatively innocent to extremely serious in nature. As a result, one knows one is taking what may be a serious risk, and that moral luck is, therefore, a factor in how one will be judged (Nagel, 1979). Even if such risks may, legally speaking, be legitimated by provisions such as the development risk defense (Newdick, 1988; and Lee and Petts in Chapter 8), there remains a moral question here that bears on legal matters.

If we recognize that we may be held morally responsible for the results of innovation, we still need a way of thinking about how to actively *take* responsibility that escapes the problems associated with consequentialism that we have described. This is by no means easy. Given the “autonomy” of naturalized technologies, we can expect that any interference effects may be extended in space, be latent for long periods of time (perhaps even lifetimes), and may be cumulative, perhaps persistent, and possibly irreversible. The release of a technology into the world, through the journey from research, through innovation, to market, becomes itself a kind of experiment, with the world as the laboratory (Krohn and Weyer, 1994). The spatial scales at which interactions between technologies, and between technologies and nature, occur may extend to encompass the entire globe (as evidenced by the traces of persistent organic pollutants and other chemicals found in species from the Poles to the Equator). Further, the timescape (Adam, 1998) of some of these interactions – the reach of their intended and unintended effects into the future – may be extremely long, or even, in principle, unending, as in the case of genetic engineering’s irreversible transformation of biological entities.

In summary, if, as we have suggested, a “modern” conceptualization of responsibility makes us into consequentialists, then we have to exercise foresight about our actions, particularly when seeking to innovate. But insofar as we rely on knowledge (and in particular science) for foresight, we appear to *lack the capacity* for foresight – because our attempts to understand the consequences of our actions are simultaneously undercut by attempts, whether by ourselves or by others, to change the world we are trying to understand. If, as Hacking suggests, we know we can “expect the unexpected,” what – as consequentialist moral agents – do we do **with this** knowledge? Because of their potential seriousness, we are called upon to anticipate global and long-term future consequences of our actions. People and non-humans distant in time and space become our moral neighbors. But how can we act responsibly toward them if we concentrate on the consequences of our actions? If these extend beyond our foresight, then are we pushed toward paralysis (Grinbaum, 2006) rather than toward action?

### 7.2.3 Reciprocal and Non-Reciprocal Responsibility

One response to this consequentialist crisis is to propose that we become deontologists again and search for a set of fixed rules on which we can rely in the face of uncertainty. To do this, however, we have to find an immovable point on which to stand, one which does not refer, as consequentialists must, to some method for calculating the relative balance of the benefits and costs of an action. Perhaps we could rely on an account of a “special moral sense” that human beings possess and which guides them in the face of uncertainty, or rely on “the intuitions of the ‘best people’,” those (like scientists



*What Is "Responsible" about Responsible Innovation? Understanding the Ethical Issues* 127

or business people) whose points of view tend to be accorded a special status and authority in modern societies? (Richardson, 1999, p. 232). Such suggestions, however, are question-begging ones in an age characterized by Weber's *Zweckrationalität*, where rationality depends on results, and any other form of justification must appear to be only an illegitimate argument from authority.

We could consider another kind of deontological ethics, however. We could look for guidance on how to *act responsibly* within the structure of our basic relationships with other humans. For example, a series of classical liberal and republican philosophers, including Locke and Rousseau (along with earlier thinkers like Hobbes), depicted our social relationship with others as being, essentially, contractual – although they did this in different ways, which philosophers classify into “contractarian” and “contractualist” positions (Darwall, 2003). The existence of society assumes an implicit and fundamental contract in which reciprocal rights are guaranteed and enforced by legal authority. Under the terms of such a contract, I demonstrate my responsibility as a citizen by ensuring I act only in ways that do not infringe other citizens' legal and moral entitlements – in the expectation that they will reciprocate by doing the same. Beneath all the role responsibilities we may have, then, we are fundamentally citizens *whose fundamental responsibility is to reciprocally respect each others' rights*.

Does this help, however, when we are thinking about humans outside our community, non-humans, and future generations of people and/or non-humans? Here, the assumption that morally-significant relationships are reciprocal ones falls down. Contractarianism, for example, represents the basis of reciprocity as mutual advantage – that, in other words, I gain from not harming you because then you will not harm me. However, people on the other side of the world who we may be able to harm via the interference effects of our technologies may well not be able to “harm us back”: if there is no possibility of reciprocity, then no obligation falls upon us. Even more problems are encountered with respect to non-humans who cannot, by their nature, enter into such contracts. The problem is yet more intractable when considering future humans, who are logically unable to enter into such contracts. We do not and cannot, in principle, stand in a reciprocal relationship to the future generations of people who will inhabit the world shaped by the innovations we create. Our relationship to them is entirely non-reciprocal (Groves, 2009), as we have the power to completely change the conditions in which they will live, while they enjoy no comparable power over us.

This observation does, however, give us a chink of light, through which a better path toward understanding the “responsibility” in responsible innovation might be glimpsed. Our power over future generations – amplified by our reliance on technological innovation and the likelihood of interference effects – is what, according to the German philosopher Hans Jonas (1984), places upon those who live in technological societies a historically unprecedented responsibility for the future. Robert E. Goodin (1985, pp. 170–172) agrees, and spells out some implications of our power over future people: any responsibility we have to future people derives from their *vulnerability* to our actions, and not from a social contract. The position in which we stand vis-à-vis future people is analogous to our relationship with a dependent, such as a child or someone in desperate need. More widely, being vulnerable to others – yet not necessarily reciprocally so – may be seen as a condition which derives from human finitude (Løgstrup, 1997),

and as such is hardly unprecedented in and of itself. This condition imposes upon us, as individuals apart from whatever particular social roles we may play, a basic obligation to care for others whom our actions may affect.

