

Introduction

John Rawls famously stated that justice is the first virtue of social institutions as truth is of systems of thought (Rawls 1971). In the context of technological innovation, which forms the background of my paper, we could paraphrase Rawls and say that *responsibility* is the first virtue of innovation as justice is of social institutions. The paraphrase is not supposed to work as a motto for my paper. Rather, it is a problematic statement relating to an empirical fact. A press release issued by the European Commission in November last year may serve as an illustration. The headline of the press release read as follows: 'EU-wide poll shows public support for responsible research and innovation'. According to the poll, most Europeans (76 percent of the respondents) want science and research to be carried out 'with due attention to ethical principles' (European Commission 2013). The overarching principle supposed to ensure this is, as the headline indicates, that of *responsibility*. This is also indicated by university strategic commitments and large scale funding of research programs for "responsible research and innovation" (RRI) in Europe.

The programmatic stress on responsible (research and) innovation in European institutions, and the widespread use of the concept of responsibility in the formation of public opinion (cf. the EC press release), seems to demand some clarity as to what responsibility as an ethical concept means in this context. However, as we shall see in the following, it is quite unclear what responsibility means and *can mean* as an ethical concept in this context – and thus how it is supposed to work as a kind of first virtue in the ethics of innovation.

The obscurity of the references to responsibility in this context is problematic. In this paper I want to focus mainly on some suggestions as to what responsibility *can mean* – on how a philosophically sound concept of responsibility can be conceived of – in this context. The discussion unfolds on the background of Dieter Birnbacher's distinction between two types of responsibility: 1) *ex post responsibility*, meaning responsibility ascribed to an agent for an act committed, and 2) *ex ante responsibility*, which means responsibility ascribed to an agent for the production of a certain state of affairs, with the acts realizing this state of affairs lying in the future (Birnbacher 2001). On the basis of an exposure of the shortcomings of *ex post* responsibility in the context of innovation, followed by an explication of the necessity of an *ex*

ante concept, I look into different ways of conceiving of a form of *ex ante* responsibility suitable for our current technological situation. Outlining some criteria for this form of responsibility, I identify two positions with very different approaches to the question of the ethical status of responsibility. One is the position of Hans Jonas, who conceives of responsibility as an ethical principle structuring moral behavior. The other is found in the reflections on Responsible Research and Innovation (RRI) by René von Schomberg, who has been an influential figure in the European Commission's work on the ethical issues of science and technology. Here responsibility is conceived of as a kind of "tool" for responsive communication and collaboration among stakeholders on ethical issues related to research and innovation. As this conception depends on external ethical principles, I argue that it represents a form of demoralization of responsibility, which leaves it standing on a shaky moral ground. Suspending this concept, I argue in favor of a critical rehabilitation of some basic thoughts in the philosophy of Jonas, which I suggest makes better sense of the idea of a responsible subject within the field of innovation. Finally, I suggest taking a step from the Jonasian ethics of responsibility, with its strong Kantian influence, towards the Hegelian concept of *Sittlichkeit* – a concrete social morality supposed to disentangle responsibility from the paradoxes of subjectivist morality.

Two problems: the subject and object of responsibility

To get a good grip of what it is that makes it so difficult to conceive of responsibility in the context of technological innovation, let us first have a quick glance at what we may call the classical, juridical concept of responsibility. While the theoretical debate on responsibility can be traced back at least as far as Aristotle's discussions in the *Nicomachean Ethics* – and onward through the philosophical and theological debates on free will and determinism by Empiricus, Cicero, Augustine, Erasmus and Luther, leading up to the classical discussions in Hume, Kant and Hegel, among others – the modern meaning of the term, as it develops during the 19th and 20th centuries in particular, is most clearly expressed through its juridical usage. In civil law, responsibility comes to be understood as the obligation to make up or to compensate for the harm one has caused through one's own fault (which in certain cases is

defined by law). In penal law it comes to be understood as the obligation to accept punishment. This way of conceiving of responsibility, i.e. as a retrospective principle evoked after the events have taken place, corresponds to what Dieter Birnbacher (2001) calls *ex post responsibility*, and can even today be considered to be the standard way of understanding and using the terms "responsibility", "responsible" etc. in a juridical context.[1]

