médecine/sciences 2015 ; 31 (hors série n° 2) : 15-20

The old-new meaning of the researcher's responsibility[1]

Alexei Grinbaum

Chercheur au CEA-Saclay/Larsim, CEA-Saclay, SPEC/LARSIM, 91191 Gif-sur-Yvette Cedex, France.
alexei.grinbaum@cea.fr

The problem of responsible innovation stems from a contradiction that affects social authority possessed by scientific knowledge. This is the contradiction between the promise of knowledge without intrinsic limits, and the limits that emerge when the products of this knowledge, in the form of technology, generate uncertainty about the future. Human desire is yet another part of 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. Their contribution to moral judgement stands on a par with a consequentialist calculation of the costs and benefits of one's action. Responsible innovation must learn to address passions, too.

The 20th 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 haemophilia blood products in the 1980s to cardiac drugs in the 2010s) on the technological side, all of these catastrophic events involved the unforeseen consequences of technological innovation as either leading vectors or helpful mediators of evil. The psychological and moral experiences induced by these events had a traumatic character [1], in which the claims of a trusted authority were unexpectedly found to be untrustworthy – thus disrupting widely held cultural assumptions about science as a reliable interpreter of the world. 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.

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 science' and the past 'ivory-tower science.' It is tempting to believe that it discloses insurmountable gaps in knowledge: we cannot predict what our actions will mean to future generations. The existence of this radical separation is one way of reading Hans Jonas's [2] remark that we, as members of technological societies, bear a historically unprecedented responsibility to future generations. If this responsibility is genuinely unprecedented, then, if we wish to be genuinely responsible innovators, do we need to innovate in the realm of responsibility, too? A positive answer may lead us nowhere. If we decide to install an entirely new kind of responsibility, then our decision leaves us amidst the ruins of old ethical concepts 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 that innovators provide evidence that their innovations are safe with complete certainty. This is, however,

[1] Il ne s'agit pas de la transcription directe de l'exposé d'Alexei Grinbaum, mais ce texte, qui est basé sur un chapitre d'Alexei Grinbaum paru dans [19], reprend l'essentiel des idées que l'auteur a developpées au cours de son intervention.

insufficient for an account of moral responsibility. Although the precautionary principle is based on an acknowledgement 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 judgement whether the benefits of acting are "proportionately" better than the potential hazards of doing so [3].

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. Radical voices may be heard 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 is precisely what is at issue.

What is needed is a recognition that our situation, although technologically unprecedented in the history of humanity, is not ethically unprecedented, and rather than inventing new decision rules, 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.

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. Hence the need for ethical thinking, once the moral dimensions of vulnerability explored, to move beyond a mere observation of this condition.

A parental duty of care can be compared with a similar duty for innovators, based on an analogy between naturalized technologies (technologies 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 behaviours aligned with social norms. 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. Are innovators parents of their innovations in this sense? As technologies are unleashed to live a "life" of their own in a complex world, they 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 human procreation does not. However, technological virtues can be formulated in the ordinary language of moral values and, more importantly, in translation into 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 societal maturity.

The analogy with parental care 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 potentially paralyzing 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 artefacts. 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? 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 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 endeavour.

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 conduct is good for the world he lives in, and she emphasized the political dimension of her thought: "In the centre of interest is the world and not the self" ([4], p. 151). The social structures observed by Arendt directed her 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 specialists in quantum theory bear political (but not a legal)

responsibility for the shared human condition in a world full of atomic power plants and nuclear weapons, notwithstanding their degree of personal involvement in the industry; and 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, Karl Jaspers [5], Arendt maintains that collective responsibility is a concept quite distinct from the concept of guilt. She argued 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 the changing technological realities.

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. Arendt's conditions for collective responsibility fully apply to such naturalised technologies, whose perceived autonomy 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 artefacts, i.e., 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. The layperson, then, assigns them collective responsibility: as seen by society, any scientist is engaged in "secret", i.e. opaque for laypeople, production of artefacts that will leave a deep mark on everyone's life. 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.

We have argued so far that responsible innovation means taking responsibility in ways that are, respectively, quasi-parental and political in nature. Quasi-parental responsibility in particular relies on 'teaching' certain virtues. This is fundamental, as moral judgement that depends on passion as well as reason includes an emotional evaluation of the technological artefact 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 a framework for thinking about them.