At first sight, this might not seem to help us much. We have reached what appears to be a troubling philosophical insight – that the power of technology potentially creates novel (and problematic) relationships between those alive now and future people. In other circumstances, relationships between present and future people might be mediated by institutions with stability designed into them – such as investment funds, state welfare services, and so on. But intergenerational relationships shaped by our reliance on technology are mediated only by “naturalized” artifacts that seem to have an uncanny life of their own, beyond our control. What can the innovator do, faced with this insight? When consequentialism is of no help, must s/he therefore stand paralyzed, facing the gap between his/her own finite capacities and the seemingly logical demand that s/he take on unlimited responsibility (European Commission, 2008), exposed to the uncertainties of moral luck and the possibility of future blame and liability? In the next section we explore a different approach to responsibility. We show how to draw from the vulnerability of future people some useful insights that help us think about the prospective responsibilities implied by responsible innovation.

### **7.3 Beyond the Sovereign Individual: Collective Responsibility, Desire, and Cultural Narratives**

#### **7.3.1 Passion Sits Alongside Reason**

The problem of responsible innovation stems, as we have seen, from a contradiction that affects the social authority possessed by scientific knowledge. This is the contradiction between the promise of knowledge without intrinsic limits, and the limits that actually emerge when the products of this knowledge, in the form of technology, generate radical uncertainty about the future. Hence, the downfall of consequentialist ethics in addressing the actual problems of responsible innovation that we have discussed above. Human desire is yet another part of the human condition that also possesses no intrinsic limits. Unconstrained by the voice of reason, passions can reinforce themselves with unbounded strength for an indefinite period of time. They provoke emotions, whose contribution to moral judgment stands on a par with the input of a consequentialist calculation of costs and benefits associated with one’s action: hence, the necessity to take them into account in addressing the problems of responsible innovation. We begin by a brief sketch of emotional attitudes that society exhibits, and suffers, with regard to technology.

The twentieth century was replete with reflections and reactions on what was perceived as the failed promises of rational science and the enlightened man. From the Holocaust and the Gulag on the social side, to Chernobyl, Fukushima, the debacle of GMOs, intrusions of privacy, and contaminated medicines (from hemophilia blood products in the 1980s to cardiac drugs in the 2010s) on the technological side, all of these catastrophic events involved unforeseen consequences of technological innovation as either leading vectors or helpful mediators of evil. The psychological and moral experiences that these events induced in human societies had a traumatic character (Frankl, 1946), in which

*What Is "Responsible" about Responsible Innovation? Understanding the Ethical Issues* 129

the claims of a trusted authority were unexpectedly found to be untrustworthy – thus disrupting widely held cultural assumptions about the place of science in society as a reliable interpreter of the world (see Chapter 6, for further discussions on trust). Yet most societies still react with hope and excitement to the promise of new technologies, even as they preserve the memory of past technological catastrophes. Thus the traumatic loss of trust continues to reverberate to this day.

One of the temptations of this trauma is to see it as total and irrevocable, marking out a radical discontinuity between the present of “open-source” science and the past “ivory-tower” science, and one that – to return to epistemological issues – discloses insurmountable gaps in knowledge that mean we cannot predict what our actions will mean to future generations. The existence of this radical separation between us and them is one way of reading Hans Jonas’s remark (Jonas, 1984) that we, as members of technological societies, bear a historically unprecedented responsibility to future generations. If this responsibility is genuinely unprecedented then, to be genuinely responsible innovators, do we need to innovate in the realm of responsibility too, perhaps to the extent of entirely jettisoning the ethical traditions we have inherited? Yet such a conclusion may lead us nowhere. If we decide we need an entirely new kind of responsibility, then our decision leaves us amidst the ruins of old concepts of responsibility, carefully distilled in the course of human history, without any obvious path to take.

As an alternative, one option is to simply reverse the standards of proof, and make “being responsible” identical with “being precautionary.” One could demand (to take things to an extreme) that innovators provide evidence that their innovations are safe with complete certainty, even where potential harms are scientifically recognized to be merely plausible. This is, however, insufficient for an account of moral responsibility. Although the precautionary principle is based on an acknowledgment of the pervasiveness of uncertainty, by itself it represents little more than a negative version of foresight-based consequentialism: grounding decisions on worst-case scenarios still requires that we foresee what these might be, and that we make a judgment whether the benefits of acting are “proportionately” better than the potential hazards of doing so (Dupuy, 2007; Marchant, 2003).

Without any firm guidelines on when exactly precaution must end, we are left caught between an unjustifiable policy of precaution and an unjustifiable policy of *laissez-faire*. The ever-present possibility of paralysis between these options drives home the importance of *emotional* reaction to innovation, and with it the analogy between the trauma of scientific authority and individual trauma: just as an individual trauma sufferer may retreat entirely from the world, unable to deal with its demands, societies may demand that technologies – or individual researchers – be certified as being entirely innocent, as being either free from potential hazards or being required to extend their foresight through additional research to ferret out every possibility of harm. Even more radically, voices may be raised against technological rationality as such, representing scientific knowledge as infected with a will to exploitation, and even totalitarianism. Counterposed to these demands are often found consequentialist responses from exasperated scientists: there is no innovation without risk, so if you want the benefits of new technologies, accept the risks! Such responses, however, miss the point: in technological experiences of trauma, the authority of confident pronouncements about the balance between “benefits” and “risks” is precisely what is at issue. Where the criteria provided by instrumental and consequentialist reason fail, *we are left with emotion or passion*

130 *Responsible Innovation*

as the basis of judgment. Consequently, we cannot expect that trying to decide how to proceed on a rule derived from decision theory, based on exploring the rationality of different ways of weighing benefits against risks, can be an adequate response, even if the decision rule we come up with is beautifully precautionary in nature. What is needed is a recognition that our situation, although technologically unprecedented in the history of humanity, *is not ethically unprecedented*, and rather than simply inventing new decision rules which individuals or organizations can apply, we need to look back and seek to learn from the moral thought of the past about styles of ethical thought which can be effectively applied to our present, “passionate” situation. As we shall see, these lessons may be more complex and demanding than simply learning a rule.

### 7.3.2 Non-Consequentialist Individual Responsibility

If technological societies are marked by experiences that tempt us with fantasies of radical discontinuity – that things are no longer as they were, that “we” and our contemporaries are special just because of *when* we are alive – then one suitably sceptical path forward may be to carefully explore *continuities*, rather than demanding the wholesale invention of a new morality entirely innocent of the alleged sins of scientific rationalism. Given that, as we have seen, the responsibilities of innovators derive from the vulnerability of future people to their actions, can we draw any useful analogies with other roles or forms of subjectivity whose identity revolves around handling vulnerability?

