(c) 2014. This is the author's version of the work. It is posted here by permission of ACM for your personal use. Not for redistribution. The definitive version was published in Proc. of *CHI 2014*, April 26 – May 01 2014, Toronto, ON, Canada. http://dx.doi.org/10.1145/2556288.2557364.

Towards a Closer Dialogue between Policy and Practice: Responsible Design in HCI

Barbara Grimpe
Dept. of Computer Science
University of Oxford
barbara.grimpe@cs.ox.ac.uk

Mark Hartswood
Dept. of Computer Science
University of Oxford
mark.hartswood@cs.ox.ac.uk

Marina Jirotka
Dept. of Computer Science
University of Oxford
marina.jirotka@cs.ox.ac.uk

ABSTRACT
Given the potent and pervasive nature of modern technologies, this paper lays out the complexities involved in achieving responsible design. In order to do this we will first compare an emerging policy-oriented programme of research known as RRI (Responsible Research and Innovation) with initiatives in HCI. A focus on the similarities and differences may highlight to what extent responsibility is already and successfully embedded within the concerns and practices of design and use, and what may yet need to be incorporated for responsible design. The paper then discusses the challenges of 'naturalising' the very ambitious programme of RRI within specific design activities and concerns, through the lens of four analytic concepts: reflexivity; responsiveness; inclusion; and anticipation. Finally, we make a case for a pragmatic, 'unromantic', but engaged reinterpretation of RRI for HCI.

Author Keywords
Responsible design; ethics; innovation; risk society; governance; participatory design; user-centered design; critical design; value-sensitive design

ACM Classification Keywords
K.4.1 Computers and Society, Public Policy Issues (Ethics); K.4.2 Computers and Society, Social Issues; K.7.4 The Computing Profession, Professional Ethics; J.4 Social and Behavioral Sciences, Sociology

INTRODUCTION
Responsible design practice has been a recurrent theme of research in HCI for many years and resonates with a broader movement to influence policy in favour of inculcating greater responsibility within scientific and innovation processes for their outcomes. A programme of work related to this concern is Responsible Research and Innovation (RRI), a policy-oriented research agenda motivated by wider societal concerns of actual or alleged hazards in a number of domains including nanotechnology, geoengineering and synthetic biology [e.g. 45]. Now this research agenda is extending into ICT [6]. While RRI is partly supported by European funding agencies [50], the debate has also been important in the US, although often appearing under different headings such as, "responsible development", "responsible governance" or "broader impacts" [37, 40]. In addition, research in RRI is currently underway to extend the agenda world-wide [36].

In this paper we compare the RRI agenda to extant practices in HCI research and design. To our knowledge this is the first comparative discussion ever across these two broad fields. The fact that RRI, like design, is far from uniform and rather an ongoing debate with many voices complicates the picture. However, the following basic ideas may be regarded as crucial. First, apart from the prevention of any harm occurring to society or nature through their innovations, researchers and innovators are required to proactively seek positive social and environmental benefits. Second, users and consumers should have a voice in all stages of a given research and innovation process. This is quite a challenge: in practice, much HCI work involves these groups only at certain stages, for example, in the first phases of design or in the final showcase event. Third, beyond accounting for such immediate user or consumer needs, RRI proponents envisage taking into account the views of the public and society at large, i.e. any direct or indirect stakeholders [45, 50]. In the age of globalization, ubiquitous computing and the Internet of Things, this far-reaching requirement will become more of a necessity but be difficult to realize in many HCI projects. Fourth, the RRI agenda questions familiar distinctions such as, the traditional "moral division of labour" between researchers/innovators and the wider public, or regulatory bodies [44, p. 1569]. Regulation would often focus on accountability and liability 'after the fact' ("consequentialist model of responsibility") and, hence miss most of the actual sources of irresponsible outcomes [44, p. 1569]. Instead, many RRI proponents wish to establish a more future-oriented responsibility as well as "decentralized and open-ended governance" right at the heart of research and innovation processes [44, p. 1569].

While the RRI agenda is evidently well motivated, it is also quite comprehensive and may actually be intrusive by definition. Thus, it runs the risk of being regarded as a 'political' top-down approach that researchers, innovators and other stakeholders in HCI cannot adopt easily, or are understandably not willing to adopt. For instance, many

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than the author(s) must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from Permissions@acm.org.
CHI 2014, April 26 - May 01 2014, Toronto, ON, Canada.
Copyright is held by the owner/author(s). Publication rights licensed to ACM.
ACM 978-1-4503-2473-1/14/04…$15.00.
http://dx.doi.org/10.1145/2556288.2557364.

researchers and practitioners already face various responsibilities in their daily lives and work, including numerous domain-specific requirements, and may view extra activities resulting from this agenda as quite onerous.

Moreover, the identification of design as *ir*responsible is not straight-forward. Fundamental uncertainty [1, p. 979] and conflicts in values are just two reasons. For instance, various "desirable" qualities of 3-D printers are crucial for producing guns, which is of doubtful value [39, p. 80]. RRI proponents acknowledge these ambiguities, at least to some degree. While von Schomberg [50] discusses various cases of supposed "irresponsible innovation" such as, an electronic patient record system being abandoned due to privacy issues, others elaborate on the difficulties of making such judgments. For instance, "irresponsibility" is often an emergent property at the system level. It may involve the creation by many actors, but is rarely clearly attributable to them one by one [32].

Similarly, in this paper we focus on an entire ecosystem of actors across time and space and try to avoid adjudicating behavior. From standpoint theory we take the position that any discussion of (ir)responsible design should not aim at a supposedly infallible "view from above", but be a (self-) critical collection of "views from somewhere" [15, p. 196]. So we wish to lay out what may be required to 'naturalise' the ambitious programme of RRI within specific design activities and concerns in HCI, and we also hope to provide some elements to 'talk back' to the RRI discourse where necessary.

