RRI measurement and assessment

Some pitfalls and a proposed way forward

van de Poel, I.R.

DOI
10.4324/9780429298998-25

Publication date
2021

Document Version
Final published version

Published in
Assessment of Responsible Innovation

14 RRI measurement and assessment

Some pitfalls and a proposed way forward

Ibo van de Poel

14.1 Introduction

In the last decade, there has been increasing attention on responsible innovation, or Responsible Research and Innovation (RRI), as a way to increase the responsible development and deployment of new technology in society. RRI refers to a more reflexive form of research and development (R&D) and innovation that is aimed at better aligning research and innovation with the "values, needs and expectations of society" (European Commission, 2014: 73).

Uptake of RRI by industry and research organizations appears to be a slow and gradual process. One underlying factor is that the vocabulary of RRI is sometimes perceived as academic or "foreign" by innovators (Dreyer et al., 2017). This does, however, not mean that the underlying rationale and motives are not recognized. Many of the actions and activities that are now promoted as RRI are already undertaken by companies and research organizations, albeit often under different heading such as corporate social responsibility (CSR), social innovation and sustainable innovation (Lubberink, Blok, van Ophem, & Omta, 2017; van de Poel et al., 2017). RRI may perhaps be best construed as an attempt to broaden and systematize such activities. What RRI seems to add is particularly a broadening of values considered (e.g. not only safety, sustainability and privacy), more attention for stakeholder involvement and public engagement and a more proactive stance where such issues are already addressed during the early phases of R&D and innovation (Stilgoe, Owen, & Macnaghten, 2013; Van den Hoven, 2013). Some have also argued that responsible innovation starts from societal challenges and needs rather than from technical opportunities, as in traditional innovation (Von Schomberg, 2019).

It has been pointed out that the uptake of RRI is constrained by the fact that current incentives for research organizations and companies point in other directions. This has led to pleas for incentivizing RRI (Gurzawska, Mäkinen, & Brey, 2017). Such incentivizing may take a multiplicity of forms, from legal requirements to financial incentives. Examples are the development of quality marks for RRI, or governments requiring RRI compliance as a condition for allowing new products on to the market.

The above considerations have led to an increasing attention for the development of methods and tools to measure, assess and monitor RRI performance. Examples can be found in the various chapters in this book (e.g. Tharani, Jarmai, Schönherr, & Urban, Chapter 8; Klaassen, Verwoerd, Kupper, & Regeer, Chapter 9; Verburg, Rook, & Pesch, Chapter 13). These tools may, for example, be used to assess where organizations stand with respect to RRI activities and awareness, even if such activities may not always be defined as RRI by the organization itself. Such assessment tools may point out possibilities for improvement, but they may also play a role in incentivizing schemes intended to promote the uptake of RRI.

If tools for RRI measurement, assessment and incentivizing are to be effective, they need not only to result in a measurement of RRI performance that is reliable and valid, but also to contribute to behavior incentives that contribute to an increased uptake of RRI and its underlying aims. However, this is by no means straightforward. Performance measurement in general is a messy and difficult process, and it may occasionally result in incentives and behavior effects that are contrary to what was intended (De Bruijn, 2007).

The aim of this chapter therefore is to highlight some of the potential pitfalls of RRI measurement, assessment and incentivizing. As we will see, avoiding these pitfalls may not be so easy (although not necessarily impossible) and there may be trade-offs between avoiding different types of pitfalls. In order to better deal with these pitfalls and dilemmas, I will propose a potential way forward.

The chapter starts with sketching the constellation of actors that may be typically involved in RRI measurement, assessment and incentivizing (Section 14.2). Next, I discuss typical motivations that may exist for doing RRI assessment (Section 14.3). These first two sections not only provide an introduction to the topic, but also provide the analytical tools that can be used to analyze in more detail concrete situations. Such analysis is a prerequisite to uncovering potential pitfalls of RRI measurement, assessment and incentivizing. Section 14.4 gives a brief overview of existing RRI assessment methods, which will also be used to illustrate some of the pitfalls in the next sections. Section 14.5 focuses on pitfalls due to measurement problems and Section 14.6 focuses on pitfalls due to behavior effects. In Section 14.7, I provide a more general discussion of the pitfalls and dilemmas in RRI assessment and I argue that the dilemmas are to a large extent due to the fact that there are different, conflicting rationales for doing RRI assessment. I suggest that getting the rationale for an RRI assessment method clear is a first step to better navigating the earlier distinguished pitfalls and dilemmas.

14.2 The constellation of actors

Before discussing the motivations behind RRI assessment and some of its potential pitfalls, it is useful to sketch an ideal-typical constellation of actors

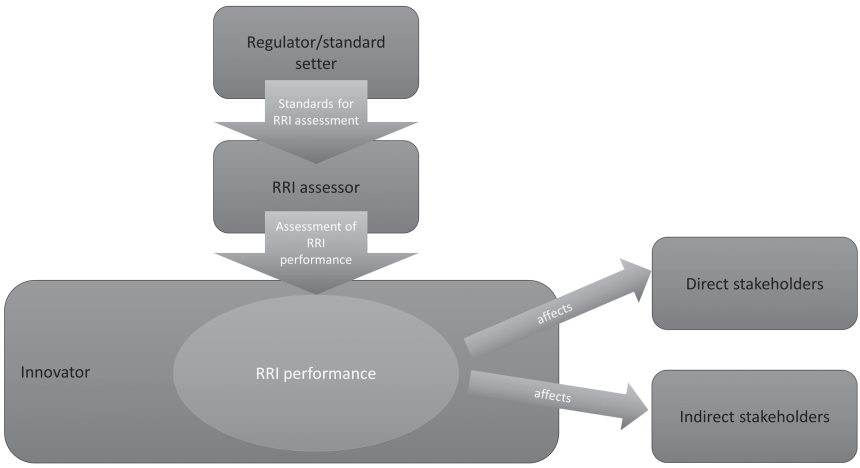

Figure 14.1 Actor roles in Responsible Research and Innovation (RRI) assessment.

that is somehow involved in, or affected by RRI assessment. Figure 14.1 lays out the main actor roles that I propose to distinguish. I discuss each of these briefly below.

