Intelligent Evaluations

12.1 Background

According to Mulgan (2014), there is a need for a new discipline that helps us to be collectively intelligent about our own collective intelligence. In bringing in this metal-level perspective, we are challenged to ask what this implies. If we look to individual learning for inspiration, we know that metacognition, or the ability to choose efficient learning strategies and evaluate your own individual learning, is essential among good learners (Flavell, 1979). At a collective level, the processes of planning, monitoring, and evaluating collective work will be equally important. Collective metacognition has been proposed as a term that describes how collective intelligence can think about its collective intelligence (Schuler, 2015).

Because of the digitization of information and the online setting, evaluations are also infiltrating almost every area of human life. The first section of this chapter describes the rise of the reputation society, which centers on evaluating persons. Online reputation is now not only important in human work settings, but it is at the center of our lives in social media. Individuals are constantly getting feedback from others in the form of quantifiable ratings of different activities.

The second section focus on evaluation of collective work. Digital technology makes it possible to design metacommunicative feedback loops in most group work and organizational work, sometimes labeled as triple loop learning (Tosey, Visser, & Saunders, 2012). This section discusses the potential of implementing a more systematic level of metacommunication in collective work.

The third section addresses intelligent evaluations in the political domain. One could claim that evaluations are at the core of any well-functioning democratic system. If we look back in time, the ancient Athenians were the first to institutionalize evaluations with the

nomothetai. Today, the Citizens' Assembly in Ireland is a fascinating new way of institutionalizing citizen evaluations in democratic systems.

12.2   The Reputation Society

12.2.1   The Emergence of the Reputation Society

In recent years, scholars have suggested that that we are moving into a new type of reputation-centered society, largely triggered by how reputation systems have become important in an increasingly number of different online environments. In contrast to an offline setting, online individual reputation becomes visible in a new way when it is assumed to be measurable as a reputation score (Gandini, 2016).

Because social and socioeconomic interactions require trust, reputation can help actors make decisions in situations where they do not have direct knowledge of other persons. In online settings, trust becomes even more important when people do not meet in the same physical co-located setting. We cannot rely on local knowledge or word of mouth. If we do not have any previous knowledge of persons, this creates uncertainty whether they actually are reliable. Online reputation system has been designed to facilitate trust or remove the need for it. Reputation scores are typically used to assess how trustworthy individuals are, building on various types of digital data, both active user data like ratings and reviews, but also passive user data like interaction histories. Algorithms and metrics automatically aggregate the data into a one-dimensional quantitative score that describes an individual's trustworthiness (A. Wilson & De Paoli, 2019).

In a historical perspective, this new reputation economy emerged with the growth of a largely individualized workforce of knowledge workers, the freelancers and self-employed workers who very early began to engage in online social networks. In the late 1990s, the notion of self-branding and self-promotion began to spread as key activities for career development. Cultivation of a professional image became essential in the new knowledge economy, and the increasing popularity of social media amplified the importance of these self-branding practices. In this context, the notion of reputation takes a prominent position as a shared cultural conception of value that bridge the offline and the online setting. It becomes strategically important to manage reputation in the network of professional contacts, as it is decisive to get jobs. This new source of trust is not only reliant on interaction in an offline setting (Gandini, 2016).

In the online setting, reputation was reinvented as a new type of social capital by imitating the logic of trustworthiness used in the Google PageRank algorithm. The huge success of this search engine builds on an analogy between hyperlinks and academic citations; the idea that the "citation of the Internet," the link, was the most informational resource. The PageRank calculated the number of links pointing to one page from other pages. In a similar way, reputation was recreated as a performance metric that could calculate a reputation score by informing about the trustworthiness of individuals in a very easy and reliable way. Although reputation across online environments can be regarded as the digitization of word of mouth, it is also an economic asset. Reputation becomes an object, a form of individual social capital that includes both offline and online networks, and it represents an investment in social relations with expectations of economic return and future job procurement in an increasingly freelance-based labor market (Gandini, 2016).

Today, the number of freelancers in the economy is increasing, with as many as one of seven workers in the UK. In addition, wide varieties of economic transactions are now dependent on reviews and feedback systems with elaborate rankings and reputation scores of various sorts. This includes an increasing number of sharing economy sites within holiday accommodation like Airbnb and online retailers such as eBay. In these systems, personal reputation functions as a networked asset that favor some persons in economic transactions of information, services, or goods (Gandini, 2016). Most of the online systems that use reputation scores build on an economically orientated competitive logic, with an emphasis on methodological individualism.