METHODOLOGY

We analysed the RRI programme by reviewing all articles known that deal explicitly with RRI (and related US and other approaches). Stilgoe et al. [44] have a long track record of RRI publications and also provide the most comprehensive and up-to-date review of existing approaches. They discuss four key concepts: reflexivity; responsiveness; inclusion; and anticipation. These have not yet been tailored to HCI concerns, hence our paper attempts to provide a domain-specific application. We sought out papers on design framing ethical issues and, for the purposes of this paper, selected those that already resonate with RRI ideas to some degree; or that, instead, can inform policy discourse; and/or those that address 'sensitive' areas such as, ICT in health care. Instead of a broad survey, we favour discussing a few pertinent papers in greater detail.

We develop our discussion along three lines of enquiry. *Perspectives* captures complex issues such as, designer versus user and other stakeholders' perspectives. *Scope* addresses challenges in terms of the scale of HCI design. This may be for example, problem-solving in local situations versus larger organizational and institutional structures that designers and other participants may need to respond to. The third challenge is *temporality*. For instance, much of HCI design does not only contribute to, but also faces ongoing change in technological development and user contexts. Studying design issues through these three analytical foci makes clear that new responsibilities are difficult to naturalise in practice, but also, and more importantly, *how* this is the case.

RETROSPECTIVE REVIEW OF RESPONSIBLE PRACTICES IN DESIGN

In HCI and related fields, researchers already orient to 'responsible design' across at least four (partly overlapping) dimensions. The first dimension concerns use of reflective practice in order to avoid parochial tendencies of design. Baumer et al. are concerned for design to be aware of the broader systemic context of the 'problem' that is to be addressed, and to be self-critical of design's propensity to implement sophisticated technological 'solutions' when perhaps a low-tech, or no-tech approach could be most effective or appropriate [2]. Sengers et al. on the other hand look to critical reflection as a means to make visible and subvert dominant metaphors and values that otherwise might be silently absorbed into the design process [42]. The next section on "reflexivity" follows up on these aspects.

A second dimension is concerned with participation as a channel for bringing ideals of social responsibility into design, and is most visible within the Participatory Design (PD) movement, especially the social democratic underpinning of the Scandinavian tradition of PD [11]. As we discuss further in the section on "inclusion", participatory approaches resonate strongly with RRI's agenda which would arguably seek to incorporate a more thorough democratisation of design than even PD.

A third dimension for responsible design includes those approaches that highlight human values in design, with value-sensitive design being the most notable [8].

Sellen et al [41] locate the need for a values-aware approach within a series of contemporary techno-cultural trends that each underscore how technologies are increasingly woven into fundamental social fabrics. This leads us to a fourth dimension for responsible design – the myriad studies that document the social transformations wrought by our increasingly technologised ways of life that spell out unforeseen consequences and reverberations of modern digital tools and services. To give but a few examples: technologies that share location can become ones of accountability when partners demand to know why it has been switched off [33]; the algorithms underpinning internet search can come to embed negatively regarded views such as racial biases or prejudices in ways that are hard to determine [46]; practices such as 'friending' and 'unfriending' on Facebook can transform the moral responsibilities of friendship [20]. Knowledge of these sorts of effects adds to our sensitivity (and presumably influences our design decisions) when we attend to our projects.

Through the work discussed previously we can trace an expansion of responsibility concerns in design aligned with the increasing potential for far reaching, yet hard to control

impacts. Value-sensitive design's concept of the 'indirect stakeholder' is perhaps a motif for this expansion, signalling the need for engagement with an ever wider circle of entangled people, places and cultures as side-effects of design. Thus, as technology achieves greater potency and reach, then it would seem the designer's conscience needs also to extend to take in the wider knock-on effects of their creations, and to consider consequences across a greater numbers of settings, people and circumstances in which unintended transformations are possible. On the face of it at least, design then would appear to be increasingly in need of RRI's proposed remedies. However, we think that this is a hasty conclusion and propose to compare the two fields in greater detail, in order to provide a more nuanced picture.

THE RRI AGENDA: FOUR KEY CONCEPTS AND DESIGN ISSUES

Reflexivity

According to the RRI agenda there is an increased need for reflective practice and critical reflexion within processes of research and innovation. The backdrop for this requirement is Beck's diagnosis of the hypermodern "risk society" [3] characterized by complexity, dynamic change and unintended consequences. This term epitomises the view that society has faced the repercussions of its 'own' earlier techno-scientific advances that got out of hand; and that it tends to cope with these by ever more sophisticated techno-scientific efforts. Consequently society risks being caught in a propelling feedback loop by producing large-scale unpredictable "hybrids" such as climate change [26].

Beck's diagnosis stems from the 1990s but still applies to contemporary society and technologies, as RRI proponents have stressed [37; 44; 34]. Many request a corresponding counter-reflexivity. Lynch has shown, though, that the notion of "reflexivity" has many possible meanings across different disciplines and methodologies [27], and we also find this variety, and occasional vagueness, in RRI approaches. However, following Stilgoe et al. we may at least identify the following features: pausing for a moment; holding "a mirror up to one's own activities, commitments and assumptions"; considering the "limits of knowledge"; being "mindful that a particular framing of an issue may not be universally held" [44, p. 1571]. It is argued that this mindfulness is required from the individual scientist and may be supported by mechanisms such as, codes of conducts. Likewise, institutions (e.g. funding and regulatory agencies) are also asked to develop reflective capacities.

This demand for institutional reflexivity is clearly justified given that many researchers, 'innovators' and designers, including those in HCI, are actually subject to a complex ecosystem of different institutional forces. Accordingly, the principle of reflexivity and the three other RRI principles (responsiveness, inclusion and anticipation) are techniques, tasks or attitudes to be shared and distributed across many actors and levels. One aspect of such a wider institutional reflexivity applies even to the RRI agenda itself. Accompanied by existing governance bodies RRI proponents may indeed try to establish themselves as moral gatekeepers who presumably know responsible innovation best. Such an 'ethics' lined up with existing authorities may be termed *particularly* irresponsible and undesirable.