14.2.1 Innovator

This is the actor that is the object of the RRI assessment. I assume here that RRI assessment is applied to a specific organization rather than to the entire knowledge or innovation system. Given the nature of RRI, I will assume that this is an organization that is developing innovative "products" (including knowledge) and that is doing R&D-like activities. I take products here broadly; basically any kind of output that can be used by another actor for another end. These products need to be somehow new and innovative, although I will refrain here from providing a precise definition of "innovation". The innovation needs to be the result of some deliberate knowledge generation activity, i.e. from R&D, although I will assume that the organization doing RRI can also focus only on research, or only on development. This first actor role of an innovator is typically played by organizations such as research laboratories, universities and companies.

14.2.2 Regulator or standard setter

This is the actor that is regulating or setting (RRI) standards for the innovator. I understand here regulation and standards very broadly. It refers to any kinds of requirements for the RRI behavior and performance of the assessed organization. Such standards can be compulsory (e.g. legal requirements

by the government) or voluntary (e.g. a non-binding code of ethics), or in between (e.g. required to receive a quality mark). They can relate to the innovative products developed by the innovator, but also to organizational requirements or procedures (e.g. the requirement to have a safety officer); they can also concern how RRI assessment is to be carried out and by whom. Although I focus here on requirements and standards related to RRI, this does not always mean that they themselves have to explicitly mention RRI. Think for example of requirements for safety and sustainability that are clearly related to RRI but need not mention RRI. Typical actors that can fulfill this second actor role are the government, standard-setting organizations like ISO and its national counterparts, branch organizations and organizations formulating quality marks. Also innovators may formulate their own rules and standards.

14.2.3 RRI assessor

This is the actor doing the actual RRI assessment. Since RRI is still in its infancy, in many cases this is not yet a specialized role that is played by a separate actor. Instead, this role is often played either by the regulator or standard setter or by the innovator itself. It may currently also be played by a temporary organization, like a European RRI project that is involved in developing (tools for) RRI assessment. However, this role can in principle also be played by separate actors like, for example, a consultancy firm or an independent organization set up for such purposes.

14.2.4 Direct stakeholders

I will here understand the direct stakeholders as those actors that use the (innovative) products of the assessed organizations.[1] Presumably, the direct stakeholders have an interest in the RRI performance of the innovator. This will, for example, reassure (or even guarantee) them that the products they use from the assessed organization meet certain safety or sustainability standards or do not invade the privacy of certain groups.

14.2.5 Indirect stakeholders

Indirect stakeholders are actors that are somehow affected by the activities of the innovator or by the use of the innovative products developed by the innovator (and used by the direct stakeholders). Indirect stakeholders thus include actors further along the value chain of an innovative product (who do not directly use the product itself) or so-called bystanders. Given the understanding of RRI as developing products that meet "the needs and values of society" (European Commission, 2014: 73), the indirect stakeholders are an important category in RRI assessment, as the ultimate aim of RRI would seem to be to assure that also values and needs of those

indirect stakeholders are properly addressed by the innovator in its innovation and R&D activities.

The above description of actors' roles is useful to better understand some of the motivations behind RRI assessment and some of its potential pitfalls. Figure 14.1, however, merely describes a conceptual framework with ideal-typical actor roles. It does not yet describe the situation in a specific case, with specific actors fulfilling specific actor roles. This specific constellation will be different from case to case, and describing it may occasionally also require distinguishing additional actor roles. Also, as already alluded to, in some cases actors' roles may be combined by one actor. For example, the innovator may set its own standards and assess them itself (i.e. some form of self-regulation).

Detailing Figure 14.1 for a specific situation (like a specific innovator or for a specific industry branch or technological domain) is useful because it helps to see where roles and interests of the various actors are overlapping, complementary or (potentially) conflicting. This in turn may point to certain (potential) weaknesses and pitfalls of RRI assessment in the specific situation.

14.3 Potential motivations for doing RRI assessment

There are various potential motivations for doing RRI assessment, and these may be different for different actors. I will here discuss some main motivations for doing RRI assessment without claiming to be complete. The motivations I will discuss are: (1) compliance; (2) increasing transparency, accountability and trust; (3) improving (RRI) performance; and (4) the desire to avoid (unnecessary) regulation.

14.3.1 Compliance

A first motivation for doing RRI assessment might be to check whether the innovator is compliant with existing rules and standards that somehow pertain to RRI issues. As indicated, such rules and standards need not be formulated in straightforward RRI language (as that is still rare), but could pertain to issues such as safety, sustainability, privacy, transparency, integrity, and so on. For the company, compliance may be a way to show to regulators and direct and indirect stakeholders that they meet the relevant standards. For the regulator, RRI assessment may be a means to check whether an innovator indeed is compliant, and for direct and indirect stakeholders it may be a means to ensure that they can rely on the products of the innovator.

14.3.2 Accountability, transparency and trust

Compliance requires explicit rules and standards against which compliance can be checked. In many cases, such rules and standards will be lacking

for innovative products. One important reason for this is that regulation (and standardization) tend to lag behind the development of new technologies. RRI was in fact in part proposed to deal with issues that are not (yet) regulated. So while RRI assessment may be used to check compliance, in many cases RRI will be applied in situations in which (new) rules and standards have not been set yet or in which they are still being established or evolving. In such situations, RRI assessment may still be relevant to show not just that RRI procedures are in place with the innovator but also that these are functioning reasonably well. This may then make the innovation processes of the innovator more transparent, contribute to the accountability of the innovator and so create trust among other actors, like regulators and stakeholders.

14.3.3 Learning and improving performance

A third motivation for doing RRI assessment may to learn from it and to improve (RRI) performance. This is primarily a motive for the actor doing RRI, i.e. the innovator, but learning and improved (RRI) performance is likely also in the interest of the other actors, like regulators and stakeholders. The reasons why RRI assessment can contribute to learning and help to improve performance are quite straightforward. RRI assessment will show on which RRI parameters the innovator is performing well and on which not so well, which indicates possibilities for improving performance. Moreover, RRI assessment may help to gain insight in the (cost)effectiveness of various RRI measures and procedures and so may help to increase performance without necessarily increasing the costs of RRI.

