Research Policy 42 (2013) 1568–1580

Contents lists available at ScienceDirect

Research Policy

journal homepage: www.elsevier.com/locate/respol

Developing a framework for responsible innovation☆

Jack Stilgoe[a,*], Richard Owen[b,1], Phil Macnaghten[c,d]

[a] University of Exeter Business School/Department of Science and Technology Studies, University College London, Gower Street, London WC1E 6BT, UK
[b] University of Exeter Business School, Rennes Drive, Exeter EX4 4PU, UK
[c] Department of Geography, Science Laboratories, Durham University, South Road, Durham DH1 3LE, UK
[d] Department of Science and Technology Policy, Institute of Geosciences, P.O. Box 6152, State University of Campinas – UNICAMP, 13083-970 Campinas, SP, Brazil

ARTICLE INFO

Article history:
Received 16 July 2012
Received in revised form 7 May 2013
Accepted 17 May 2013
Available online 13 June 2013

Keywords:
Responsible innovation
Governance
Emerging technologies
Ethics
Geoengineering

ABSTRACT

The governance of emerging science and innovation is a major challenge for contemporary democracies. In this paper we present a framework for understanding and supporting efforts aimed at 'responsible innovation'. The framework was developed in part through work with one of the first major research projects in the controversial area of geoengineering, funded by the UK Research Councils. We describe this case study, and how this became a location to articulate and explore four integrated dimensions of responsible innovation: anticipation, reflexivity, inclusion and responsiveness. Although the framework for responsible innovation was designed for use by the UK Research Councils and the scientific communities they support, we argue that it has more general application and relevance.

1. Introduction

1.1. Responsibility, science and innovation

Responsible innovation is an idea that is both old and new. Responsibility has always been an important theme of research and innovation practice, although how it has been framed has varied with time and place. Francis Bacon's imperative to support science 'for the relief of man's estate', the institutionalisation and professionalisation of science from the 17th century onwards, Vannevar Bush's (1945) 'Endless Frontier', JD Bernal's (1939) arguments for science in the service of society and Michael Polanyi's (1962) 'Republic of Science' counter-argument have all contained particular notions of responsibility.

Science has been conventionally invoked by policy as emancipatory. This has allowed scientists and innovators considerable freedom from political accountability. From this perspective, the role responsibilities of scientists – to produce reliable knowledge – and their wider moral responsibilities to society are imagined to be conflicted. The perceived high value of knowledge to society means that such role responsibilities typically trump any wider social or moral obligations (Douglas, 2003). Although frequent objections from university scientists suggest a permanent assault on their autonomy, much of the constitution of Polanyi's (1962) self-governing 'Republic of Science' survives to this day.

In the second half of the 20th century, as science and innovation have become increasingly intertwined and formalised within research policy (Kearnes and Wienroth, 2011), and as the power of technology to produce both benefit and harm has become clearer, debates concerning responsibility have broadened (Jonas, 1984; Collingridge, 1980; Beck, 1992; Groves, 2006). We have seen recognition and negotiation of the responsibilities of scientists beyond those associated with their professional roles (e.g. Douglas, 2003; Mitcham, 2003). We have seen scientists' own ideas of 'research integrity' change in response to societal concerns (Mitcham, 2003; Steneck, 2006). In the 1970s, biologists in the nascent field of recombinant DNA research sought to 'take responsibility' for the possible hazards their research might unleash, with a meeting at Asilomar in 1975 and a subsequent moratorium.[2] Concerns about the 'dual use' of emerging technologies and the limits of self-regulation, visible in physicists' agonising about nuclear fission prior to the Manhattan project (Weart, 1976), resurfaced in 2012 with the recent controversy over the publishing of potentially

* Corresponding author. Tel.: +44 020 7679 7197.
E-mail addresses: j.stilgoe@ucl.ac.uk, jackstilgoe@gmail.com (J. Stilgoe), r.j.owen@exeter.ac.uk (R. Owen), p.m.macnaghten@durham.ac.uk (P. Macnaghten).
[1] Tel.: +44 01392 723458.

[2] We should point out that this meeting was criticised, both at the time (Rogers, 1975) and in later scholarship (Wright, 2001; Nelkin, 2001) as being motivated by an attempt to escape top-down regulation rather than to 'take responsibility'

0048-7333/$ – see front matter © 2013 The Authors. Published by Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.respol.2013.05.008

dangerous research on flu viruses (Kaiser and Moreno, 2012). The negotiation of responsibility between practicing scientists, innovators and the outside world remains an important and contested area of debate to this day.

Research in Science and Technology Studies (STS) suggests that conceptions of responsibility should build on the understanding that science and technology are not only technically but also socially and politically constituted (e.g. Winner, 1977). Latour (2008) suggests that science does not straightforwardly reveal reality through techniques of simplification and purification aimed at further mastery. As Callon et al. (2009) point out, science and technology can, paradoxically, add to our sense of uncertainty and ignorance. They tend to produce a "*continuous* movement toward a greater and greater level of *attachments* of things and people at an ever expanding scale and at an ever increasing degree of intimacy" (Latour, 2008, p. 4, italics in original). These observations suggest that unforeseen impacts – potentially harmful, potentially transformative – will be not just possible but probable (Hacking, 1986).