We now carve out the challenges of realizing such a multi-actor and multi-level reflexivity in practice by help of an example: wearable computing, and wearable cameras in particular [24, 16]. Kelly et al. have developed an "ethical framework" for wearable cameras in health research, which we consider a code of conduct as requested by Stilgoe et al. Such conventions seem necessary: cameras are relatively intrusive and imply many unpredictable situations of potential privacy and confidentiality 'breach' [24]. Such case-based considerations allow us to draw some preliminary lessons that may hold not only for wearable cameras but for other pervasive technologies too.

Perspectives, scope and temporality

If "reflexivity" means bearing in mind the limits of one's own knowledge, then this requirement is, in the case of wearable cameras, built into the approach, at least to some extent. Health data generated this way entails the chance for a paradigmatic shift from a "researcher-centric construction of the social world to that of the participants" [24, p. 315]. Unfortunately, this enhanced sense of perspective in health research comes at a high price, as the camera bearer's own concerns and the perspectives of many indirect stakeholders are to be considered (e.g. when going to the bathroom, or at work). All these stakeholders are likely to have differing views of what should be kept in private, anonymised, or handled confidentially. Since the camera is supposed to capture as much of everyday life as possible the devil is in the detail. For instance, full informed consent or compromises would have to be negotiated with all the family members, colleagues, and people on the street that an individual meets spontaneously when using the camera.

This sounds pessimistic, and one may argue that not all bearers of wearable cameras, and indirectly affected parties, care so much about embarrassment, privacy or confidentiality. At first sight, Harper et al. [16] offer a more optimistic reading of the technology. They investigated the ways in which households used, and made sense of, wearable cameras for recording aspects of their daily life. Reportedly, some of the participants were positively surprised that the devices helped them to discover unnoticed and unvalued aspects of daily routines, and supported them in changing their lives to the better. Thus, in this specific research context the camera apparently did not surface ethical issues but stimulated individual reflexivity in a fruitful way. But the situation is perhaps more complicated when the use of the devices is scaled up.

With so many different stakeholders in a single observation study, scope is obviously an important issue. However, beyond 'counting' all the implicated individuals we would like to draw attention to more systemic levels of society, i.e.

existing organisational and institutional structures that challenge all the involved parties mentioned, pushing them to their limits of individual reflexive capacity, and foregrounding the necessity of a more distributed, "institutional" reflexivity as claimed by Stilgoe et al. Wearable cameras will be used at workplaces that have their own confidentiality rules, such as banks. Or their bearers will pass security systems, such as at airports that follow their own monitoring procedures that may prohibit photography. These are all instances of higher-order existing societal 'regimes' (formal authorities and authoritative procedures) that challenge the capacities of researchers, software/camera designers, and users in thinking through the potential undesirable consequences of using wearable cameras and related data 'in the wild'.

Researchers' and designers' reflexive capacities are challenged even more due to a number of temporal disjunctions. This is probably an aspect that needs to be added to the RRI definition introduced earlier, but it is a challenging one: reflexivity may not only mean to acknowledge the limits of a given state of knowledge, but also its continuous change. In this regard the code of conduct of Kelly et al. is again illuminative [24]. It is argued that the current guidelines may not cover all (future) issues and hence should be revised and adapted when new challenges emerge. Unfortunately, in practice such an ethos means a lot of work. Many of the previously mentioned organisational and institutional 'regimes' feature their own short-term temporal dynamics (e.g. different scheduled and spontaneous meetings at the workplace; time-slots for security control before boarding) that require participants to take situated actions, and change decisions concerning privacy, anonymisation or confidentiality arrangements. In addition, technologies often evolve faster than the legal regimes, and with new technologies new sorts of responsibilities emerge.

All these practical issues in terms of temporality, perspective and scope challenge the individual and institutional reflexive capacities of all who are implicated in the design, research and use of wearable computing and possibly other contemporary technologies. How to reconceptualise the RRI principle of reflexivity under these circumstances? Following up on Lynch we argue that reflexivity is not an absolute virtue, e.g. the antithesis of naivety and ignorance. Instead, the meaning of the term, and the degree of its realization, are "relational" to the "circumscribed problems" and opportunities of a given community of practice [27, p. 41, 46] such as, a department, or a work group funded through a temporary project. We elaborate on this pragmatic take on reflexivity in the conclusions, as it mirrors a great deal of our understanding of the entire RRI agenda in relation to HCI.

Responsibility and responsiveness

This section considers the RRI concept of responsiveness in the context of innovative technologies of care. Caring for an elder person with multiple diseases or ailments requires knowing, and 'making use' of various interdependent responsibilities [48, p. 19]. A dedicated and concerned relative, a doctor or paid care worker (i.e. virtue responsibility and/or role responsibility) who are formally trained or have informally learned to interpret combined illnesses (capacity responsibility) will try to judge the effects of medication as far as possible (causal responsibility). If they have allegedly failed to do so in an act of negligence, and the person dies prematurely, they may be sued legally (outcome and liability responsibility). However, when auxiliary devices such as a wheelchair, or sensor technologies, are used (e.g. fall detectors, PDAs sampling physiological parameters such as, heart rate) the situation becomes more complex. Given these first rudimentary components of an agent-based health care system, the causal responsibility may fully or partially be imputed to these agents, and hence their producers too.

At this point 'responsibility' becomes obviously an issue for system designers, and hence their need to be exempted from liability. Indeed, the "spheres of risk" expand in today's society so searching for a responsible person or organization who indemnifies others for a loss becomes increasingly important. However, holding companies accountable "for any future consequence of their present choices" may be financially unsustainable [34, p. 553].

Nevertheless, RRI proponents argue that there is still room for "forward-looking" responsibility, such as, "responsiveness" [48, p. 17; 44, p. 1572; 34]. What does "responsiveness" mean from a conceptual point of view, and what does it imply for HCI design in practice?