14.3.4 Avoiding regulation

A fourth, more controversial, motivation for RRI assessment may be to avoid, or postpone, regulation. By doing RRI assessment innovators might want to show that they are taking their responsibility and that no government regulation is needed. One might argue that this is not a proper reason for RRI assessment but rather amounts to a pitfall. It is indeed true that innovators may misuse RRI assessment for purposes like avoiding regulation (cf. the discussion below on window dressing). However, it should be noted that regulation is not always desirable or in the interests of stakeholders (and the regulator). Government regulation comes with its own pitfalls, and if RRI can be achieved without explicit government regulation it may be more desirable than having regulation. The point, of course, is that from a societal point of view, regulation is sometimes needed.

Understanding the motivation for RRI assessment in a concrete situation is important because it will to an important extent determine whether, and which of, the pitfalls that I discuss below will actually materialize. I would therefore suggest that an analysis of concrete situations should start with

both a mapping of the actors involved (Section 14.2) and their motivations behind doing RRI assessment (Section 14.3), in order to analyze which pitfalls of RRI assessment can be expected (Sections 14.5 and 14.6) and should be prevented if possible. But before we turn to the pitfalls, it is useful to briefly discuss some existing RRI assessment methods.

14.4 An overview of some existing RRI metrics and assessment methods

Table 14.1 provides an overview of some existing RRI metrics and assessment methods (see Chapters 8, 9 and 13; Flipse, Dam, Stragier, Vrielink, & Sanden, 2015; Ravn, Nielsen, & Mejlgaard, 2015; Stahl et al., 2017; Strand et al., 2015; Wickson & Carew, 2014). The dimensions have been filled out by the author on the basis of the mentioned references; in most cases this was rather straightforward; but in a number of cases it required some interpretation.[2] As the table shows, existing RRI assessment methods have a range of different aims, including monitoring, measuring RRI levels, comparison and learning. In terms of the motivations for doing RRI assessment that I discussed in Section 14.2, these aims are mainly related to the second (accountability) and third (learning and improving performance) motivation mentioned. It is not very surprising that none is related to compliance, since there are not (yet) legal requirements or quality marks for RRI, which can serve as a basis for compliance. Similarly, it is not surprising that avoiding regulation is not an (official) aim of the proposed methods. This does not rule out that use of these methods may sometimes be motivated by the desire to avoid regulation, but it is obviously not an aim of the methods.

The different methods are geared towards different objects of assessment, from employees to countries; most are however aimed at either the organizational level (company) or the project level. What is further striking is that most are based on some form of self-assessment. This is probably explained by two factors. One is that RRI is a complex notion with many dimensions; I will further reflect on this feature in the next section. The other is that RRI assessment, and in the sense the whole field of RRI, is a relatively recent endeavor and it usually takes time to operationalize complex notions like RRI into features or items that are objectively measurable; although approaches 2 and 3 in the table are clearly attempts to do so.

I will reflect on these and other features of the existing RRI assessment in the next section when I discuss typical measurement problems that the notion of RRI gives rise to.

14.5 Measurement problems

Since RRI is a complex notion, RRI (performance) cannot be directly measured. Moreover, RRI is a normative or value-laden notion, i.e. it expresses what is a (more) desirable form of innovation, which may make it

Table 14.1 Overview of some Responsible Research and Innovation (RRI) assessment methods proposed in the literature

	Reference	*Unit of assessment*	*Assessor*	*Aim of assessment*	*Based on 6 EU keys*[a]	*Based on AREA framework*[b]	*Structure*	*Type of measurement*	*Aggregate score*
1	Wickson and Carew (2014)	Project	Multiple	Multiple	No	Yes	7 criteria, rubric for each criterion	Judgement	No
2	Ravn, Nielsen, and Mejlgaard (2015)	Country	Independent assessor	Monitoring; comparison	Yes	No	6 dimensions, 36 indicators	Objective	No
3	Strand et al. (2015)	RRI initiative	Independent assessor	Monitor and assess the impacts of RRI initiatives	Yes	Yes	6 dimensions, each with performance (process and product) indicators and perception indicators[c]	Objective	No
4	Stahl et al. (2017)	Company	RRI researchers; self-assessment[d]	Assessing RRI level, monitoring	No	Yes	3 RRI categories, 14 RRI components, each scored on 5-point scale (maturity levels)	Judgement	No

5	Flipse et al. (2015)	Project (within a company)	Self-assessment	Monitoring; decision support for managers	No	Yes	8 Key performance indicators (KPIs), each consisting of several items	Judgement	No
6	Tharani et al. (Chapter 8, this book)	Company	Self-assessment	Learning	No	Yes	4 sections; 43 questions	Judgement	No
7	Klaassen et al. (Chapter 9, this book)	Project	Self-assessment	Learning	No	Yes	4 process dimensions, each with criteria, subcriteria and inviting questions	Judgement	No
8	Verburg, Rook, and Pesch (Chapter 13, this book)	Employee (in a company)	Self-assessment	Assessing RRI level	No	Yes	7 items	Judgement	no

Notes
a The six RRI keys of the EU are: engagement; gender equality; science education; ethics; open access; governance(European Commission, 2012).
b This framework identifies four procedural dimensions for RRI: anticipation, reflexivity, engagement (or inclusiveness), action (or responsiveness) (EPSRC, 2019; Owen et al., 2013).
c Based on Table 3.1 in the report with the prioritized indicators. The report suggests two more potential dimensions (and indicators).
d In the publication it is applied by the RRI researcher but self-assessment is suggested as a possible way to scale up the method.

even more difficult to measure. This means that the notion of RRI, and RRI performance, first needs to be operationalized before it can be measured.