These systems draw on data about a user's activities to produce information about that individual's standing in the online community. They resemble the point systems and leaderboards in online games, where the "capital" is the opportunities for gamers to "spend" these points in different ways within the game-world. The main difference from a game is that the scoring metric will typically combines many different types of data. Data can be generated directly from user's activities, such as frequency of visits, how much time they spend on the site, how many transactions they complete, how many contributions they make to a discussion, how many network ties they have, and so on. It typically also includes how others rate the contributions, through likes, up- and downvotes, or more specific assessments regarding helpfulness, reliability, promptness etc., and qualitative feedback like review comments (A. Wilson & De Paoli, 2019).

The metrics may be different depending on the online site. It often intends to serve as proxies for prior experience and personal knowledge that can predict future behavior. The different factors in a reputation system will typically be used to generate a numerical measure of the user's overall behavior/reputation/ranking within the relevant community. Reputation scores are aggregates or averages and the data can be weighted in a range of different ways. Scores are often also made public to other community members, so individuals can make decisions whom they want to interact with. In other cases, the scores will only be available to site administrators (or an automated process) and allow privileges or give access to services within the space (e.g., using star ratings or badges) (A. Wilson & De Paoli, 2019).

When reputation systems are intended to support transactions of a trading nature, an entity's reputation score depend on customer feedback about reliability, product quality, speed of response, etc. In other areas, these scores can include the number of contributions and other users' explicit evaluation of individuals. Users can judge others in a range of different ways, by awarding stars, writing feedback, favorite, up- and down-vote. Consequently, one can both accumulate and lose reputation, often spend it, and sometimes even speculate on it. However, it is difficult to defend against unfair assessments or being able to explain choices. Once a seller receives a negative rating, this can easily lead to more negative ratings because these systems tend to amplify biased up- or downward spirals (A. Wilson & De Paoli, 2019).

Furthermore, studies on eBay find that giving feedback is not motivated by altruism, but by an expectation of reciprocity. Users lose interest in receiving feedback once they have accumulated experience and a good reputation score. Then they will no longer need to elicit ratings form others by rating them (A. Wilson & De Paoli, 2019).

Reputation manipulation is also a significant threat against these systems. Unscrupulous participants may find ways to manipulate the reputation scores in dishonest ways so they can earn more money in e-commerce platforms. A well-established, high reputation can provide a better price. Similarly, participants may try to damage the reputation of others, leading to rivals losing customers (A. Wilson & De Paoli, 2019). When reputation is viewed as a currency or marketable commodity, it is exposed to the same problems as in financial markets, like questions of ownership, fairness, and control. Collusion is one threat, as a group of people contributes to boost or undermine a reputation score. Badmouthing can produce unfairly negative ratings and damage reputation, with negative economic

consequences. Bots have also been linked with manipulation of reputation scores (A. Wilson & De Paoli, 2019).

12.2.2   *Online Reputations Moving into New Domains*

The reputation score systems are becoming increasingly ubiquitous, as they move out of business sites, into an increasing number of new domains. For example, many expert Q&A sites, which are primarily a discussion forum instead of a trading environment, have begun to employ reputation systems so that users asking questions can judge the trustworthiness of an answer, or community members can build up their own reputation as experts. Here, high reputation scores is motivated by kudos and honor, for example by receiving badges of achievement in the community. This can be regarded as a type of gamification, which motivates knowledge sharing in the community (A. Wilson & De Paoli, 2019). StackOverflow is one such example of an expert Q&A site where programmers can ask and answer questions relating to technical issues. Users can up-vote and down-vote others' questions and answers. These actions not only organize what is visible, but they also contribute to reputation building. In general, there are many more ways to gain reputation than to lose it. However, one loses reputation points if a post is flagged as offensive or spam.

Points can also be "spent" (transferred to another user) in a bounty system for those seeking quick answers to complex questions (A. Wilson & De Paoli, 2019). In addition, points are converted into privileges at the site: for example, once a user has 15 points, they can vote up a question or answer; once they have, 20, they can talk in a chat; once they have 125, they can vote down questions or answers; and so on. At 1,500 points, users are allowed to add new tags to the site (questions are tagged as corresponding to particular topic areas, such as SQL or Java); at 2,000, users can edit other users' questions and answers. At 10,000 points users gain moderation rights; at 25,000, they have access to the site's analytics. Thus, there are incentives to build one's reputation that go beyond the acquisition of reputation for its own sake, (A. Wilson & De Paoli, 2019).

Another area for reputation score systems is scholarly social networks for scientists. These measure a scientist's reputation and scientific impact (e.g., altmetrics) in new ways (Fecher & Friesike, 2014). For instance, ResearchGate aims to be the prime deliverer of scholarly reputation by designing reputational metrics that builds on a richer

amount of data compared with only measuring the number of citations of an article, which has been the most important influence on a researcher's reputation regarding funding and career opportunities. Today, the scholarly workflow in the online setting of formerly hidden actions like reading, bookmarking, sharing, discussing, and rating are leaving traces online and offer new ways of measuring scientific impact. However, studies show that alternative engagement metrics, such as Q&A and follower data, can also lead to reputational anomalies (Nicholas, Clark, & Herman, 2016).