First of all, it literally means to respond, i.e. to react and answer. This attitude, in turn, entails "previous listening", and "openness, a willingness to understand and confront the other's commitments and concerns with ours" [34, p. 546, 557]. Thus, responsiveness has clearly a situational and interactional dimension: it requires heedfulness in face-to-face interactions. Second, responsiveness includes sensitivity for new knowledge and perspectives that are not known in advance but *emerge* in ongoing interactions [44, p. 1572]. Third, similar to the institutional reflexivity discussed in the last section, RRI proponents see the need of a broader 'institutional' responsiveness. So far, this term has been scarcely defined, though. For Pellizzoni, it refers to institutions such as, the EU Commission or the "Corporate Social Responsibility" initiative. Such institutions are considered as merely "reactive", not responsive, when they "attempt to avoid any substantial change in their agenda, structure and identity" [34, p. 558].

Let us apply some of these aspects to the example introduced earlier. In care for elder people, presence and movement sensors provide data about positions and daily activities such as sleep, with the aim of estimating the health state of an individual. On an aggregate level, these sensors seem to allow for the identification of different

classes of people, i.e. "healthy" or "normal people" and others such as, "nocturnally overactive" ones [43, p. 448-450]. Moreover, it is argued that a combination of sensors may provide "information of a higher semantic level" such as, that the patient is eating [25, p. 4]. However, both this deduction ('the patient is eating') and the more aggregate classifications for entire populations ("normal" and "healthy" versus "over" or "underactive") can be questioned. For instance, a sensor on a box of medication may assure that the box was *accessed* [31], but it cannot tell whether the medication was actually taken, or how much. In a similar way, movement and presence sensors can indeed only deliver information about physical movement or presence, but nothing about the actual well-being of the person, let alone his or her feelings. Thus, the "availability of instant or continuous signs" about the elder person "means that care *appears* proximal and continuous" [31, p. 153; emphasis added], but individual and local idiosyncrasies (habits etc.) and the emotional state of the person may be completely overlooked.

Sensor-based telecare architectures entail promises such as, shorter reaction times in case of incidents and cost-efficiency. However, the issues just raised suggest that technological innovation in this area may also need to include *responsiveness*. Generally it may be argued that any new design effort should be geared to "person-centered care" [18]. In practice, this vision may be translated into an ongoing situated engagement referred to as "bricolage": the "pragmatic customization" of technologies, including 'old' ones such as telephones still being used, *with* the elder person and the carers [10]. Thus, responsiveness would mean the provision of flexible components to 'think with things'" – not once and for all, but always anew, for example, when one component breaks down [10]. This focus on emerging issues is how the concept of (flexible and situated) responsiveness is distinguished from the common-sense understanding of responsibility as a rather fixed state or attitude. In terms of the sensor technologies discussed previously this conceptual shift may mean: *first* investigating the living conditions and habits of elder people in their individual homes, and then designing (and installing, testing, redesigning etc.) appropriate sensors. While this approach is part of user-centered design, reconsidering it through the conceptual lens of responsiveness transforms it from an on-off binary matter into an *ongoing* interactive process throughout the (remaining) life-span of the main users and stakeholders.

Perspectives, scope and temporality

At its core, responsiveness is a situated, interactional accomplishment, so it has a very dynamic temporal structure including many short-term feedback loops over a (hopefully) longer time span. But it is easy to argue theoretically for ongoing responsiveness to different stakeholder needs whereas, in practice, some of these may be extremely difficult to grasp and to be sympathetic with – especially if they are accompanied by 'bewildering' reactions such as, becoming restless or aggressive, as seems to be true for many dementia patients. In the future, more and more of us will live longer, and more and more carers will have to work with elder people with reduced or transformed cognitive capacities – and hence find new ways of perceiving their social world, and expressing their needs. It may be argued that understanding stakeholders' perspectives in this and other domains (for example, developing learning devices for children), challenges researchers and designers alike. There is an increase in literature, documentaries and arts projects that try to familiarize a broader audience with the social worlds as perceived by elder people with cognitive disabilities, and as perceived by those who work or live with them [e.g. 22]. Designers' sense of responsiveness may be enhanced by such more creative, less strictly scientific insights.

As suggested above, a crucial question concerns how responsiveness copes with scale. An important driver of healthcare technology innovation is to realize economies of scale in a climate of increasing demand and constrained resources. This, however, strains designs' ability to be individually responsive and the problems of 'scaling-up' from telehealth pilot to national provision are becoming evident. These problems often arise because of the difficulties of maintaining the levels of responsiveness that were possible within the context of a local trial [47]. One responsibility incumbent on design would perhaps be an awareness of the implications of the effects of use at scale, where the transformative implications can be amplified and the technology characteristics may appear very different to the end users. However, similar to the argument of "institutional reflexivity" we need to appeal to a broader ecosystem of regulatory bodies and funding institutions to find their ways of "institutional responsiveness".

Inclusion

Similar to responsiveness being ideally an incessant interactive process and not just a one-off effort of user-centered design, many RRI proponents envision "inclusion" as nothing less than the "ongoing involvement of society" at all stages of any research or innovation process – without "wasting their time" and taxpayers' money [45, p. 12]. Achieving this 'efficiency' in terms of time and money is quite a challenge in practice.

Such practical difficulties with inclusion are well-known in RRI. For instance, some authors have observed a "proliferation of participatory approaches" such as, "consensus conferences", "citizen's juries", "deliberative mapping", and "focus groups" [44, p. 1571; 44] and hence a need for clarifying – in each case anew – what the purpose, the criteria and the methods of participation are. Stilgoe et al. also point to a more fundamental criticism. For instance, engagement processes would often "construct particular kinds of public" [44, p. 1572]. Thus, "inclusion" has a performative dimension and implies power. 'Society' and 'the public' are shaped by elicitation techniques. Radical

critics have argued that there is "a veritable *extractive* industry" which seeks to "generate certified 'public opinion'" in favour of existing formal authorities such as, the government [29, p. 280; emphasis in the original).