Operationalization of RRI performance may, for example, follow a two-step procedure (cf. Keeney, 1992; Kroes & van de Poel, 2015). First, a number of dimensions (or evaluation criteria or objectives) are associated with RRI, which as such may also not be directly measurable. To make these dimensions measurable, a number of measurable items (attributes, indices or questions) may be associated with each dimension. Often, these items will not exactly measure the relevant dimensions, but only by approximation, so that they are best seen as proxies for the chosen RRI construct.

The need to operationalize RRI is clearly visible from Table 14.1. As the table shows, in six out of eight RRI assessment methods (nos. 2–7) RRI is at least a two-level construct (in one case (no. 7) it even has four levels), with typically at least ten items at the lowest level. Only two methods (nos. 1 and 8) have only one level and fewer than ten items.

If RRI performance is operationalized by defining a range of items (attributes, indices or questions), as is the case with all methods summarized in Table 14.1, a next issue is whether – and if so, how – to aggregate scores on these into an overall score. One option is to limit oneself to scores on individual items, or a number of (aggregated) indices, and to refrain from an overall score. Indeed all existing RRI assessment approaches considered here refrain from calculating an overall score. This avoids the problem of aggregation but may make it difficult to judge whether RRI performance has improved (over time), particularly if performance on some items increases while it decreases on other items (cf. Bradburn, Cartwright, & Fuller, 2017). Some of the methods propose the use of spider diagrams that can help to show which dimensions of RRI improve over time, and which dimensions decline.

Two general concerns may arise with respect to the measurement of RRI performance, namely reliability and validity (Carmines & Zeller, 1979). Reliability means that the measurement measures correctly, i.e. that it measures the "real value" of an item. This is often understood as implying that if the measurement were done again it would measure the same value for an item, or – if the measurement is done by somebody else – it will result in the same value. It also typically implies that if two items intend to measure the same phenomenon, their measured value should be the same.

There are a number of reasons why metrics for RRI may result in unreliable measurements. One is that the attributes may often not be objectively measurable, but rather are items or questions that require a judgment by the one filling out the questionnaire. Indeed, only two of the assessment methods in Table 14.1 (i.e. nos. 2 and 3) make use of objectively measurable indicators; all others require some form of judgment. The subjectivity of judgment may be further aggravated by the fact that items or questions sometimes contain vague or ambivalent terms like, for example, "sufficient", "better" and "relevant". These terms may be interpreted differently

by different persons. And even one and the same person may interpret these terms differently at different points in time, which may lead to unreliable results.

Take for example a question like "Are there procedures in place to address the relevant ethical issues raised by the innovation?" A growing awareness of ethical issues may result in that the respondent answers at T_1 yes and at T_2 no, not because fewer ethical issues are addressed at T_2, but simply because the respondent has become aware of more ethical issues. So, while it may be argued that actual RRI performance has increased from T_1 to T_2 because there is more awareness of ethical issues, and because more issues are addressed at T_2 than T_1, on the basis of the answers to the indicated question the suggestion may arise that the actual RRI performance has decreased.

One way to reduce the subjectivity of judgment in RRI assessment is to provide a rubrics that gives guidance how to score questions or items, as the first method in Table 14.1 indeed does (Wickson & Carew, 2014). Another way in which subjectivity may be diminished is by involving more people in scoring items, or using questions that are a starting point for discussion, as seems to be the aim in methods 6 and 7 in Table 14.1,[3] but which may also work with some of the other methods.

A related potential source of unreliability is that RRI assessment is often a form of self-assessment, i.e. the innovator him- or herself has to fill out a questionnaire, or has to score items on a scale. This is indeed the case for six out of eight of the RRI assessment methods listed in Table 14.1. This may result in biased measurements, in particular when the innovator has a strategic interest in the outcome of the measurement. (This issue will be further discussed in the next section.)

Apart from reliability, validity may be an issue. A measurement is usually seen as valid if it measures what it intends to measure (rather than something else). As we have seen, RRI, or RRI performance, is not directly measurable. We first need to operationalize it to make it measurable. However, it is very conceivable that the chosen operationalization does not exactly or completely cover the underlying concept. The issue here is one of content or construct validity (Carmines & Zeller, 1979). By operationalizing RRI (performance), we create a construct that is measurable, but this construct may lack content or construct validity. This may be particularly so because in choosing a particular construct we may have good reasons to choose it in such a way that it is reliably measurable. However, the attributes that are most reliably measurable may not be the ones that are also most relevant for RRI, so diminishing construct validity.

Of course, the problem of content or construct validity is not unique to RRI; it applies to any complex social (or psychological) notion that we intend to measure (Carmines & Zeller, 1979). Examples are notions like intelligence or well-being. However, there seem to be a few underlying reasons why RRI may be particularly difficult to measure.

A first reason is that there is no consensus on the definition of RRI. Some definitions of RRI for example are based on the so-called AREA framework and they stress four procedural criteria for RRI: anticipation, reflection, engagement (or inclusiveness) and action (or responsiveness) (EPSRC, 2019; Owen et al., 2013; Stilgoe et al., 2013). But one may also place more emphasis on the six keys (engagement; gender equality; science education; ethics; open access; governance) for RRI defined by the European Union (EU) (European Commission, 2012). Yet other definitions of RRI stress the outcome or product dimension of RRI, for example emphasizing that innovative products should respect certain values (Van den Hoven, 2013). Or they may place emphasis on whether innovations contribute to the sustainable developments goals of the United Nations.

A look at Table 14.1 shows that all considered approaches were somehow inspired by the four procedural criteria for RRI; only two used the six EU keys for RRI. However, what is perhaps most remarkable is that the RRI constructs for the eight methods shown in Table 14.1 are very different. Methods 2 and 3 use the same six dimensions (i.e. the EU RRI keys) but somewhat different indicators. All the others RRI constructs are rather different from each other, already at the highest level, but certainly in terms of more detailed items. One might wonder how what is supposedly one concept can lead to such diverse constructs.