The first problem this form of *ex post* responsibility faces in the context of innovation concerns the *subject* of responsibility. In the classical juridical sense, responsibility is to be ascribed to an agent capable of recognizing himself as the author of the act in question (cf. Kant). If the agent cannot do this, he cannot assume responsibility for the act, and it would therefore be problematic to punish him or make him compensate for it. Conversely, if the agent *can* recognize himself as the author of the act, he will be able and obligated to assume responsibility for it. In the context of technological innovation, this criterion for ascribing responsibility in the *ex post* sense faces a series of problems. First of all, the processes of innovation often involve so many people or groups of people that it is impossible to point out a single subject – whether collective or individual – responsible for initiating the action that led to the condemnable consequences. Consider, for example, the famous case of the invention of nuclear science, which finally led to the catastrophes of Nagasaki and Hiroshima. Who are we to hold responsible for this catastrophe: the scientists, the developers of the bomb, the politicians and military strategists, the pilot who pushed the button? They all undoubtedly played a role in this catastrophic event. But we would hardly hold the innovators of nuclear science and all the people involved in the development of this science responsible for the catastrophes involved in the use of nuclear weapons. So in what sense can we decide whether their research and innovation entail responsibility or not?

While the problem regarding the subject of responsibility is serious and difficult to solve, it is not principally insoluble within the framework of *ex post* responsibility, since in theory these agents could all recognize themselves as authors of actions leading to the catastrophic events. A second problem, however, which concerns a displacement of the *object* of responsibility, tears this framework apart. Within the framework of *ex post* responsibility, the object of responsibility is the harm caused by the action of an agent. To ascribe responsibility therefore requires a clear view of the action and its (actual or possible) consequences: the agent is declared responsible with reference to the harm his action causes or may cause. In the case of technological innovation, however, we are dealing with possible effects that are

largely unknown to us, and that may stretch far into the future. The case of GMOs (genetically modified organisms) may illustrate this point. Currently there exists no scientific certainty about what the impact of GMOs being released into the environment may be, and possible negative effects – ecological as well as social and economic – may not reveal themselves for generations. The same holds for several other forms of biotechnology and for nanotechnologies. Responsibility in such cases concerns some possible, future state of affairs that may or may not prove harmful to someone. So when we raise the demand that those involved in the development of new technology act responsibly, we do not call on them merely to make up or to compensate for harm done, or to accept punishment for their deeds. On the contrary, we call on them to *act responsibly* in the sense that their actions *do not produce undesirable consequences* – so that their actions may not produce a state of affairs causing harm to anyone. We call on the innovators in the field of biotechnology to be careful so that their activities do not alter human beings and their environment in ways that could prove harmful to them; we call on corporations not to use risky technology; we want politicians to regulate and monitor the access and circulation of potentially harmful technology on the market, and so on.

In short, in addition to taking responsibility for one's actions, responsibility in the context of innovation means acting in such a way that one's actions do not produce a state of affairs causing harm to any present or future being. And for this purpose, any concept of responsibility that focuses solely on the imputation of an action to an agent, as is the case with the *ex post* concept, will not suffice.

Reframing responsibility: Hans Jonas vs. the European Commission

Through the explication of the shortcomings of the *ex post* concept we see the contours of a different form of responsibility, that we with Birnbacher can label *ex ante* responsibility: A form of prospective responsibility that is to be evoked *before* the events take place. Rather than focusing mainly on ascription, imputation, accountability etc. – which are all typical focuses for *ex post* responsibility – the main focus here is on the relation between the agent

and the people (present or future) potentially affected by his or her actions.

An interesting question at this point concerns the ethical status of this concept. We can identify two poles in the literature on this topic:

1. On the one side, we have responsibility as an *ethical principle*, canonized in the philosophy of Hans Jonas (2003).
2. On the other we have the reflections on RRI by René von Schomberg (especially in Schomberg 2011). Here responsibility is thought of as a principle, or rather a kind of tool, supposed to ensure rational communication and collaboration between stakeholders on ethical issues.

In the first case, we are clearly dealing with a *moral* concept of responsibility; for Jonas, responsibility is a supreme ethical principle imposing on us specific moral obligations (we will return to this below). In the second, however, it seems as if responsibility is in a sense *demoralized*: Responsibility in von Schomberg's version of RRI is not a moral concept in itself, but a tool for structuring action according to external ethical principles, as can be seen in his proposed definition of RRI:

Responsible Research and Innovation is a transparent, interactive process by which societal actors and innovators become mutually responsive to each other with a view to the (ethical) acceptability, sustainability and societal desirability of the innovation process and its marketable products (in order to allow a proper embedding of scientific and technological advances in our society). (Schomberg 2011:9)

The formal character of the procedure seems to suggest that it is in a sense morally neutral. However, as it is developed and initiated out of moral concerns – one of its main aims is to make science and technology compatible with shared moral values – and relies totally on external ethical principles, it is of course embedded in ethics. Resting on a contingent moral ground, however, it stands in danger of becoming just another tool for structuring ethical discourse without any real moral obligation imposed on the parties involved.