Moreover, moving further away from the RRI literature and closer to existing HCI approaches, participatory and user-centered tropes stand in contrast to "critical" [5] or "ludic" [9] design traditions aimed overtly at challenging users' sense of possibility beyond the horizon of the unconsidered routine and normative aspects of their lives. These practices perhaps also invert our notion of responsiveness (see last section): It is the users, not the designers, who are led to 'respond' to the design in ways that cause them to reflect upon their own taken for granted practices and cultural assumptions. In the same way that Bowen argues that 'disruptive' practices are compatible with (and enrich) participatory approaches by "opening up the design space" [5, p. 5], we can make the symmetrical argument that design always carries with it an implication for existing social orders that become apparent in the process of making. One author of this paper has observed that the design of a mammography training tool prised clinical resources from their usual regulated use context, making visible, and to an extent threatening, the existing clinical hierarchy. This happened because some staff could potentially make clinical decisions in the simulated environment that they could not normally legitimately make in real life practice. This underlines that design *per se* has a disruptive potential, irrespective of whether the design method is fine-tuned to bring this potential to the fore.

Perspectives, scope and temporality

As for reflexivity and responsiveness, we promote a pragmatic interpretation of the RRI principle of inclusion. Inclusiveness might be limited by practical constraints such as, time available and need to make progress (for example, reaching the next 'milestone' in an externally financed project). Inclusiveness on a local scale often involves users in the moment of design and other direct stakeholders. Value sensitive design (VSD) points to indirect stakeholders as an important category of 'user', yet the broad potential for social transformations of digital technologies implies that increasingly, the circle of indirect stakeholders becomes ever wider. Yetim criticises VSD for not providing a principled way of identifying indirect stakeholders and argues that the boundaries of the system can be determined (or at least critically explored) using a discursive approach [49]. But with widening scope, inclusiveness takes on a different complexion, particularly in contexts where design is concerned with the production and refinement of generic software packages aimed to be used across a large number of customer organisations. Here the competing demands of customer-users have to be managed and smoothed to prevent the software becoming overly complicated by functions specific to particular use contexts [35]. In such design settings, inclusiveness moves away from an ideal and becomes entangled in the pragmatic business concerns of 'supplying software' and 'making a profit'. Moreover, even where user engagement is prioritised, technological processes can resist incorporation of user-generated insights through a plethora of reasons that have often to do with the materiality of the artefact [21].

Temporality is also an important dimension of inclusion. One example from Hyysalo concerns how a company developing a wearable healthcare technology adjusted their user engagement strategy to be more participative when they were trying to establish their product. But then they adopted more 'arm's length' relationships with users in later phases when aiming at the generification of their product for a range of markets [21]. Thus, in commercial contexts, relationships with users may be more strategic and dictated by ever changing business necessities, rather than conforming always to the highest standards of user-centred design. Also, with time, the technology available for engaging users evolves, opening up new sorts of possibilities for mass monitoring of application use (raising new sorts of ethical concerns) and new forms of engagement 'in-the-wild' via 'living labs', for example. In the temporality dimension then both the reasons for motivating user involvement and the means by which this might be achieved are continuously shifting.

Anticipation

Interestingly, the idea of participation scrutinized in the previous section is linked to the RRI concept of "anticipation". It has been argued that responsible innovation includes forward-looking activities. However, and according to RRI proponents, these are not foresight and prediction based on (quantitative) risk assessment, which would often fail in providing early warnings for future hazards, but "anticipatory discussions of possible and desirable futures" with different stakeholders. Anticipation includes 'what if' questions; considering what is plausible, possible, and contingent; and being cautious about optimistic techno-scientific promises, trying to disentangle "present hype from future reality" [44, p. 1570-1571].

Anticipation thus, has much in common with reflexivity – but with a stronger emphasis on possible and even desirable futures. However, finding this (societal) desirability is a tightrope walk [e.g. 1, p. 994]. There is no straight-forward version of what 'the public' needs or wants. The moral impetus in the RRI agenda also risks downplaying existing efforts in research and design, e.g. the numerous 'small' anticipatory activities (thoughts, memos, coordinating time tables among colleagues) regarding next day's or next week's challenges – everyday responsible innovation in lower-case, not capital letters.

In other words, anticipatory practices are not only the province of RRI, nor are they always enacted through more calculated and formal risk assessments. Here we want to draw closer attention to the range of anticipatory practices that show that anticipation is a normal aspect of many spheres of life. We also want to draw attention to the

relationship between anticipation and governance. While "anticipatory governance" is an explicit RRI theme [e.g. 1], we wish to show that this is not a new or very 'special activity'. It is rather the 'thick end' of naturally occurring and multi-situated practices of technology selection and regulation across a range of contexts.

An important arena where anticipation is a key practice is in support of making business decisions, for example, in predicting which technologies to invest in, how markets might move and so on. Strategic decisions about possible technology futures are made within corporations at a high level by Technology Officer board members, aided by a sophisticated 'prediction industry' exemplified by management consultants such as Gartner. In these arenas anticipation is about seeking security in decisions made now that depend upon external factors known to be volatile. One aspect of anticipation from an RRI perspective concerns anticipating 'user response' to novel technologies and their potential, in particular to head off the sorts of outcry that erupted over the SPICE project to test the feasibility of cooling the earth by injecting particulates into the atmosphere [44]. Interestingly, this concern over 'user response', often framed in terms of "public concern" [12], actually shows how ordinary members of the public are themselves quite able to anticipate consequences of technology and technologically mediated social change. While these anticipations are perhaps most visible for controversial technologies and when aggregated and intensified via media or pressure group (re)presentation, perhaps they indicate more subtle and rarely investigated dimensions to anticipatory practices in everyday life in relation to technology. That is, how people make decisions about every day, mundane plans for the future with respect to their anticipations of technology and its trends and effects. Part of these will concern straightforward decisions that perhaps mirror corporate technology selection strategies, for example, predicting when a new model may become available, which vendor to commit to, which standards might survive. But there are also likely to be thoughts as to when to adopt or reject technologies with a view to their anticipated consequences on various modes of life or future plans, for example, the choice of vocation [23], or relying on the increasing availability of Internet reach when deciding to move to a rural location.