The underlying reason here seems to be that RRI is what Bradburn et al. (2017) call a *Ballung* concept and which they distinguish from a *pinpoint* concept. *Ballung* refers to the German word for congestion, which is used because a lot is packed into a concept. Bradburn et al. (2017: 76) say about such concepts:

> There is often no central core without which one does not merit the label, different clusterings of features among the congestion (*Ballung*) can matter for different uses, and whether a feature counts as being or outside the concept – and how far outside – is context and use dependent.

This indeed seems true of RRI. Some authors have noted that the more specific content of RRI is largely left open (Oftedal, 2014). As we have also seen, different definitions of RRI have been proposed stressing different features and some features (like science education) that are inside RRI according to some definitions (e.g. the six keys to RRI of the EU) are outside the concept in other definitions.

Ballung concepts are notoriously difficult to operationalize, also because operationalization requires first of all a clear definition. However, this does not imply that it is necessarily impossible to measure *specific* notions of RRI. Rather it becomes mandatory to make explicit what notion of RRI is operationalized and measured. Another consequence is that assessments based on different notions of RRI cannot be compared with each other. This is exactly what surfaces from Table 14.1. Different authors come with

different operationalizations of RRI depending on the specific aims they have in mind (implicitly or explicitly) for RRI assessment and depending on what context and unit of assessment they assume. The consequence is also clearly that RRI levels measured with one method cannot be compared with measurements from other methods (with the possible exception of methods 2 and 3).

One may also wonder whether it is always productive to exactly pinpoint more exact notions of responsible innovation.[4] While this may make the notion easier to measure, it might also mean that it loses its function as communicative device among diverse groups. The reason is that responsible innovation often functions as what has been called a boundary object (Star & Griesemer, 1989). Boundary objects are concepts that have a common meaning among groups, or across sites, but also have some interpretative flexibility, so that different actors can adapt them to their specific local needs. Due to this combination, they can help to foster communication and cooperation between groups. As a *Ballung* concept, responsible innovation may well function as a boundary object. However, attempts to completely specify its meaning might mean that it loses some of its interpretative flexibility and that some of the involved groups may no longer subscribe to it.

A second reason why RRI may be particularly difficult to measure and assess is that RRI interventions seem often aimed at somehow improving the innovation process from a societal point of view, rather than at attaining some predefined "absolute" level of "responsibility". Of course, for an innovation process to be "responsible", some common criteria apply, like the need to include stakeholders. But what is the "right" level of stakeholder involvement may not be the same for every innovation, or for each technical domain.

Two issues are at stake here. One is that what counts as more responsible (in terms of RRI) may depend on context. In one particular RRI project, stakeholder involvement may be a main way to improve RRI performance while in other cases the emphasis is on anticipation or responsiveness. Another one is that often RRI seems to be aimed at improving performance, and hence seems to be a moving target rather than a completely predefined notion. It seems like methods 6 and 7 in particular try to address this issue. They both contain questions, and are primarily aimed at learning rather than measuring predefined levels of RRI.

A third reason why RRI may be particularly difficult to measure is that it is a normative notion. This differentiates it from other complex notions (and *Ballung* concepts) like, for example, intelligence. RRI expresses what is desirable, not what is factually the case. What seems particularly relevant here is the so-called naturalistic fallacy, i.e. the impossibility to understand normative notions fully in descriptive terms. According to the philosopher G. E. Moore, descriptive definitions of normative terms are always vulnerable to what he calls the open-question argument (Moore, 1903). If we provide a descriptive definition of a normative term like good, we can always

ask the question: "It meets these and these descriptive requirements, but is it good?" Similarly, if we understand RRI fully in descriptive terms (or attributes), we can ask the question: "But is the innovation (process) really responsible?" It follows that if we are to measure a normative concept like RRI, at least some of the attributes need to be normative or involve normative judgments.

14.6 Strategic behavior

Performance measurement will almost always influence behavior, and in many cases it will invite strategic behavior (De Bruijn, 2007). This is most obviously the case if the performance measurement is directly connected to rewards or punishments. But even without such a direct connection, it is likely that performance measurement will have (behavioral) effects because (almost) nobody wants to perform poorly.

The fact that performance measurement affects behavior is in itself not bad; it may even be desirable or intended. After all, one reason to do RRI performance measurement may be to increase RRI performance. Nevertheless, it is important to realize that if we measure the RRI performance of an innovator, we do not just carry out a measurement but make an intervention. This intervention will have effects – desirable ones, but potentially also undesirable ones.

Moreover, the fact that there is not just measurement but intervention may affect the quality of the measurement. For one, it may affect the reliability of the measurement. This is most obviously the case if RRI performance measurement depends on self-assessment. Similarly, if the innovator is also the standard setter, there may be an inclination to choose RRI performance measures on which the innovator scores relatively well. In such cases, strategic behavior may potentially affect the chosen RRI construct, which may affect validity.

If the innovator is also the RRI assessor and/or the standard setter, there is a danger of *window dressing*. Window dressing is the phenomenon where an agent pretends to meet certain ethical standards (and makes efforts to show that) while in reality these standards are not met, or at least not to the extent pretended. Certain combinations of actor roles increase the risk of window dressing; in particular there is a risk of window dressing if the innovator is also the RRI assessor and/or is also the standard setter (Figure 14.1). This not to say that if actors combine these roles, it will always or necessary result in window dressing. An innovator may be genuine and even self-critical in an RRI self-assessment; it may potentially even be more critical than an outsider. Conversely, window dressing may also occur with an external assessor and standard setter. Even if the innovator, assessor and regulator are different persons or organizations they may still be closely aligned.[5] There may be collaborations between these actors, or economic or political dependencies, or they may share mutual ideological commitments.

Such collaborations, dependencies and shared commitments are indeed quite common in today's innovations systems, as for example underlined by such notions as the triple-helix model of innovations, that assumes close collaboration between universities, innovating companies and governments (Etzkowitz & Leydesdorff, 1997). Still, *ceteris paribus*, the risk of window dressing seems lower when the different roles outlined in Section 14.2 are separated rather than combined. However, as Table 14.1 testifies, most current RRI assessments are based on self-assessment and thus combine at least the role of innovator, or the one being assessed, with that of RRI assessor.