At this point we might speak of the adventure of responsibility becoming a misadventure: Devoid of any internal moral value, it ventures into a sort of ethical overinflation where it can

mean everything and nothing.

Jonas, on the other hand, promotes responsibility as an ethical principle stating that we are *obligated* to act in such a way that our actions ensure the continuing existence of human life on earth – which also implies intervening when the risk occurs of endangering humanity in any way and in any prospect of time. This imperative – understood not in the Kantian sense of a principle it would be logically contradictory to go against, but as a kind of “axiom” (Jonas 2003:36), a necessary point of departure of ethics given the situation we are in, with our stage of technological development giving us an unprecedented power to change and control our environment – thus imposes on us a duty to take all measures in answering the “call” of the fragile other (fragility may be considered the primary object of responsibility in Jonas) to take care of it, and to protect it against every possible risk regarding its further existence.

An obvious challenge related to this approach is that the imperative of absolute precaution requires substantial knowledge about possible risks and benefits, knowledge that is often lacking in the complex field of technological innovation. Still, the imperative has a strong appeal, both intellectually (it seems unreasonable to put the future of humanity to risk) and to our feelings of solidarity, compassion with others and so on. The EU-poll referred to earlier, for example, indicates a strong positive evaluation of precautionary concerns in the European population. (At the same time, it is striking that many Europeans seem *not* to support responsible research and innovation. Here I can only speculate, but one might suspect that some of the skepticism can be due to the dubious status of ethics in this context.) Furthermore, the precautionary principle has a strong position in EU laws on matters of science and technology. However, the precautionary principle as applied in this way is attached to specific areas of innovation. It does not, for example, have a strong position within the field of economic innovation. But does the answer lie in the Jonasian universalization or totalization of responsibility? As Paul Ricoeur (2000:32) observes, the final consequence of a totalization of responsibility may be unlimited responsibility of everyone for everything. Paradoxically, this might lead to a kind of fatalism making the subject truly responsible only for his intentions, not for his actions (I will return to this dilemma, identified by Hegel in his *Philosophy of Right*, shortly).

Jonas, however, puts a limit to responsibility by way of his principle that knowledge together with power implies responsibility: We are responsible to the degree that we have sufficient

knowledge about the actualities and possibilities of the situation coupled with the power to do something about it. (Or applied differently: Knowledge about the destructive powers of our actions obligates us not to proceed with the action.)

This also gives us a clue as to how the *subject* of (*ex ante*) responsibility can be conceived of. The subject of this form of responsibility is none other than the subject (individual or collective) possessing knowledge about the risks involved in an action and the powers to either proceed with it or to restrain it.

Responsibility and beyond: A passage to Hegel

We have seen that the shortcomings of the *ex post* concept of responsibility in the context of innovation requires that we conceive of new ways of thinking about responsibility. For responsibility to work as a first virtue of innovation, as seems to be the ambition of the EU with its programmatic stress on responsible research and innovation, we need a theory of responsibility that clarifies how responsibility can be evoked as an ethical principle before the harmful actions and events take place, while at the same time maintaining the idea of a responsible subject.

I have suggested that this concept of responsibility can be developed on the background of a critical rehabilitation of the Jonasian concept of responsibility: A concept that, through its coupling of the imperative of precaution with the idea of the fragile other as the object of responsibility, speaks both to our intellects and to our moral feelings of solidarity with and compassion for others – and further provides us with an idea of the subject of responsibility as the subject capable of causing harm.