Perspectives, scope and temporality

An important aspect of anticipation for RRI is that it enables a link to regulating technology production through the shaping of research agendas to avoid unpalatable directions. One side of regulation that is related to scope and perhaps has less attention within RRI discourse, but is highly familiar to design, is the myriad contexts in which technology is used. Just as innovation studies show how technologies are configured over time, in and through use, to become 'domesticated' into local ecologies of artefacts and practices, we might expect also to find that local regulatory orders respond to technologies and their transformative potential in similar ways. Thus, once a technology is released into the wild, its totalising effects are not necessarily a foregone conclusion. As Healy says: "Sometimes a technology is so pervasive that we cannot escape it, but often we have the freedom to choose" [17].

We go further, however, and suggest there are commonly occurring responses to technology that lie between the poles of rejection or resigned acceptance. In other words there is a range of possible ways that technologies can be incorporated into local regulatory orders according to local values and practices and in anticipation of likely consequences. An example of this is how parents enforce time limits for children's use of games consoles or the Internet as part of the regulation of domestic life.

These instances of accommodating the benefits and pitfalls of different technologies with other aspects of everyday life are included in the 'Slow Tech' movement, for instance, establishing rules for mobile phone at dinner [e.g. 19]. Such local forms of anticipatory governance imply the arduous negotiation of different perspectives (for instance, the conflicting interests and prospects of family members). Moreover, in the time dimension, this 'small' (mundane) anticipatory work will consist of ever new corrections and adaptations in numerous small feedback loops.

Design occurring at larger scales, at the level of communities [28, 7], also has some interesting properties with respect to regulation. One is how communities and supportive technologies can co-emerge in ways that embed the collective's emerging norms, values and regulatory principles. A second is how, at the same time, the developing community engages in loose, collectively constituted governance processes that shape socio-technical configurations, allowing anticipation over technological possibilities and affording consensual processes of decision making over design choices [4]. For example, Botero's case study shows how technology arrangements supportive of a group of seniors choosing to live together can be assembled through communal design practices as part of, and in step with the nascent community's own emerging social practices. This represents a sort of informal 'RRI in action' and shows regulation to be a nuanced and multi-level phenomenon implicated in routine settings of use as much as within innovator's labs or legislative processes. Thus, elements of pacing and control, and the interweaving of technology design with the design of its (situated) regulation are perhaps significant, and stand opposed to the image of technology-driven social transformation as riding a wild horse where the best one can do is try to hold on.

CONCLUSIONS

This paper has provided a detailed comparison of the RRI agenda with HCI research and design in order to find out where the new external requirements are already met, where practical difficulties lie (in terms of perspectives, time and scope), but also, where HCI could (critically) 'talk

back'. We have discussed four RRI key notions: reflexivity, responsiveness, inclusion and anticipation.

We have explained the ways in which implementing the new and growing RRI agenda in HCI design is challenging. Despite this empirically grounded scepticism we have not rejected the RRI agenda entirely, in a radical or fatalist fashion. Instead, we have developed a pragmatic, 'unromantic' understanding of responsible design in HCI as an ongoing distributed effort of many parties at many levels, i.e. a continuous strive for shared responsibility. However, this envisaged collective responsibility is not a paradise on earth. It is hard work and does not imply ideal solutions for everybody, once and for all; and it may still result in harmful outcomes at system levels [32].

Following up on Lynch's work [27] we suggest a pragmatic interpretation of the ideal of *reflexivity*. It is not an absolute virtue but an attitude and practice whose precise contents, and the degree to which these can be enacted, depend on the concrete problems and opportunities of a given community of practice such as, a department, or a work group that has been shuffled together temporarily for a project. This relational concept includes further aspects. First, any kind of reflexivity expected from HCI researchers and designers hinges on the (limited) availability of resources. Mundane things like money, time and enabling work conditions are crucial. Second, "reflective design" [42] of a given technology cannot be achieved once and for all but is an ongoing learning effort, i.e. a "collaborative process" [42, p. 56] of both designers and users – and, as we wish to emphasize, also of funders, policy makers and other institutional stakeholders. Thus, reflexivity is spatially dispersed and temporally stretched, and it is "reflection-in-action". Technology is not conceptualized as a product to be designed, delivered and then to be 'ticked off', but as an "open space" that different users and other stakeholders continually engage with, and that they may challenge at times [42, p. 55]. Accordingly, designers are asked to rethink the technology periodically and to feed back experiences into the (re)design process, as far as possible (given the limited resources discussed previously). Third, we think that "reflexivity" is not *per se* good, and it is not always productive. A lot of design work successfully relies on the *un*reflected, i.e. tacit, embodied knowledge an HCI researcher or designer has acquired over time. Obsessive reflection, then, may distract and confuse with little benefit.

This lesson is inspired by Hallnäs' and Redström's account of "slow technologies" [13]. They designed a system that collects and rearranges information from e-mails and other sources at a given place such as, an office, and then reprojects it to the local audience. While the original design idea was to entertain and increase workplace awareness, in the office context (and less so, for instance, at a reception party) the technology-induced "reflexivity" was perceived as inefficient and rather uninspiring. Accordingly, we consider these and other forms of reflexivity such as, "group reflexivity" [30], not as a panacea, but as occasional work technique that should be well-dosed.

"Responsibility" is different from the RRI concept of *responsiveness*. It is a more forward-looking activity – although not too far, we suggest: responsiveness is a heedfulness that matters mostly in the here and now, a situated, interactive and face-to-face-accomplishment. We discussed the care for elder people as a sensitive area where such a local face-to-face heedfulness may be 'endangered' by relying too much on sophisticated technologies designed in a one-off way. We ask designers to perhaps think of design as a longer term process that is more sensitive, more responsive to locally emerging issues, and less 'mediated'.

However, such an engaged responsive design is a difficult task. It may be *especially* hard to realise in those contexts were it is *most* needed, e.g. with so-called 'vulnerable' people who may challenge the capacities of all involved. Again, a corresponding institutional responsiveness may be required, e.g. more organisational or financial support.