It is first of all important to note that, contrary to a commonly-held view that a victim of wrongdoing of any kind enjoys a superior moral standing over the ordinary man, being vulnerable does not equate with being right, nor being righteous. Victimhood caused by the feeling of vulnerability does not suffice on its own to determine the meaning of action, but merely serves as one of its motivations. As Robert Musil wrote, “Even a sex-murderer is, in some cranny of his soul, full of inner hurt and hidden appeals; somehow the world is wronging him like a child, and he does not have the capacity to express this in any other way than the way he has found works for him. In the criminal there is both a vulnerability and a resistance against the world, and both are present in every person who has a powerful moral destiny” (Musil, 1990/1913, p. 39). Hence the need, once the moral dimensions of vulnerability are explored, to add to our ethical thinking a different component taking us beyond a mere observation of this condition. However, we first focus on vulnerability as a factor of individual responsibility.

We mentioned previously that the relationship between present and future people was more akin to that between caregivers and dependents or children than between adult contemporaries, given that it generates non-reciprocal responsibilities. Yet this is not perfectly comparable, by any means. For example, the dependents of caregivers (and children in particular) are not “future people,” but present people whose potentialities we assume are particularly sensitive and malleable, and who are vulnerable as a result. This point, however, means we can make a useful comparison between a parental duty of care and a similar duty which might be appropriate for innovators, based on an analogy between “naturalized” technologies (technologies, that is, with something like an autonomous existence of their own) and children.

We might say that parents are required to care for children in such a way as to encourage certain kinds of character traits and behaviors aligned with social norms.

*What Is "Responsible" about Responsible Innovation? Understanding the Ethical Issues* 131

Parents may be thought of as having a duty to future people (and to their children's contemporaries) not to raise offspring who ignore their responsibilities to others. Their responsibility to future people is, therefore, mediated by their responsibilities to their children, and vice versa. Further, the role of parent is without the kinds of well-defined limits we may expect to find drawn around an occupation, to which explicit role responsibilities might attach. 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 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 future consequences, which, as we have seen, is an expectation that technological societies both promote and undermine. The capacities required are other than this, belonging to the set of personal characteristics that has traditionally been called "virtues." The purpose of raising children on this pattern is to make them fit them for adulthood, for relative, then full, autonomy, and for taking responsibility on their own account.

Are innovators parents of their innovations in this sense? As they are unleashed to live a "life" of their own in a complex world, technologies can develop a certain autonomy – though one that obviously does not make them identical to human children, particularly because technological innovation carries an aura of novelty that, arguably, human procreation does not. During early elaboration stages, technologies are dependents with sensitive and malleable potentialities, in which it is hoped the ingenuity of innovators will help produce – in concert with the sleeping powers of nature that are called to life by the innovator – particular virtues. These typically include the efficient production of "right impacts"; that is beneficial, intended effects (see Chapter 3), together with the kind of stability set out in ideas of "safety by design" (Kelty, 2009), and the broader social character of a technology implied by the concept of the "social constitution of technologies" (Grove-White, Macnaghten, and Wynne, 2000). Such obvious virtues can even be formulated in the ordinary language of moral values rather than technical specifications: to rephrase the slogan promoted by one technological company, not only the innovator but technology itself must not "be evil." Those involved in research and development therefore partake, in many ways, in preparing their technological children for maturity – although we should note that if naturalized technological artifacts with the potential for unforeseeable interference effects are children, then they are certainly endowed with a special social status – not truly alive yet decidedly not simply inanimate slaves . . . They may not be capable of maturity in the sense that we take human children to be, but they cannot simply be fixed in the position of eternal childhood either – for to do so is to be unjust to their caring and responsible "parents."

We see the problems that arise from such an "infantilization" of technology in the precautionary tendencies of contemporary governance of innovation. For example, the European Code of Conduct for Responsible Nanosciences and Nanotechnologies Research contains a principle of accountability that reflects the traumatized precautionary stance we examined above: "Researchers and research organizations should remain accountable for the social, environmental, and human health impacts that their research (in nanosciences and nanotechnologies) may impose on present and future generations" (European Commission, 2008). This reference to future generations without any time limit for accountability, from the point of view of consequentialist ethics, subjects the



innovator-parent to unbounded hazards of moral luck. It is notable that we do not accept such infantilization of technology in relation to many past inventions. Today, we do not hold the inventor of the locomotive, telephone, or internal combustion engine responsible for any negative impacts that these technologies have exercised on humanity since the nineteenth century. It does not occur to us to blame high costs of electricity and an ensuing increase in poverty on Edison or Faraday, or illnesses caused by inhalation of particulates on Rudolf Diesel. However, if “should remain accountable” were replaced with “care for the vulnerability of technology users” within a finite time span of individual quasi-parental responsibility, then such an excessive extension of responsibility would be altogether removed.

Let us pause at this point to review the usefulness of the analogy with parental care. It is informative to the extent that it illustrates how the future-oriented responsibilities of innovators, based on vulnerability, can be understood in terms of what mediates between present and future for caregivers and for innovators – respectively, children and technological innovations. The duty of care for the vulnerability of future people becomes concrete, as a result, in the idea of a duty to “teach” or “encode” the virtues in children or in created artifacts. But before we ask which virtues should be taught in this way, the analogy itself leads us to examine, beyond the individual dimension, the relationship between the innovator and the wider social order. To what extent does the technological innovator adopt a particular *collective, hence political*, role, in addition to the “parental” one?

### 7.3.3 Collective Political Responsibility

Our analogy with parental responsibilities was presented to illustrate an important continuity between responsible innovation and extant conceptions of responsibility, to avoid an excessive and paralyzing concentration on the uniqueness of technological responsibility. We now examine a further continuity: between responsible innovation and *collective, political responsibility*. This is relevant to innovation as bringing new social practices and even institutions that transform the ways in which human beings interact with their peers and the world around them.