As a final (but none the less, on my behalf, experimental) move, I suggest taking a step beyond the Jonasian ethics, with its Kantian influence, towards the Hegelian concept of *Sittlichkeit*. In the *Philosophy of Right*, Hegel places responsibility – in the Kantian sense as the ability of an agent to recognize itself as the author of the act causing the events in

question – under the category of *Moralität*, i.e. the individual's capacity to make moral decisions or to draw up a moral vision of the world in a kind of social vacuum. It is here that Hegel identifies the "paradox" or dilemma referred to above: Action involves the externalization of subjective will (intention), which means that when one acts in the world, the will, or rather the action constituting the will, is instantly subjected to external necessity, and is therefore brought out of the control of the subject (meaning the subjective will). Eventually we get what might be called "side effects" of the action: effects neither intended nor foreseen by the subject. These effects might be catastrophic, but since they are unintended and unforeseen, they are outside the realm of the subjective will, which means that it cannot recognize itself as their true author and therefore cannot accept being held responsible^[2] for them (Hegel 1986:203–292). Still, someone must be held responsible; and who could that be other than the subject initiating the action? But if we really were to hold it responsible, this would mean burdening it with an unlimited responsibility, making it impossible to act at all. This dilemma of the side effects of actions is especially pertinent in the case of technological innovation, where the actual catastrophic effects produced by the use of the technology seldom coincide, and can even be in direct conflict, with the intentions of its innovators. For Hegel, this dilemma cannot be solved within the framework of subjective morality (*Moralität*), because it is produced by the fact that this framework isolates the subjective will from the (social/objective) sphere in which its actions take place. This in turn necessitates the move from *Moralität* to *Sittlichkeit*.^[3]

I cannot give a detailed account of Hegel's theory of *Sittlichkeit* here, but the point is that responsibility is moved from the sphere of subjective morality to the objective sphere of society. In *Sittlichkeit*, subjective will is dialectically sublated by the logic of the collective. This means that there are no purely subjective actions, and therefore no purely subjective form of responsibility: The subject is always already a part of (or embedded in) the social sphere – with its objectivity in the form of mores, customs, shared beliefs, laws, institutions and so on – as are its actions and their effects. The actions and their effects are thereby subjected to a necessity not purely external, but constituted by the subject as an integral part of this sphere of necessity, which at the ontological level of *Sittlichkeit* is the sphere of sociality (which is also the sphere of true freedom, in the Hegelian sense of freedom as constituted by the social/moral-psychological process of recognition, *Anerkennung*).

This may appear overly abstract, but in fact it is meant to show how morality changes from

being expressed in an abstract "ought" (*Sollen*), drawn from the subject's relating to itself and its idea of the good, to being played out in the concrete social setting of society (*bürgerliche Gesellschaft*). From the Hegelian perspective, what is abstract is rather any theory of moral responsibility presenting it as something that can be ascribed to an agent principally existing outside the social sphere of morality – as would be the case in the RRI scheme proposed by von Schomberg, that is based on the metaphysical idea of an agent acting on the world of ethics and morality from the outside. Hegel, on the other hand, places the individual in the social sphere from the beginning, realizing that any concept of a subjective will acting on the world from the outside – or rather, from the inside of its isolated existence – is bound to be contradictory, or at least insufficient, and that the attempts to ascribe responsibility for an action and its effects is absurd within this framework. Instead, we must start from the observation of the subject unfolding in the social sphere, and, according to Hegel at least, base our conception on the dialectic of subjective and objective morality.

In short, the reason why I want to explore this path in my investigations into the concept of responsibility is that I see a clear need to make a move from abstract ethics – with its dualism of the subject of intentions and the external sphere of ethical principles – to a concrete social morality. I would like to see where it leads to when, on the basis of a critique of responsibility in RRI-schemes and moral philosophies of responsibility such as Jonas's, we pay renewed attention to Hegel's social theory and philosophy of moral agency; not accepting the whole theory at face value, but concentrating on the sublation of classical (*ex post*) responsibility into the ethico-institutional realm of *Sittlichkeit* – a realm where no action is considered purely subjective, but is instead conceived of as the subject's constitutive self-investing in the (social) world of institutions, laws, mores, customs, shared beliefs etc. – making the consequences of its actions, whether intended or unintended, future or present, intelligible as moral aspects of the actions themselves.

Concluding remarks: Hegel with Jonas?

An important question will be to what extent the *ex ante* responsible subject can be

incorporated into the Hegelian ethics of *Sittlichkeit*. How would the responsible subject of Jonas's ethics, for example - making individual moral decisions on the basis of a strict imperative of precaution - fit in with Hegel's basic thesis of the fundamental social character of moral agency? I suspect that it would be difficult, if not impossible, to make Jonas's responsible subject fit seamlessly into the Hegelian scheme of ethical life. Jonas's subject does not make its moral decisions in a social vacuum - here he is on a par with Hegel - but the Jonasian "call of the other" does not seem to be conditioned by any form of *Sittlichkeit*. Rather, this call seems to be the (ontological) source of moral responsibility itself (Jonas 2003, especially pp. 234-240.). Thus, the subject's morality is not necessarily linked to its belonging to an existing social order, but rather has its origin in singular encounters with fragile others (archetypically exemplified by the encounter with the extreme fragility of the newborn child).