A silent alchemist who once unleashed natural processes in the darkness of a laboratory has, with the centrality of innovation to globalised, technological societies, become a political individual. The political question asked 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 judgement of moral responsibility. As Davies and Macnaghten [6] 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 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 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" ([7], p. 7) If responsible innovation is something more than a rephrased safety protocol, it must inevitably address, not just reason, but also 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 favourable business opportunities. We learn from literature that the latter aren't morally innocent: by saying "O opportunity, thy guilt is great", Shakespeare famously made in *The Rape of Lucrece* a moral judgement 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. These circumstances exist when what *may* potentially be wished for is itself boundless, like the never-ending technological progress [8]. 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 inde-

pendently in many different cultures and is deeply rooted in human nature ([9], 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 towards *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 with 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" ([10], pp. 103-104, our translation). 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 fervour", 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" ([11], our translation). Thus perfect success forbids peace of mind, and, by way of analogy, 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 escalation 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.

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 [6, 8, 12, 13]. Ancient and modern narratives become part and parcel of the social reading of technology, making it impossible to tackle ethical questions without an evocation of mythological personifications of various technical feats and the ensuing moral punishment, *e.g.*, 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 judgement that typically are not a part of his or her rational toolkit.

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" ([14] , 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 fulfilment 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 ten 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: (a) 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; (b) 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; (c) 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; (d) responsibility: when 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 [15]. Unlike Judah Loew, Frankenstein, who created a monster, cannot undo what he had done: the monster wouldn't 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" ([16], chapter 35), 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: 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 started to secretly 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. 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 behaviour. 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 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 judgement 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 puts to a halt 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 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 [17, 18]. Frankenstein knows that his political responsibility for what the monster will do to the 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 judgement 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 dialogue 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. ◊

LIENS D'INTÉRÊT

L'auteur déclare n'avoir aucun lien d'intérêt concernant les données publiées dans cet article.

REFERENCES

1. Frankl V (1946). *Man's search for meaning*. Boston : Beacon Press, 2006.
2. Jonas H. *The imperative of responsibility: in search of an ethics for the technological age*. Chicago-London : University of Chicago Press, 1984.
3. Grinbaum A, Dupuy JP. Living with uncertainty: toward a normative assessment of nanotechnology. *Techné* 2004 ; 8 : 4-25.
4. Arendt H (1968). Collective responsibility. In : Arendt H, ed. *Responsibility and judgment*. New York : Schocken Books, 2003.
5. Jaspers K (1947). *The question of German guilt*. New York : Fordham University Press, 2000.
6. Davies SR, Macnaghten P. Narratives of mastery and resistance: lay ethics of nanotechnology. *Nanoethics* 2010 ; 4 : 141-51.
7. Auden WH (1947). *The age of anxiety*. Princeton, NJ : Princeton University Press, 2011.
8. Dupuy JP. The narratology of lay ethics. *Nanoethics* 2010 ; 4: 153-70.
9. Dodds ER. *The Greeks and the irrational*. Berkeley : University of California Press, 1951.
10. Gangloff A. *Dion Chrysostome et les mythes*. Paris : Éditions Jérôme Million, 2006.
11. Festugière AJ (ed). *Corpus hermeticum*. Paris : Les Belles Lettres, 1954.
12. Ferrari A, Nordmann A. Beyond conversation: some lessons for nanoethics. *Nanoethics* 2010 ; 4 : 171-81.
13. Grinbaum A. The nanotechnological golem. *Nanoethics* 2010 ; 4 : 191-8.
14. Sherwin B. *Golems among us: how a Jewish legend can help us navigate the biotech century*. Chicago : Ivan R. Dee, 2004.
15. Shelley M (1818). *Frankenstein, or the modern Prometheus*. Oxford : Oxford University Press, 2009.
16. St Augustine (398). *The Confessions*. Oxford : Clarendon Press, 2000.
17. Melville H (1924). *Billy Budd*. New York : Tor Books, 1992.
18. Rey O. *Le testament de Melville : penser le bien et le mal avec Billy Budd*. Paris : Gallimard, 2011.
19. Owen R, Bessant J, Heintz M (eds). *Responsible innovation: managing the responsible emergence of science and innovation in society*. Wiley, 2013.

TIRÉS À PART

A. Grinbaum