Society often believes that the innovator creates in order to serve some identified social need. The innovator herself may indeed be motivated by the desire to mend social injustice or to do social good. In her innovation, however, she contributes to a process that often reshapes the social and natural worlds in unforeseeable ways. In this sense, innovators are (*pace* Shelley, 1970/1821) the unacknowledged legislators and co-creators of the world. They thus adopt a political responsibility as part of a particular professional group engaged in a collective endeavor.

Hannah Arendt’s analysis focuses on the collective aspect of responsibility and its implications for social groups. A German Jew who fled Germany in 1933, Arendt chose to turn her experience of ethical and social trauma into something beyond a question of her own, or anyone’s, individual ethics. She did not ask whether an individual is good but whether his or her conduct is good for the world s/he lives in, and she emphasized the political dimension of her thought: “In the centre of interest is the world and not the self” (Arendt, 1968, p. 151). The social structures observed by Arendt directed her

*What Is "Responsible" about Responsible Innovation? Understanding the Ethical Issues* 133

attention to the notion of collective responsibility. By definition, collective responsibility occurs if the following two conditions are met: a person must be held responsible for something she has not done, and the reason for her responsibility must be her membership in a group which no voluntary act of hers can dissolve. Thus, all nuclear scientists share political responsibility for the human condition ~~that we share~~ in a world full of atomic power plants and nuclear weapons, irrespective of their degree of personal involvement in the industry; or all scientists in general partake in shaping the world, whatever their individual research disciplines might be. Collective responsibility looms large, not in considerations regarding individual actions based on personal convictions about what is right, but in *political considerations of a group's conduct*. In contrast to, for example, KJaspers (1947), Arendt maintains that collective responsibility is a concept quite distinct from the concept of guilt. She argues that the notion of collective guilt only serves to exculpate those individuals who are actually legally guilty of specific evils, while collective responsibility is a moral and political, but not a legal, phenomenon, and relates to collective well-being under changing technological realities. This form of responsibility arises when complex social forces become important that cannot be reduced to individual will or intent, but the responsibility for the consequences of the action of such forces is, nevertheless, attributed to individuals who compose the group: this is the case for innovation!

Collective political responsibility rests on historical continuity, flowing from the past and reaching out to embrace the future. The individual who recognizes his or her implication in collective responsibility recognizes that who s/he herself is, what s/he values and how s/he acts – in other words, his or her own identity – is inseparable from a social identity that is historical in nature, and rooted within particular institutions. Taking responsibility means assuming an account of the history of the institutions to which one belongs, and which shape who one is, as much as it means assuming responsibility for shaping the future consequences of what they do with an eye on wider well-being. Weighing the rights and wrongs of an action here goes beyond, for example, simply concentrating on the narrow technical benefits of an invention. The innovator, as bearer of a political responsibility specific to his or her social role, has to ask herself about the *wider social and political significance of what she intends to accomplish, and what her actions may accomplish despite her intentions*. Here, the limitations of traditional considerations of role responsibility are again apparent: the responsibilities one assumes when playing a role are a medium through which one enters the political sphere, and the role itself is but a gate that leads to an arena where moral judgment takes place.

On a group level, Arendt's conditions for collective responsibility fully apply to Nordmann's "naturalized technologies." The perceived autonomy of these artifacts is due to the fact that the internal functioning of complex technological devices remains opaque for the layperson. Science is perceived as a mysterious force that produces useful artifacts, that is, a kind of modern magic. However, even if we acknowledge the autonomy and ubiquity of modern technology, the layperson will distinguish it from magic or fairy tales in that s/he knows that there exist living people, namely scientists and engineers, whose participation in the inner workings of science and technology is direct: they are the initiates. Hence, their collective responsibility: as seen by society, any scientist is engaged in "secret" production of artifacts that will leave a deep mark

134 *Responsible Innovation*

on every man's life. No act of his or her own, even if s/he exits the institution of science completely, will return the scientist to the status of layperson, as long as his or her past scientific training and occupation remain known to the group. Particularly paradoxical cases of political responsibility arise when scientists working in a certain discipline are held responsible for what has been done in a very different domain. The intra-scientific differences that are evident to the initiates remain socially invisible, and, as a consequence, politically irrelevant.

### 7.3.4 The Virtues of Responsible Innovation

We have argued, so far, that responsible innovation means taking responsibility in ways that are, respectively, quasi-parental and collectively political in nature. The quasi-parental way of taking responsibility implies a limited kind of individual responsibility that focuses upon the duty of care for the malleability of technological artifacts and the vulnerability of their future users. However, being vulnerable is no guarantee that either the technological artifact or the person will act virtuously: a victim may turn into a persecutor of those who took care of him/her, given the right circumstances. If the innovator who takes care of the "virtues" of his or her creations is blamed for any unintended consequences, the reaction will primarily be internal rather than external: s/he will feel ashamed and in conflict with his or her own conscience, rather than being liable in the eyes of society and perhaps punished under the law. The latter, political form of responsibility is collective, and the "politics" to which it refers is not the usual kind. The relevant line of division within society runs, not between opposed alliances (such as those on the political left and the right) that represent contrasting interpretations of the "public good," but between technological initiates and technological laypeople, who may very well vote identically at elections, but may well not form the same opinion where the contribution of technological innovation is at stake. How these two groups can live together, and what they ought to do in order to maintain peaceful coexistence, is the central challenge posed by the innovator's specific political vocation. Its collective dimension implies that no one person, however knowledgeable about science and technology, can single-handedly answer this question. The timescale over which evolves the division line between laypeople and the initiates is larger than the scale of usual politics and its actors are diffused and impersonal: on the left (or right), naturalized technologies, on the opposite side, a world-wide constituency of their users crossing all national and natural frontiers. Decision-making in this political configuration is often implicit, and its consequences take years, if not decades, to become visible.