A theoretical affiliation can, however - with some adjustments on both sides - be construed regarding one crucial matter at stake here: the matter of the side effects of actions. Trying to solve the (Hegelian) problem of the side effects of actions, Jonas evokes the idea of (what Ricoeur calls) the succession of generations (Ricoeur 2000:31). The impacts - intended or unintended - of an action (*Tat*, in Hegel) on future generations are linked to the action - the *Handlung* manifested by the externalization of subjective will in Hegel's theory - and thereby to our subjectivity by the way the action (*Tat*) is carried on into the future by the succession of generations. In other words: The action (*Handlung* as well as *Tat*) unfolds, through its effects, as part of an inter-generational chain of being. This makes responsibility a matter more of ethical life as a whole - i.e. as a system irreducible to its (at any time T) existing parts - than of a subjective will acting upon the matter of ethics and morality from the supra-moral standpoint of the "rational agent" or the like - which, as implied above, can be said to be the metaphysical starting point of von Schomberg's theory of the (co-)responsible subject; a subject that seems stripped bare of any moral capacity besides that of the (displaced?) intention or interest to act according to external ethical principles.

Whether this theoretical affiliation stands the test of philosophical scrutinizing remains to be seen. In any case, the demands for responsibility in our time make urgent the need for further reflection on and clarification of the meaning of the concept - not least if it is to continue functioning as a "first virtue" in the ethics of technology and innovation.

References

Birnbacher, D. (2001): "Philosophical Foundations of Responsibility". In *Responsibility: The Many Faces of a Social Phenomenon*. Auhagen, A.E. et al. (eds.). London/NY: Routledge.

European Commission (2013): "Press release 14 November 2013: Eu-wide poll shows public support for responsible research and innovation." http://europa.eu/rapid/press-release_IP-13-1075_en.htm (accessed 25 November 2013).

— (2013): *Options for Strengthening Responsible Research and Innovation. Report of the Expert Group on the State of Art in Europe on Responsible Research and Innovation*. http://ec.europa.eu/research/science-society/document_library/pdf_06/options-for-strengthening_en.pdf (accessed 21 July 2014).

Hegel, G.W.F (1986): *Grundlinien der Philosophie des Rechts*. Frankfurt am Main: Suhrkamp Verlag.

Jonas, H. (2003): *Das Prinzip Verantwortung. Versuch einer Ethik für die technologische Zivilisation*. Frankfurt am Main: Suhrkamp Verlag.

Rawls, John (1971): *A Theory of Justice*. Cambridge, MA: Harvard UP.

Ricoeur, Paul (2000): *The Just*. Chicago: The University of Chicago Press.

Schomberg, R.v. (2007): *From the Ethics of Technology towards an Ethics of Knowledge Policy and Knowledge Assessment*. A working document from the European Commission Services, Directorate General for Research, Brussels.

— (2011): "Prospects for technology assessment in a framework of responsible research and innovation". In: Dusseldorp, M., and Beecroft, R. (eds) *Technikfolgen abschätzen lehren. Bildungspotenziale transdisziplinäre Methoden*, pp. 39–61. Wiesbaden: Vs Verlag.

[1] There are some prominent exceptions, notably: 1) *obligations of means* (as opposed to obligations of result, cf. in particular French civil and (presumably) penal codes, 2) *state responsibility* (international law, especially human rights obligations), and 3) *duty of care* (Anglophone common law, especially English). While responsibility in these cases are revealed in law courts ex post, i.e. in case of a breach of them, they are present in the law in the prospective sense as duties whose fulfillment is expected of the relevant agents (private individuals, corporations or organizations, states). I am indebted to Professor Giorgio Baruchello (University of Akureyri) for this remark.

[2] Hegel does not use the contemporary German word for responsibility, *Verantwortung*, but uses related terms such as *zurechnen* (ascribe) (1986:218) and *imbutabel* (1986:212) where we today would speak of responsibility/*Verantwortung*.

[3] That is, it does not alone necessitate this move, but is one of the "contradictions" of subjective morality necessitating it.