Also, when innovators cannot influence the construct chosen or the measurement, RRI measurement and assessment are likely to have behavioral effects. More specifically, innovators may make efforts to score better on the measured RRI attributes. This is in itself of course not undesirable, and in many cases, it is even desirable. However, in some cases, it may have detrimental effects.[6] This is particularly the case if the chosen RRI construct is not completely valid. In that case, an increase in performance in terms of the chosen RRI attributes may not signal a real increase in RRI performance. For example, certain RRI aspects may not be included in the measured attributes because they are difficult to measure. It is perfectly conceivable that increased performance in the measured RRI attributes goes hand in hand with decreased performance in aspects of RRI that are not measured, so that an increase in measured RRI performance does not signal an increase in the "real" RRI performance of the innovator.

Such effects may be particularly apparent if RRI performance measurements come with strong incentives. The effect will be, in general, that what is incentivized is the particular construct of RRI that is being measured. As long as construct validity is high, this is not a big problem. But since, as we have seen, RRI is a complex notion, it is not unlikely that what is being incentivized is actually a particular, somewhat narrow, interpretation of RRI. While strong incentives make it more likely that the particular construct of RRI that is measured is achieved, they also seem to make it likely that other aspects of RRI – that are not being measured and incentivized – are ignored. The reason for this is that resources are limited, so that an increased performance in some respects is likely to come with less attention for, or even decreased performance in, other aspects.

Incentivizing RRI may be problematic for other reasons as well. One such reason is the phenomenon known as "crowding out" (see e.g. Gneezy, Meier, & Rey-Biel, 2011). Crowding out occurs if an intrinsic motivation to achieve some good (in our case RRI) is replaced by external incentives, so that the initial intrinsic motivation decreases, or even disappears. For example, if sustainable behavior is incentivized with financial incentives, people may start to behave sustainably because of the financial gains rather than because they believe it to be good to behave sustainably. This may "crowd out" their intrinsic motivation to do good. Paradoxically, the effect may be that they

start to behave less sustainably (cf. Gneezy & Rustichini, 2000). There are different possible explanations for this phenomenon (Gneezy et al., 2011). One possible explanation is that financial incentives may erode social norms. Another possible explanation is that by introducing (financial) incentives, people start to see sustainable behavior as something that can be traded for financial gains (and losses) rather than as something done for intrinsic moral reasons. As a consequence of this change in perspective, they might be willing to bear the financial costs that come with unsustainable behavior, for example because it brings them comfort, while – previously – the unsustainable behavior might have come with a moral guilt, that they wanted to avoid. So, incentivizing may sometimes make moral goods tradable, so that they lose their special moral status.

Crowding out may also be a concern when contemplating assessment and incentivizing schemes for RRI. There may be many motives for an innovator to engage in RRI, but the desire "to do good" is certainly one of them. On the other hand, the currently limited uptake of RRI suggests that moral motivations alone may not be enough, and that some form of incentivizing with connected RRI assessment or measurement schemes may be required, also to overcome some of the barriers for RRI uptake (Gurzawska et al., 2017). Nevertheless, an awareness of the risk of crowding out would be helpful in shaping assessment and incentivizing schemes for RRI.

14.7 Discussion

We have seen that measuring and assessing RRI performance are not straightforward. RRI is a complex and multidimensional notion. Moreover, RRI is what has been called a *Ballung* concept, a concept without clear borders, and such concepts are notoriously difficult to operationalize and measure. These measurement problems are further aggregated by the fact that there is not an agreed definition of RRI and that RRI is a normative notion, not just a descriptive one.

Navigating the various pitfalls of RRI assessment is not straightforward. Choices or directions that avoid certain pitfalls may well increase the likelihood of other pitfalls. Two tensions stand out in particular. One is that between reliability and validity. In order to make RRI measurement and assessment more reliable, one might want to aim for RRI attributes that are (more) objectively measurable. However, since RRI is a complex and normative notion, a focus on only attributes that can be objectively measured is likely to decrease construct validity as it will leave out aspects of RRI that are less tangible or more contextual, but are not less important. The other tension is that between avoiding window dressing and avoiding crowding out. The risk of window dressing can be reduced by a clearer division of roles and also by focusing more on the quantitative, easily measurable aspects of RRI. Both, however, may well increase the risk of crowding out. A sharper division of roles may make RRI an external obligation or constraint, rather

than something that is aimed at for intrinsic reasons. Also a focus on objectively measurable numbers seems to increase the risk of crowding out, certainly if it is coupled with predictable rewards and punishments.

How are we to move forward given these pitfalls and dilemmas? Without suggesting that all dilemmas can be avoided, I think that major headway can be made by distinguishing more clearly between different rationales for doing RRI assessment. The rationales I have in mind are learning, accountability and incentivizing. I will explain these in more detail below and will argue that these three rationales are very hard – if not impossible – to combine in one RRI assessment method. As a consequence, in developing and deploying RRI assessment we are best advised to deliberately focus on one of these rationales, rather than trying to combine all three in one RRI assessment approach or tool. This does not rule out the use of *different* RRI assessment approaches (or tools) for different rationales, but it makes it advisable not to try to combine different rationales in one RRI assessment approach (or tool).

Learning is aimed at improving RRI performance through gaining new insights and learning new skills. For example, monitoring the efforts of certain RRI actions can provide insights in what RRI actions are (cost) effective and so help to improve RRI performance. But learning may take others forms as well. It may lead to a greater awareness and sensitivity of RRI issues. Learning may also relate to the ability to work with different stakeholders and the ability to connect research and innovation to the values, expectations and needs of society (cf. Klaassen et al., Chapter 9). Also what has been called second-order learning (Argyris & Schön, 1978) is relevant. While first-order learning is about learning how to better achieve given goals, second-order learning is learning about what goals to achieve and puts in question existing value and belief systems. If RRI assessment is to support second-order learning, it should probably leave room for changing goals and perspectives rather than being based on pregiven targets.