Among the two types of non-consequentialist responsibility we have discussed, quasi-parental responsibility in particular relies on "teaching" certain virtues. This is fundamental, as moral judgment that depends on passion as well as reason includes an emotional evaluation of the technological artifact and the innovator who created it. Preparing to assume such kinds of responsibility is typically not a part of the training received by scientists, industrial entrepreneurs, or managers of scientific institutions. The non-consequentialist character of responsible innovation we have suggested must require particular forms of education. Without wanting to map out in detail what the virtues of the innovator might be, or the educational means of creating them, we present in the closing pages of this chapter a framework for thinking about them.

### 7.3.5 Narratives Take over Where Cost–Benefit Analysis Fails

#### 7.3.5.1 *Desire and Motivation*

A silent alchemist who once unleashed natural processes in the darkness of a laboratory has, with the centrality of innovation to globalized, technological societies, become a political individual whom we call an innovator or technological entrepreneur. The political question asked by contemporary technological societies of such individuals is: in the first place, why would innovators *wish* to make a pact with the sleeping powers of Nature? What did they *want to achieve*? Both the goal and the very *desire* to achieve it are ethically suspect and subject to scrutiny. Here, we should pause and reflect on the problem of desire as such, and more generally on the place of passion in the judgment of moral responsibility; for when human desire is implicated within any ethical framework governing one's actions, what counts is whether there are clearly demarcated limits to it, and not simply whether acting brings results which fulfill the desire. As Davies and Macnaghten (2010) note in a seemingly paradoxical finding in their study of lay perceptions of technology, "getting exactly what you want may not ultimately be good for you." What exactly does this imply for responsible innovation?

We answer this question from a philosophical and practical point of view that rests on two pillars. The first, as we have already discussed, is that by their very nature science and technology, like any creative process, exceed the limits of prudence. There is continuity between the human condition that they contribute to create and the condition, explored in literature, of a hero who confronts powerful natural forces. In his well-known poem "The Age of Anxiety," W. H. Auden contrasts the demands of pure engineering: "The prudent atom//Simply insists upon its safety now, //Security at all costs," with the forces that govern and reward desire and ambition: "Nature rewards//Perilous leaps" (Auden, 2011/1947, p. 7) Hence, if responsible innovation is something more than a rephrased safety protocol, it must inevitably address, not just reason, but also the passion which inhabits a courageous innovator preparing to make a perilous leap.

This analogy between modern innovator and literary hero might help to reveal unexpected moral difficulties to be faced by the former. Scientific discovery and its ensuing transformation into successful technology depend on multiple factors: assiduous research, for sure, but also serendipity and favorable business opportunities. We learn from literature that the latter are not morally innocent: by saying "O opportunity, thy guilt is great," Shakespeare famously made in *The Rape of Lucrece* a moral judgment so puzzling that it either calls for a mythological personification of "guilty" chance (his own solution) or, for the analytic mind, it reveals the need to open up the Shakespearean shortcut from "opportunity" to "guilt," by spelling out what elements may form this chain of logic. This is where the moral suspiciousness of desire comes into play. Under some circumstances, getting exactly what one wants may lead one to unforeseen disasters and catastrophes: "be wary of what you wish for..." These circumstances exist when what *may* potentially be wished for is itself boundless, like the never-ending technological progress, which can be the cause both of great expectations and of the disruption of the community (Dupuy, 2010).

This moral conundrum is not unknown in history. The notion that too much success incurs a supernatural danger, especially if one brags about it, has appeared independently in many different cultures and is deeply rooted in human nature (Dodds, 1951, p. 30).

Ancient Greek mythology, and later Greek thought, distinguish between four different kinds of circumstances: successful action may provoke jealousy of the gods (*phthonos*), it may lead to divine retribution (*nemesis*), it may cause complacency of the man who has done too well (*koros*), or it may lead to arrogance in word, deed, or thought (*hubris*). *Hubris* is condemned by the Greek society and punished by law, but reaction to the other three is more subtle. *Phthonos* and *nemesis* are dangerous and must be feared. The attitude that the Greeks have toward *koros* is rather ambivalent: the complacency assumed in this notion makes someone's life untenable, however *koros* can hardly be avoided, for it goes hand in hand with ambition, or the inability to put an end to one's desire of great achievements, called *philotimia*. In a telling example, a modern commentator connects Ulysses's hardships with his *philotimia* in a way that bears striking resemblance to the innovator and his or her limitless desire to bring new technologies to life: "[it] condemns Ulysses to a hard life, for he must constantly live up to the height of new dangers, unless the reputation of his past deeds be tarnished. Peace of mind is forbidden to him, because he depends on a reputation placed under continuous threat" (Gangloff, 2006, pp. 103–104, our translation from French). In the later centuries of Greek thought we find an explicit argument describing the moral condition of a man who has achieved great technical feats as "always on fire from fervor," his soul "consumed by a continuous suite of loves, hopes, and desires," the reason being that "the sweetness of success lures him into a painful ordeal of the worst misfortunes" (Festugière, 1954, our translation from French). Thus perfect success forbids peace of mind, and, by way of analogy between ancient and modern ethical thought, this is at the same time a part of the innovator's human condition and a moral problem of its own. The impossibility to limit one's desire endlessly amplifies ambition, and the only way to escape from this eternal fire is via balancing one's desire with humility that would help to restore one's mind to peace. How exactly this can be achieved, and whether this is at all possible, cannot be answered in full generality; what needs to be done instead is an educational effort that would teach the individual to compensate his or her own virtue of scientific ambition with virtuous lucidity, inasmuch as the moral standing of this ambition is concerned, thus contributing to an accrued sense of innovator's responsibility.

#### 7.3.5.2 *Cultural Narratives*

The second pillar is the importance of *stories* for ethical thinking. Several recent publications insist on their relevance, both practically observed and theoretically motivated, for understanding public perception of new technologies (Davies and Macnaghten, 2010; Dupuy, 2010; Ferrari and Nordmann, 2010; Grinbaum, 2010). Ancient and modern narratives become part and parcel of the social reading of technology, making it impossible to tackle ethical questions that it raises without an evocation of mythological personifications of various technical feats and the ensuing moral punishment, for example, Prometheus, Daedalus, or Pandora. Thinking about moral questions with the help of stories is to virtue ethics what cost–benefit analysis is to consequentialism, and the ever more evident irrelevance of consequentialism to the present science–society situation makes it urgent to resort to other tools of dealing with the growing number of problems. We survey here two such stories that are particularly relevant for the analysis of responsibility. As with all myths or narratives, they do not contain a direct answer to the moral question that they explore. Rather, they proceed



by encouraging the scientist and the innovator to reflect on the sides of moral judgment that typically are not a part of his or her rational toolkit. If consequentialism cannot but fail to predict the exact future consequences because of their high uncertainty, one can still imagine what future ethical implications might be by resorting to cultural narratives. Their interpretation may reveal a surprising degree of analogy with the scientific and technological work of modern scientists and engineers.