Inclusiveness is a further RRI concept that is also embedded within various traditions of design – namely user-centred and participatory design – often for comparable reasons, such as, ideals of social democracy, or protecting against negative side effects by taking into account the concerns of more distantly implicated populations. Based on our own research on training tools designed for, and used in clinical settings, we pointed to the positive disruptive potential of critical design – and argued that the social role of the designer as provocateur is perhaps as significant a responsibility as accounting for (unchallenged) user needs as emphasized by participatory and user-centred traditions. Thus, this concept of the (productively) provocative designer may be considered as one of contributions the HCI community can make to the RRI literature.

While precisely defining 'indirect stakeholders' both for RRI and design is problematic, design at least has grappled with this problem in the form of approaches to 'boundary identification'. Design too has a grasp of how user engagement at scale changes complexion due to the economics of software provision, where 'the user' becomes harder to identify and competing user needs have to be managed. It is likely that RRI will also be challenged by populations with a diversity of aspirations, values and concerns, and it will be interesting to see what approaches might be used to manage these.

In terms of *anticipation* we focused on the relationship with governance, a link also made by some RRI proponents. By reference to the 'Slow Tech' movement we have argued, though, that anticipatory governance is less of a special than a mundane activity, from local to larger scale-community level. In our view, supposedly harmful technology development may be outweighed by user agency, that is, the various ways in which users already regulate technological impacts on their everyday life, or try to do so.

Thus, some aspects of anticipation are already embedded, but perhaps there is an opportunity for designers to take greater advantage of this, and assist users further by creating designs that help afford regulation in use. A simple example would be the light on a camera that indicates when it is recording, and lets people adapt their behaviour accordingly. Such design for anticipation departs from experiences with extant technologies in the here and now, and consists of rather small feedback loops of (re)adaptions.

In this regard, we agree with Reeves' analysis of "envisioning" in design [38]. Envisionings may speak about the future but actually reflect the concerns of the present. We think that this implicit grounding in the present could actually be seized more proactively: study present user problems in order to improve on your design principles; but take these rather as "questions" than as cemented instructions, given that uncertainty is such a basic feature of reality. "Trial and error", then, is the (pragmatic) way to move forward into the future, in small steps – "a way of continually recalibrating and refining" the original principles [38; p. 1580]. Since present user problems are heterogeneous, the envisionings should also be manifold. But this brings us back to a fundamental problem discussed earlier: envisioning for diversity requires sufficient resources (time, money, etc.). These structural questions exceed the duration of projects and need to be addressed by the entire ecosystem of funding agencies and other organizations making up the institutional landscape of HCI.

ACKNOWLEDGMENTS

We thank our reviewers for their very helpful suggestions. The research leading to these results has received funding from the European Community's Seventh Framework Programme (FP7/2007-2013) under grant agreement n° 321480, and grant agreement n° 600854.

REFERENCES

1. Barben, D., Fisher, E., Selin, C., and Guston, D. Anticipatory governance of nanotechnology: Foresight, Engagement, and Integration. In *The Handbook of Science and technology Studies,* Hackett, E., Amsterdamska, O., Lynch, M., Wajcman, J., eds., MIT Press, Cambridge (2007), 979-1000.
2. Baumer, E.P.S., and Silberman, M.S. When the implication is not to design (technology). In *Proc. CHI 2011*, ACM Press (2011), 2271–2274.
3. Beck, U. *Risk society: towards a new modernity*. Sage Publications, London ; Newbury Park, Calif, 1992.
4. Botero, A. Expanding design space(s) - design in communal endeavours, PhD Thesis. Aalto University, School of Arts, Design and Architecture, 2013.
5. Bowen, S. Critical theory and participatory design. In *Proc. CHI 2010*, ACM Press (2010).
6. Eden, G., Jirotka, M. and Stahl, B. Responsible Research and Innovation: Critical reflection into the potential social consequences of ICT. *IEEE RCIS* 2013 (2013).
7. Franqueira, T. Reshaping urban lives - design as a social intervention towards community networks. In *Proc. IASDR 2007* (2007).
8. Friedman, B. Value-sensitive design. *Interactions November + December* (1996), 17-23.
9. Gaver, W.W., Bowers, J., Boucher, A., Gellerson, H., Pennington, S., Schmidt, A., Steed, A., Villars, N., and Walker, B. The drift table: designing for ludic engagement. In *Proc. CHI 2004*, ACM Press (2004), 885.
10. Greenhalgh, T., Wherton, J., Sugarhood, P., Hinder, S., Procter, R., and Stones, R. What matters to older people with assisted living needs? A phenomenological analysis of the use and non-use of telehealth and telecare. *Social Science & Medicine 93* (2013), 86–94.
11. Gregory, J. Scandinavian approaches to participatory design. *Internat. J. of Engineering Education 19*, 1 (2003), 62-74.
12. Guston, D.H., and Sarewitz, D. Real-time technology assessment. *Technology in Society 24* (2002), 93–109.
13. Hallnäs, L., and Redström, J. Slow technology – designing for reflection. *Personal and Ubiquitous Computing* 5 (2001): 201-212.
14. Hanson, V.L. Influencing technology adoption by older adults. *Interacting with Computers 22* (2010), 502–509.
15. Haraway, D. Situated Knowledges: The Science Question in Feminism and the Privilege of Partial Perspective. In *Simians, Cyborgs, and Women. The Reinvention of Nature*. Free Association Books, London (1991), 183-201.
16. Harper, R., Randall, D., Smyth, N., Heledd, L. And Moore, R. The past is a different place: they do things differently there. In *Proc. DIS 2008,* ACM Press (2008), 271-280.
17. Healy, T. The unanticipated consequences of technology. In *Nanotechnology: ethical and social Implications*, A.S. Khan, ed. Taylor & Francis, Boca Raton, FL (2012), 155–174.
18. Hill, L., Roberts, G., Wildgoose, J., Perkins, R., and Hahn, S. Recovery and person-centred care in dementia: common purpose, common practice? *Advances in Psychiatric Treatment 16* (2010), 288–298.
19. Hofmann, J. B. http://www.janellburleyhofmann.com/slow-tech-manifesto/
20. Holmes, M. Emotional reflexivity in contemporary friendships: understanding it using Elias and Facebook etiquette. *Sociological Research Online 16, 1* (2011).
21. Hyysalo, S. *Health technology development and use. From practice bound imaginations to evolving impacts.* Routledge, London, 2010.
22. Inside Out of Mind. http://www.meetinggroundtheatrecompany.co.uk/projects/inside-out-of-mind