Responsibility in governance has historically been concerned with the 'products' of science and innovation, particularly impacts that are later found to be unacceptable or harmful to society or the environment. Recognition of the limitations of governance by market choice has led to the progressive introduction of post hoc, and often risk-based regulation. This has created a well-established division of labour that reflects a consequentialist framing of responsibility, as accountability or liability (Pellizzoni, 2004; Grinbaum and Groves, 2013). With innovation, the past and present however do not provide a reasonable guide to the future (Adam and Groves, 2011), so such retrospective accounts of responsibility are inherently limited. We face a dilemma of control (Collingridge, 1980), in that we lack the evidence on which to govern technologies before pathologies of path dependency (David, 2001), technological lock-in (Arthur, 1989), 'entrenchment' (Collingridge, 1980) and closure (Stirling, 2007) set in.

We have (pre-)cautionary tales of risks whose effects did not materialise for many years, where potential threats were foreseen but ignored or where only certain risks were considered relevant (Hoffmann-Riem and Wynne, 2002; EEA, 2001, 2013). Governance processes, often premised on formal risk assessment, have done little to identify in advance many of the most profound impacts that we have experienced through innovation, with the 2008 financial crisis being the most disruptive recent example (Muniesa and Lenglet, 2013). Bioethics, another major governance response, has drawn criticism for privileging individual ethical values such as autonomy over those such as solidarity that might lead to a genuine 'public ethics' (Nuffield Council on Bioethics, 2012; also Prainsack and Buyx, 2012) and, in its consequentialist version, serving to bolster the narrow instrumental expectations of innovators in some areas (Hedgecoe, 2010).

Callon et al. (2009) use the metaphor of science and technology 'overflowing' the boundaries of existing scientific regulatory institutional frameworks. They point to the need for new 'hybrid forums' that will help our democracies to be "enriched, expanded, extended and... more able to absorb the debates and controversies surrounding science and technology" (Callon et al., 2009, p. 9). Such controversies have demonstrated that public concerns cannot be reduced to questions of risk, but rather encompass a range of concerns relating to the purposes and motivations of research (Grove-White et al., 2000; Wynne, 2002; Grove-White et al., 1997; Macnaghten and Szerszynski, 2013; Stilgoe, 2011), joining a stream of policy debate about the directions of innovation (Smith et al., 2005; Stirling, 2008; Morlacchi and Martin, 2009; Fisher et al., 2006; Flanagan et al., 2011). Yet, despite efforts at enlarging participation (see, for example, RCEP, 1998; House of Lords, 2000; Wilsdon and Willis, 2004) current forms of regulatory governance offer little scope for broad ethical reflection on the purposes of science or innovation.

1.2. A new scientific governance?

One alternative to a consequentialist model of responsibility has been to succumb to moral luck (Williams, 1981), to hope that an appeal to unpredictability and an inability to 'reasonably foresee' will allow us to escape moral accountability for our actions. Dissatisfaction with both this approach and risk-based regulation has moved attention away from accountability, liability and evidence towards those future-oriented dimensions of responsibility – care and responsiveness – that offer greater potential to accommodate uncertainty and allow reflection on purposes and values (Jonas, 1984; Richardson, 1999; Pellizzoni, 2004; Groves, 2006; Adam and Groves, 2011).

Emerging technologies typically fall into what Hajer (2003) calls an 'institutional void'. There are few agreed structures or rules that govern them. They are therefore emblematic of the move from old models of governing to more decentralised and open-ended governance, which takes place in new places – markets, networks and partnerships as well as conventional policy and politics (Hajer and Wagenaar, 2003).

A number of multi-level, non-regulatory forms of science and innovation governance have taken this forward-looking view of responsibility, building on insights from STS that highlight the social and political choices that stabilise particular innovations (Williams and Edge, 1996; Pinch and Bijker, 1984; Winner, 1986). New models of anticipatory governance (Barben et al., 2008; Karinen and Guston, 2010) Constructive, Real-Time and other forms of technology assessment (Rip et al., 1995; Guston and Sarewitz, 2002; Grin and Grunwald, 2000), upstream engagement (Wynne, 2002; Wilsdon and Willis, 2004), value-sensitive design (Friedman, 1996; van den Hoven et al., 2012) and socio-technical integration (Fisher et al., 2006; Schuurbiers, 2011) have emerged. These have been complemented by policy instruments such as normative codes of conduct (see, for example, European Commission, 2008), standards, certifications and accreditations, running alongside expert reports, technology assessments and strategic roadmaps. Such initiatives have, to varying degrees, attempted to introduce broader ethical reflection into the scientific and innovation process, breaking the existing moral division of labour described above. They have attempted to open up science and innovation (Stirling, 2008) to a wider range of inputs, notably through the creation of new spaces of 'public dialogue' (Irwin, 2006).

The other important aspect of a forward-looking view of responsibility in science and innovation is that it is shared (Richardson, 1999; Mitcham, 2003; Von Schomberg, 2007). The unpredictability of innovation is inherently linked to its collective nature. Following Callon's account of innovation as 'society in the making' (Callon, 1987), we can see that implications are 'systemic', coming from the interplay of the technical and the social (Hellström, 2003). This suggests that scientists, research funders, innovators and others have a collective political responsibility (Grinbaum and Groves, 2013) or co-responsibility (Mitcham, 2003). This reflects understanding that while actors may not individually be irresponsible people, it is the often complex and coupled *systems* of science and innovation that create what Ulrich Beck (2000) calls 'organised irresponsibility'.[3] We can point to 'second-order' (Illies and Meijers, 2009) or 'meta-task' responsibilities (van den Hoven, 1998; van den Hoven et al.,

[3] von Schomberg (2013) suggests four categories of irresponsible innovation that typically manifest: Technology push, Neglect of ethical principles, Policy Pull and Lack of precaution and foresight.