The first story concerns Rabbi Judah Loew of Prague, to whom a legend ascribes the creation of an artificial man called the Golem of Prague. Rabbi Loew wrote, "Everything that God created requires repair and completion" (Sherwin, 2004, p. 53). On this interpretation of a Biblical verse in *Genesis* 2 : 3, which isn't uncommon in the Jewish tradition, the world was "created to be made": God has not finished his creation and therefore human beings receive a mandate to act as "God's partners in the act of creation," by developing raw materials and unleashing the sleeping powers of Nature. Not only is innovation *per se* free of sin; it is encouraged and praised as a mandatory activity in one's fulfillment of his human potential. Like modern technology that is said to serve societal needs, in the Jewish tradition human creativity is always purposeful: Judah Loew creates the Golem of Prague, not on a whim or for pleasure, but in order to protect the city's Jewish community from the many threats they encountered in the gloomy streets of Prague in 1580. Once unleashed, the golem obeyed Judah Loew's commands and successfully protected the Prague ghetto for about 10 years. Then, according to one popular version of the story, the golem went berserk, at which point Judah Loew was summoned and told to do something to stop the golem's wrongdoing. He "unmade" the golem by a procedure that was, "technically" speaking, the reverse of the method he had used to make him.

This legend exemplifies several typical features of the many golem stories in Jewish literature that may cast new light on modern science and technology. In reflecting on such stories, we may learn more about the complexities of moral judgment. Points of comparison between the golem legends and modern techno-science include: (i) purposefulness: a golem is made on purpose by a human creator with a specific goal in mind, while modern technology is often justified before society as being created in order to serve identified social needs; (ii) reversibility: a golem can be both made and unmade through a fixed procedure, while modern technological innovation can change the world so dramatically that one can hardly envisage going back; (iii) machine-like obedience: the creator commands his creation at will and the latter obeys the former, while modern naturalized technologies gain a form of autonomy that demands they be granted a special, intermediate social and moral status; and (iv) responsibility: when the golem's actions become harmful, the community tells Judah Loew to repair the damage. Responsibility for the golem's conduct falls upon his creator rather than upon the golem himself, and this in spite of the fact that the golem behaved and looked more or less like an autonomous human being. This is strikingly similar to the quasi-parental responsibility of the innovator we have discussed earlier.

The second story concerns Mary Shelley's novel about Victor Frankenstein, which displays a different set of characteristics (Shelley, 2009/1818). Unlike Judah Loew, Frankenstein, who created a monster, cannot undo what he had done: the monster would not obey him and escapes his power altogether. The process unleashed here is irreversible, but even as it begins to produce terrible consequences (as the story develops

the monster kills several people), Frankenstein keeps his moral perplexities to himself. He evidently refuses to acknowledge any political dimension of his action. His responsibility with regard to society, which happens not to be imposed on him by legal or any other external threat to his own person, proceeds exclusively from his own conscience. And although he is perturbed by the monster's actions, he does not reveal that he has created it, nor does he admit what he knows of its deeds, thus allowing one person falsely accused of murder to be executed. What places him in this position is the modern version of what Augustine of Hippo called the "lust of the eyes" (Augustine, 2000/398, Chapter 35; O'Neill, 1993, pp. 155–59), the desire for scientific truth and technical achievement above any other effect produced by the innovator's desire. The story then goes on to explore the consequences of Frankenstein's failure to admit his political responsibility. Soon the monster promises to put an end to both his and others' suffering if Frankenstein makes for him a second artificial creature to become his wife. Seduced by an easy technical solution to the problem of social evil, Frankenstein complies and begins to work on the second monster, only to realize a little later that by making this new creature he would unleash yet another irreversible process out of his control. He refuses to finish the second being and flees the country and all human company, apparently unable to cope with a moral burden.

Shelley's verdict is unequivocal: Frankenstein's creative activity was morally wrong, for it failed to stand up to the moral and political challenges it had itself generated. But why *precisely* was it wrong? Unlike Judah Loew, Victor Frankenstein created the monster without a particular societal goal – is this the source of evil? Or is it the lack of reversibility? Or the lack of control, whereby the monster's autonomy placed him altogether out of his creator's control? A small episode in the novel reveals further complexity by proposing a parabola about the source of evil in the monster, which we can interpret as a story about good and evil in modern technology (Pavlopoulos, Grinbaum, and Bontems, 2010, Chapter 5): after the monster's initial escape from Frankenstein, he finds refuge in a hovel next to a small house inhabited by a blind man and his two children. By observing the family and reading their books, the monster learns human language. Gradually he warms up to the poor family and starts secretly to help them. One day, longing for mutual kindness, he decides to come out to his hosts. First he enters into a conversation with the blind man and is received warmly by him. But when the children arrive and see the monster, they beat him and throw him into the street. At this moment the monster puts an end to his righteous conduct and turns to wrongdoing.

This episode mingles the usual theme inherited from the Golem legends (that social success and the moral status of one's novel creation depend on the purity of the creator's intentions), with the unpredictability of nonetheless morally relevant consequences, otherwise known as moral luck (see above). Shelley contends that evil influence in the monster is not necessarily due to a lack of reversibility in the original innovation, nor of course to Frankenstein's revealed evil intentions, but to the human conduct on which the monster models his own behavior. When the blind man's children beat the creature, he learns from experience and, henceforth, starts to spread evil himself. Taking this episode as a metaphor for the condition of modern technology, one might contend that the responsibility for misuse of technological innovation belongs with the society rather than the inventor; technology would not be prone to misuse *per se*, nor would

*What Is "Responsible" about Responsible Innovation? Understanding the Ethical Issues* 139

such misuse be inevitable. If it occurs, then it is rooted in the environment in which technology operates rather than being encoded deterministically in the technical object. In other words, as we frequently hear today, moral judgment depends on how technical objects are used, while the existence of the object itself is neither good nor bad.