23. Jackson, S., and S. Barbrow. Infrastructure and vocation: field, calling and computation in ecology. In *Proc CHI 2013*, ACM Press (2013), 2873-2882.
24. Kelly, P., Marshall, S.J., Badland, H., Kerr, J., Oliver, M., Doherty, A.R., and Foster, C. An ethical framework for automated, wearable cameras in health behavior research. *Am. J. of Prev. Medicine 44* (2013), 314–319.
25. Larab, A., Conchon, E., Bastide, R., and Singer, N. A sustainable software architecture for home care monitoring applications. In *Proc. IEEE DEST* (2012).
26. Latour, B. *We have never been modern*. Harvard University Press, Cambridge, Mass, 1993.
27. Lynch, M. Against reflexivity as an academic virtue and source of privileged knowledge. *Theory Culture Society 17* (2000): 26-54.
28. Leadbeater, C. *The user innovation revolution: how business can unlock the value of customers' ideas*. National Consumer Council, London, 2006.
29. Lezaun, J., and Soneryd, L. Consulting citizens: technologies of elicitation and the mobility of publics. *Public Understanding of Science 16* (2007), 279–297.
30. Moncur, W. The emotional wellbeing of researchers: considerations for practice. In *Proc. CHI 2013*, ACM Press (2013), 1883–1890.
31. Mort, M., Roberts, C., and Milligan, C. Telecare and older people: Re-ordering social relations. In *Towards Responsible Research and Innovation in the Information and Communication Technologies and Security Technologies Fields*, R. Schomberg, ed. Publications Office of the EU, Luxemburg, (2011), 149–164.
32. Owen, R., Stilgoe, J.,Macnaghten, P., Gorman, M., Fisher, E., Guston, D. A framework for responsible innovation. In *Responsible Innovation: Managing the Responsible Emergence of Science and Innovation in Society*, Owen, R., Bessant, J., and Heintz, M., ed., Wiley (2013), 27-50.
33. Page, X., and Kobsa, A. The circles of latitude: adoption and usage of location tracking in online social networking. In *IEEE Comp. Science & Eng.* (2009), 1027–1030.
34. Pellizzoni, L. Responsibility and environmental governance. *Environm. Politics 13* (2004), 541–565.
35. Pollock, N., and Williams, R. Who decides the shape of product markets? The knowledge institutions that name and categorise new technologies. *Information and Organization 21* (2011), 194–217.
36. ProGReSS. http://www.progressproject.eu/
37. Randles, S., Youtie, J., Guston, D., Harthorn, B., Newfield, C., Shapira, P., Wickson, F., Rip, A., von Schomberg, R., and Pidgeon, N. A trans-atlantic conversation on Responsible Innovation and Responsible Governance. 2012. http://works.bepress.com/pshapira/54/
38. Reeves, S. Envisioning ubiquitous computing. In *Proc. CHI 2012,* ACM Press (2012), 1573-1582.
39. Rosner, D., and Bean, J. The Weapons Factory in the Den. *Interactions* May + June (2013), 80-81.
40. Schienke, E.W., Tuana, N., Brown, D.A., Davis, K.J., Keller, K., Shortle, J.S., Stickler, M., and Baum, S.D. The role of the National Science Foundation Broader Impacts Criterion in enhancing research ethics pedagogy. *Social Epistemology 23* (2009), 317–336.
41. Sellen, A., Rogers, Y., Harper, R., and Rodden, T. Reflecting human values in a digital age. *Comm. of the ACM 52* (2009), 358–366.
42. Sengers, P., Boehner, K., David, S., and Kaye, J. "Jofish". Reflective design. In *Proc. Critical Computing 2005*, ACM Press (2005), 49–58.
43. Singer, N., Trouilhet, S., Rammal, A., and Pécatte, J.-M. Collecting and classifying large scale data to build an adaptive and collective memory: A case study in e-Health for a pro-active management. In *Multi-Agent Systems - Modeling, Interactions, Simulations and Case Studies*, F. Alkhateeb, E. Al Maghayreh, and I. Abu Doush, eds. (2011), 439–454.
44. Stilgoe, J., Owen, R., and Macnaghten, P. Developing a framework for responsible innovation. *Research Policy*, *42* (2013), 1568-1580.
45. Sutcliffe, H. A report on Responsible Research & Innovation., 2011. http://www.matterforall.org/pdf/RRI-Report2.pdf
46. Sweeney, L. Discrimination in Online Ad Delivery. *Queue 11*, 3 (2013).
47. Ure, J., Pinnock, H., Hanley, J., Kidd G, McCall Smith, E., Tarling, A., Pagliari, C., Sheikh, A., MacNee, W., McKinstry, B. Piloting telemonitoring in COPD: a mixed methods exploration of issues in design and implementation. *PRIM Care Resp J 21,* 1 (2012), 57-64.
48. Vincent, N.A. A structured taxonomy of responsibility concepts. In *Moral Responsibility*, N.A. Vincent, I. van de Poel, and J. Hoven, eds., Springer (2011), 15–35.
49. Yetim, F. Bringing discourse ethics to value sensitive design: pathways toward a deliberative future. *AIS Trans. on Human-Comp. Interact. 3,* 2 (2011), 133–155.
50. Von Schomberg, R. A vision of Responsible Research and Innovation. In *Responsible Innovation: Managing the Responsible Emergence of Science and Innovation in Society*, Owen, R., Bessant, J., and Heintz, M., ed., Wiley, Hoboken (2013), 51-74.