Accountability may take different forms, like for example showing that resources for RRI have been spent well, or that the organization meets a certain minimum level of RRI performance or that it can explain its choices in innovation to stakeholders. In all cases, accountability typically involves an external agent to which the innovator is accountable. Moreover, accountability in most cases assumes clear standards or expectations against which the innovator is held accountable.

Incentivizing is, like learning, aimed at improving the RRI performance of the innovator, but whereas in learning the (implicit) assumption is that the innovator is intrinsically motivated to improve RRI performance and thus wants to learn, the incentivizing rationale assumes that an (external) incentive scheme is required to motivate the innovator to do RRI and to improve RRI performance. This external perspective is somewhat similar to that in accountability, but whereas accountability is mainly backward-looking (accounting for what one has done in the past), incentivizing is forward-looking (i.e. incentivizing some future behavior).

Although learning, accountability and incentivizing may all three be proper rationales for doing RRI assessment, they seem hard to combine because there are fundamental tensions between them:

- *Learning versus accountability*: learning typically requires openness for failures and deviations; accountability makes it much harder to recognize these. In terms of RRI assessment, learning is often served by self-assessment, by subjective items that require judgment and by context-specificity, while accountability typically requires independent assessment, objectively measurable indicators and comparability (between contexts).
- *Learning versus incentivizing*: learning assumes an intrinsic motivation to do RRI and to improve; incentivizing assumes external motivation and lack of improvement without incentives. Incentivizing may lead to crowding out of intrinsic motivation and thus may diminish and undermine learning.
- *Incentivizing versus accountability*: the contrast is perhaps less stark than in the other two cases, but there are still potential tensions. Accountability requires objectively measurable indicators and comparability (between contexts); it will often lead to incentivizing what can be measured (and compared) rather than the underlying aim. Accountability may well lead to window dressing rather than a real improvement in RRI performance

Table 14.2 provides a summary of the differences between the three potential rationales for RRI assessment, including their main differences and tensions. Within each rationale, more specific aims may be formulated, like the aims mentioned in Table 14.1. It is important to note, however, that seemingly the same aim may serve different rationales. For example, an aim like monitoring RRI performance, as mentioned in Table 14.1, may be relevant in all three rationales. However, what is a good way to monitor RRI and what are appropriate RRI assessment approaches and tools heavily depend on the underlying rationale. If the underlying rationale is learning, some form of self-assessment and subjective items that require judgment, as for example in approaches 6 and 7 in Table 14.1 would be perfectly appropriate. However, for accountability these would be inappropriate methods, and approaches like 2 and 3 in Table 14.1 would be much more suitable.

The important lesson that can be drawn from this is that in developing and applying methods for RRI assessment one should not only be clear about what the aim of a specific method is, but should also be clear about the underlying rationale (learning, accountability or incentivizing). If one is vague about the underlying rationales or makes an attempt to combine different rationales, it is much more likely that an RRI assessment method is developed or applied that is not fit for purpose.

Table 14.2 Rationales for Responsible Research and Innovation (RRI) assessment

Rationale	*Main aim of RRI assessment*	*Assumed motivation for RRI*	*Self-assessment*	*Objectively measurable indicators*	*Context*
Learning	Improve RRI performance through learning	Intrinsic	Possible and even desirable	Not necessary	Need for context-specificity
Accountability	Show compliance and reliability to outside world	Can be both	Undesirable	Preferable if not required	Need for (some) comparability between contexts
Incentivizing	Improve RRI performance through external incentives	Extrinsic	Possible but usually undesirable	Preferable	Need for (some) comparability between contexts

14.8 Conclusions

The increased attention on RRI has led to proposals for assessing the RRI performance of innovators. Several methods and tools for RRI assessment are now available. I have discussed and illustrated several pitfalls of such methods. Some of these pitfalls are more specific to RRI, like the lack of a uniform definition and the normative character of RRI; others are more general, like the risks of window dressing and that of crowding out intrinsic motivation. Some pitfalls also have a dilemmatic character, in the sense that they relate to different requirements for RRI assessment that are difficult, if not impossible, to combine. To navigate these dilemmas, I have argued that it is best to start from the underlying rationales for doing RRI assessment. I have distinguished three such rationales – learning, accountability and incentivizing – and I have argued that these three rationales are by and large at tension with each other. The conclusion is that if one wants to develop and deploy RRI assessment one should make an explicit choice of one of the rationales rather than trying to serve all three at once. This is even the case if one's assessment method is geared towards an aim, like monitoring RRI performance, that is seemingly important for all three rationales.

Acknowledgment

This chapter has been written as part of the PRISMA project that received funding from the European Union's Horizon 2020 research and innovation program under grant agreement no. 710059.

Notes

1 This is similar to the definition of direct stakeholders in value-sensitive design; see e.g. Friedman, Kahn, and Borning (2006).
2 In particular the aim of assessment was not always explicitly mentioned, as well as who was the (assumed) assessor. In most cases, both could reasonably be derived from the further description or context.
3 These methods are typically aimed more at learning than at assessing RRI levels (and accountability). I will return to this issue in more detail below.
4 I would like to thank one of the reviewers for drawing my attention to this point.
5 I would like to thank one of the reviewers for pointing this out.
6 The point is similar to what is known as Goodhart's law in economics, which has been paraphrased by Strathern (1997: 308) as "When a measure becomes a target, it ceases to be a good measure."

References

Argyris, C., & Schön, D. A. (1978). *Organizational learning*. Reading, MA: Addison-Wesley.

Bradburn, N. M., Cartwright, N. L., & Fuller, J. (2017). A theory of measurement. In L. McClimans (Ed.), *Measuremnet in medicine. Philosophical essays on assessmemt and evaluation* (pp. 73–88). London: Rowman & Littlefield.

Carmines, E. G., & Zeller, R. A. (1979). *Reliability and validity assessment*. Beverly Hills, CA: Sage Publications.

De Bruijn, J. A. (2007). *Managing performance in the public sector* (second edition). London: Routledge.