Yet Shelley gives reasons to doubt this interpretation. Whether the source of the monster's wrongdoing is in his creator or in a random chain of events that happened to the monster after his escape, Victor Frankenstein still feels an unbearable responsibility that forces him to flee and abandon both his work and his world. Hence, evil done by the monster has something to do with Frankenstein himself. When the latter halts the creation of the second being, it is not because he suddenly mistrusts the monster's promise to live peacefully in the woods with his future partner. Rather, he realizes that episodes such as the meeting between the monster and the blind man's children are inevitable because they are a consequence of his own finiteness, and of the dark side that is inherent to Frankenstein as human creator. Angelic, purely righteous beings cannot subsist, as Melville will make clear a few decades after Shelley's novel by putting to death his Billy Budd (Melville, 1924; Rey, 2011). Frankenstein knows that his political responsibility for what the monster will do to society, although not limitless, is nevertheless very real: he cannot come to terms with his conscience, affirms his own responsibility, although no legal threats are made against him, and flees society.

To some extent, the innovator today is put in all these different situations at once: on the one hand, society exerts pressure on him if his work proves harmful; on the other, by turning inwards and interrogating his or her own conscience, the innovator must make a choice between his or her ambitions and desires, and face moral judgment even if (or perhaps, especially if) they are successfully realized. Yet there is no universal answer as to how to translate the lessons of old stories into action in the present. Even as one strives to possess the requisite virtues of the responsible innovator: to bind one's desire, to check ambition by humility, and to maintain both internal interrogation and external dialog about the meaning of one's actions, there is no guarantee that moral luck in the uncertain future will not mean that one's efforts to act responsibly will not turn out to have unintended consequences. Whatever choices are made, the final verdict on a distinction between responsible and irresponsible innovation is not in our capacity to make. No one can vouch that his action is an adequate expression of the virtues of a responsible innovator: rather, living up to the demands of responsibility is a lifelong process.

#### **7.4 Conclusion: Responsibility and Meaning**

Narratives teach continuity and comparison, and it is through developing such reflective skills about the cultural meaning of innovation that the virtue of responsibility can be developed. By tracing out, as they unfold, how the consequences of action ramify through time in unpredictable ways, narratives can also teach the limits of foresight and rational prudence. One finds such limits, too, by pursuing a purely analytic approach. Thus the downfall of consequentialist ethics exemplified by the notion of moral luck and the failure to take into account passion and desire call for alternative tools for ethical reflection, and virtue ethics supplies such instruments in the form of narratives. A refusal of virtue ethics to carry out moral judgment too hastily, before all its complexities can

be fully contemplated, is a necessary ingredient of the moral education that the scientist should receive in the framework of responsible innovation.

Yet responsibility is only one of the virtues. Others, relevant to the scientific endeavor, perhaps include integrity, impartiality, honesty, lucidity, or openness. But even practicing all these virtues is no ultimate guarantee that good intentions, schooled by a sensitivity to vulnerability, shaped by quasi-parental care and tempered through responsive dialog and political deliberation, will not produce bad effects. It is the meaning of action as a moral phenomenon that ultimately matters. As Musil reminds us, this meaning cannot be exhausted by excusability or inexcusability of occasional bad effects (Musil, 1990/1913, p. 38), for, on the one hand, political consequences of innovation for the ways in which people live in a society and its impact on the human condition will exceed the scope of any hasty judgment. On the other hand, the judgment itself involves two types of actors: the human innovator, of course, but also the naturalized technology in its capacity of semi-autonomous force that possesses the power to help shape our common destiny. As a result, as far as the innovator is concerned, the final verdict on whether he or she has acted responsibly or irresponsibly remains pending. As Paul Tillich wrote, “In the anxiety of guilt and condemnation doubt has not yet undermined the certainty of an ultimate responsibility.” The occurrence of bad consequences, therefore, is not a final ethical verdict. The truly final word in moral judgment will remain with the meaning of one’s action: “If, however, doubt and meaninglessness prevail, one experiences an abyss in which [...] the truth of ultimate responsibility disappears” (Tillich, 2000/1952, p. 174). The courage to deal with a complex process that creates such meaning is the best evidence for responsible innovation.

## References

- Adam, B. (1998) *Timescapes of Modernity: The Environment and Invisible Hazards*. London: Routledge.
- Adam, B. and Groves, C. (2007) *Future Matters: Action, Knowledge, Ethics*. Leiden: Brill.
- Arendt, H. (1994/1963). *Eichmann in Jerusalem: A Report on the Banality of Evil*. New York: Penguin Books.
- Arendt, H. 1998/1958. *The Human Condition*. Chicago: Chicago University Press, (2nd edn).
- Arendt H. (2003/1968) Collective responsibility, in: H. Arendt, *Responsibility and Judgment*, New York: Schocken Books.
- Auden, W. H. (2011/1947) *The Age of Anxiety*. Princeton University Press.
- Augustine (398) *Confessions*, Oxford: Clarendon Press, (2000).
- Beck, U. 1999. *World Risk Society*. London: Polity Press.
- Bradley, F. H. 1927. *Ethical Studies*. 2nd edn. Oxford: Clarendon.
- Darwall, S. L. 2003. *Contractarianism, Contractualism*. Oxford: Blackwell Science.
- Davies S. R. and P. Macnaghten, 2010. Narratives of mastery and resistance: lay ethics of nanotechnology, *Nanoethics*, **4**, 141–151.
- Davis, W. (2001) Introduction, in *Taking Responsibility: Comparative Perspectives*. (ed. Davis, W.) Charlottesville, VA: University Press of Virginia, pp. 1–27.
- Dodds, E.R. (1951). *The Greeks and the Irrational*, University of California Press.

What Is "Responsible" about Responsible Innovation? Understanding the Ethical Issues 141

- Dupuy, J.-P. (2007) Rational choice before the apocalypse. *Anthropoetics* **13**(3), 237–261.
- Dupuy J.-P., (2010) The narratology of lay ethics, *NanoEthics*, **4**, 153–170.
- European Commission (2008) Recommendation on 'A Code of Conduct for Responsible Nanosciences and Nanotechnologies Research'. C(2008) 424, Brussels.
- Ferrari A. and A. Nordmann, 2010. Beyond conversation: some lessons for nanoethics, *NanoEthics*, **4**, 171–181.
- Festugière, A.-J. (ed.) 1954. *Corpus Hermeticum*. Les Belles Lettres, Paris.
- Frankl, V., (2006/1946) *Man's Search for Meaning*, Beacon Press.
- Gangloff, A. (2006) *Dion Chrysostome et les Mythes*, Paris: Editions Jérôme Millon.
- Goodin, R. E. (1985) *Protecting the Vulnerable: A Reanalysis of Our Social Responsibilities*. Chicago: University of Chicago Press.
- Grinbaum, A. 2006. Cognitive barriers in the perception of nanotechnology, *Journal of Law, Medicine and Ethics*, **34**(4), 689–694.
- Grinbaum A., (2010) The nanotechnological golem, *Nanoethics*, **4**, 191–198.
- Grinbaum A. and J.-P. Dupuy, (2004) Living with uncertainty: toward a normative assessment of nanotechnology, *Techné*, **8**(2), 4–25.
- Groves, C. (2009) Future ethics: risk, care and non-reciprocal responsibility. *Journal of Global Ethics* **5**(1), 17–31.
- Grove-White, R., Macnaghten, P., and Wynne, B. (2000) *Wising Up: The Public and New Technologies*, IEPPP, Lancaster.
- Hacking, I. (1986) Culpable ignorance of interference effects, in MacLean, D. (ed.) *Values at Risk*. Totowa, NJ: Rowman & Allanheld, pp. 136–154.
- Hart, H. L. A. (1968) *Punishment and Responsibility*. Oxford: Clarendon Press.
- Jaspers K. (2000/1947) *The Question of German Guilt*, Fordham University Press.
- Jonas, H. (1984) *The Imperative of Responsibility: In Search of an Ethics for the Technological Age*. Chicago, London: University of Chicago Press.
- Kelty, C. 2009. Beyond implications and applications: the story of 'Safety by Design'. *NanoEthics* **3**(2), 79–96.
- Krohn, W. and Weyer, J. 1994. Society as a laboratory: the social risks of experimental research. *Science and Public Policy* **21**(3), 173–183.
- Løgstrup, K. E. (1997) *The Ethical Demand*. Notre Dame and London: University of Notre Dame Press.
- ~~Luntley, M. 2003. Ethics in the face of uncertainty: judgement not rules. *Business Ethics: A European Review*, **12**(4), 325–333.~~
- Marchant, G. E. (2003). From general policy to legal rule: aspirations and limitations of the precautionary principle. *Environmental Health Perspectives* **111**(14), 1799–1803.
- McKeon, R. 1990. *Freedom and History and Other Essays*. Chicago: University of Chicago Press.
- Melville, H. 1924 *Billy Budd*. ~~consulted edn. 1992~~, TOR Books.
- Musil R. (1990/1913) *Precision and Soul*. The University of Chicago Press.
- Nagel, T. (1979) Moral luck, in *Mortal Questions* (ed. Nagel, T.). Cambridge: Cambridge University Press, pp. 24–38.
- Newdick, C. (1988) The development risk defence of the consumer protection act 1987. *The Cambridge Law Journal* **47**(03), 455–476.



- Nordmann, A. (2005) Noumenal technology: reflections on the incredible tininess of nano. *Techne* 8(03), 3–23.
- O’Neill, J. (1993). *Ecology, Policy and Politics*. London: Routledge.
- Pavlopoulos, M., Grinbaum, A., and Bontems, V. (2010) Toolkit for Ethical Reflection and Communication on Nanoscience and Nanotechnology, CEA, Saclay, [www.observatorynano.eu/project/catalogue/4ET/](http://www.observatorynano.eu/project/catalogue/4ET/) (accessed 20 December 2010).
- Pellizzoni, L. (2004) Responsibility and environmental governance. *Environmental Politics* 13(3), 541–565.
- Rey, O. (2011) *Le Testament de Melville: Penser le Bien et le Mal avec Billy Budd*, Gallimard.
- Richardson, H. S. (1999) Institutionally divided moral responsibility. *Social Philosophy and Policy* 16(2), 218–249.
- Schneewind, J. B. (1984) The divine corporation and the history of ethics, in Rorty, R., Schneewind, J.B., and Skinner, Q., eds. *Philosophy in History: Essays on the Historiography of Philosophy*. Cambridge: Cambridge University Press, pp. 173–192.
- Shelley, P. B. (1970/1821). A defence of poetry, in *Political Writings Including “A Defence of Poetry”*. (ed. R. A. Duerksen), New York, Appleton-Century-Crofts Inc.
- Shelley M., 2009/1818. *Frankenstein, or the Modern Prometheus*. Oxford University Press.
- Sherwin B., (2004) *Golems Among Us: How a Jewish Legend Can Help Us Navigate the Biotech Century*, Chicago: Ivan R. Dee Publisher.
- Tillich P. 1952. *The Courage to Be*. Yale University Press. ~~2000~~.
- Timmons, M. 2002. *Moral Theory: An Introduction*. Totowa: Rowman & Littlefield Publishers.
- Williams B., 1981. *Moral Luck*, Cambridge, Cambridge University Press.
- Yara, T. 2011. Uncertain risks and consensus building: the HIV crisis as a case study. *Journal of Philosophy and Ethics in Health Care and Medicine* 5, 151–167.

**Queries in Chapter 7**

- Q1. Reference “Luntley (2003)” has not been cited in the text. Please indicate where it should be cited; or delete from the reference list.

UNCORRECTED PROOFS