Dreyer, M., Chefneux, L., Goldberg, A., von Heimburg, J., Patrignani, N., Schofield, M., & Shilling, C. (2017). Responsible innovation: A complementary view from industry with proposals for bridging different perspectives. *Sustainability*, *9*(10), 1719.

EPSRC. (2019). Anticipate, reflect, engage and act (AREA). Retrieved from https://epsrc.ukri.org/research/framework/area/

Etzkowitz, H., & Leydesdorff, L. A. (1997). *Universities and the global knowledge economy: A triple helix of university–industry–government relations*. London: Pinter.

European Commission. (2012). Responsible Research and Innovation. Europe's ability to respond to societal challenges. Retrieved from http://ec.europa.eu/research/science-society/document_library/pdf_06/responsible-research-and-innovation-leaflet_en.pdf

European Commission. (2014). Rome Declaration on Responsible Research and Innovation in Europe. Retrieved from https://ec.europa.eu/research/swafs/pdf/rome_declaration_RRI_final_21_November.pdf

Flipse, S. M., Dam, K. H. V., Stragier, J., Vrielink, T. J. C. O., & Sanden, M. C. A. V. d. (2015). Operationalizing responsible research and innovation in industry through decision support in innovation practice. *Journal on Chain and Network Science*, *15*(2), 135–146. doi:10.3920/JCNS2015.x004

Friedman, B., Kahn, P. H. J., & Borning, A. (2006). Value sensitive design and information systems. In P. Zhang & D. Galletta (Eds.), *Human–computer interaction in management information systems: Foundations* (Vol. 5, pp. 348–372). Armonk, NY: M.E. Sharpe.

Gneezy, U., Meier, S., & Rey-Biel, P. (2011). When and why incentives (don't) work to modify behavior. *Journal of Economic Perspectives*, *25*(4), 191–210. doi:10.1257/jep.25.4.191

Gneezy, U., & Rustichini, A. (2000). A fine is a price. *The Journal of Legal Studies*, *29*(1), 1–17. doi:10.1086/468061

Gurzawska, A., Mäkinen, M., & Brey, P. (2017). Implementation of Responsible Research and Innovation (RRI) practices in industry: Providing the right incentives. *Sustainability*, *9*(10), 1759.

Keeney, R. L. (1992). *Value-focused thinking: A path to creative decisionmaking*. Cambridge, MA: Harvard University Press.

Kroes, P., & van de Poel, I. (2015). Design for values and the definition, specification, and operationalization of values. In J. van den Hoven, P. E. Vermaas, & I. van de Poel (Eds.), *Handbook of ethics, values, and technological design: sources, theory, values and application domains* (pp. 151–178). Dordrecht: Springer Netherlands.

Lubberink, R., Blok, V., van Ophem, J., & Omta, O. (2017). Lessons for responsible innovation in the business context: A systematic literature review of responsible, social and sustainable innovation practices. *Sustainability*, *9*(9), 721.

Moore, G. E. (1903). *Principia ethica*. Cambridge: Cambridge University Press.

Oftedal, G. (2014). The role of philosophy of science in Responsible Research and Innovation (RRI): The case of nanomedicine. *Life Sciences, Society and Policy*, *10*(1), 5. doi:10.1186/s40504-014-0005-8

Owen, R., Stilgoe, J., Macnaghten, P., Gorman, M., Fisher, E., & Guston, D. (2013). A framework for responsible innovation. In R. Owen, J. Bessant, & M. Heintz (Eds.), *Responsible innovation* (pp. 27–50). Chichester: Wiley.

Ravn, T., Nielsen, M. W., & Mejlgaard, N. (2015). *Metrics and Indicators of Responsible Research and Innovation*. Retrieved from www.rri-tools.eu/documents/10184/47609/MORRI-D3.2/aa871252-6b2c-42ae-a8d8-a8c442d1d557

Stahl, B. C., Obach, M., Yaghmaei, E., Ikonen, V., Chatfield, K., & Brem, A. (2017). The Responsible Research and Innovation (RRI) maturity model: Linking theory and practice. *Sustainability*, *9*(6), 1036.

Star, S. L., & Griesemer, J. R. (1989). Institutional ecology. Translations and boundary objects: Amateurs and professionals in Berkeley's Museum of Vertebrate Zoology, 1907–39. *Social Studies of Science*, *19*(3), 387–420.

Stilgoe, J., Owen, R., & Macnaghten, P. (2013). Developing a framework for responsible innovation. *Research Policy*, *42*(9), 1568–1580. doi: http://dx.doi.org/10.1016/j.respol.2013.05.008

Strand, R., Spaapen, J., Bauer, M. W., Hogan, E., Revuelta, G., Stagl, S., … Pereira, Â. G. (2015). *Indicators for Promoting and Monitoring Responsible Research and Innovation. Report from the Expert Group on Policy Indicators for Responsible Research and Innovation*. Retrieved from: http://ec.europa.eu/research/swafs/pdf/pub_rri/rri_indicators_final_version.pdf

Strathern, M. (1997). 'Improving ratings': Audit in the British university system. *European Review*, *5*(3), 305–321. doi:10.1002/(SICI)1234-981X(199707)5:3<305::AID-EURO184>3.0.CO;2–4

van de Poel, I., Asveld, L., Flipse, S., Klaassen, P., Scholten, V., & Yaghmaei, E. (2017). Company strategies for Responsible Research and Innovation (RRI): A conceptual model. *Sustainability*, *9*(11), 2045.

Van den Hoven, J. (2013). Value sensitive design and responsible innovation. In R. Owen, J. Bessant, & M. Heintz (Eds.), *Responsible innovation* (pp. 75–84). Chichester: Wiley.

Von Schomberg, R. (2019). Why responsible innovation? In R. Von Schomberg & H. Jonathan (Eds.), *International handbook on responsible innovation* (pp. 12–33). Cheltenham (UK): Edward Elgar.

Wickson, F., & Carew, A. L. (2014). Quality criteria and indicators for responsible research and innovation: Learning from transdisciplinarity. *Journal of Responsible Innovation*, 1(3), 254–273. doi:10.1080/23299460.2014.963004