Proxy voting is another example of a reputation system in the political domain. In one proposed model, labeled liquid democracy, it is possible to transfer votes through a new type of flexible representation. In this alternative model, all members vote directly on all policy issues. Since the required voter competence is more demanding, voters can delegate their votes to a representative to vote on their behalf on a singular or several policy issues. These votes can even be further delegated to another representative, but can also be terminated by the original voter at any time. The voting model allows for area-specific representation instead of the typical political representation that intends to cover all policy issues. It enables voters to directly authorize experts to vote on their behalf. This allows for a more fine-grained account of political representation compared with the typical policy bundles political parties provide. In addition, the voting method is assumed to require a "reputation system" based on ranking and ratings. When experts have a transparent record of accomplishment, voters can easily assess the results of previous claims and use this information when they vote. Liquid democracy is proposed to be particularly relevant in legislative decision-making (Blum & Zuber, 2016).

In other areas a reputation score system might have potentially negative effects. Conventional reputation systems are loaded not just with the values they are designed for (trust, honest behavior, reliability), but also a more subtle value-system that is orientated towards a competitive, capitalist free market based on self-interested individuals. Individuals possess a capital that individuals can accumulate and lose it. This may be appropriate for an e-commerce website, but in other contexts, these systems may have potentially negative effects (A. Wilson & De Paoli, 2019). For example, in expert Q&A sites, knowledge sharing often builds on values like pride and commitment. There is less focus on promoting individual skills.

Although some of these sites operate with reputation scores, this is not necessarily at the expense of other actors who compete for the same resources. However, a potential disadvantage is that low reputation scores may have a negative psychological effect on individuals. One example is the discussion forums on the platform supportgroups.com, which is dedicated to users with financial problems, homelessness, anxiety, and other mental and physical health issues. Recently, it included an online reputation system so that users can acquire points when they publish comments in the forums. Since this environment attracts vulnerable individuals, reputation scores systems are particularly concerning since they can have a negative impact on users' self-esteem (A. Wilson & De Paoli, 2019).

In reputation systems that aim to create a supportive community of peers, it may be counterproductive to develop a quantifiable reputation system. Individualized ratings risk weakening community values and cooperative activities that are not measured by the system. While a market-based view of reputation may be desirable in a business-focused trading site, it may undermine important noncompetitive values on other sites (A. Wilson & De Paoli, 2019).

12.2.3   Reputation Score Systems in Social Media

Social media sites and professional networking sites encourage social interaction through likes, shares and mentions, and other comments. These statistical data are used to aggregate metrics, which also create individual reputation systems in the informal social domain. For instance, social influence today is increasingly regarded as something that can be measured through the number of followers you have on different sites. Reputation is becoming more important since we increasingly depend upon others to engage in transactions to employ us, befriend us, or listen to our opinions (Gandini, 2016).

In social media, people live continuously in the gaze of others through a range of informal assessments such as likes, friends, followers, and many other secret rankings. This system produces a stream of evaluative metrics that raise or lower one's social currency. As a consequence, self-presentation in social media has become an increasingly important part of people's life. The continuous "curation" of one's photos, comments, and profile with deletions, additions, and modifications, are all designed to maximize likes, being the core value indicator in this existential marketplace (Zuboff, 2019).

The most important psychological process in Facebook is "social comparison." It describes the influence from our social environment, when we tacitly apply evaluative criteria from our society, community, group, family, and friends. Ordinary young people are drawn into online communication that automatically triggers social comparison on an unprecedented scale. Both insecurity and anxiety increase when individuals constantly chase for positive feedback from others. The use of likes in social networks provide users with those variably timed dopamine shots, which further increase their efforts. A post with zero likes is not just privately painful, but it stands out as a kind of public condemnation. Still, most users are more eager for the reward than the fear of being humiliated (Zuboff, 2019).

In the social media life, there is no self independent of other's feedback. The likes provide a continuous assessment of one's value on the social market. In one study, one third of the women said that their biggest worry online was that they constantly had to compare themselves and their lives with others. The systems are designed to maximize the possible amount of users' time and consciousness, and the result are several types of emotional anguish such as addiction, boredom, distress, and isolation. Simple behavioral techniques are used, like variable reinforcement, which let the user receive small rewards every once in a while, in the form of likes and comments from others (Zuboff, 2019). Zuboff's descriptions may overemphasize the negative effects, but they are a reminder of the destructive effects reputation systems can have when they colonize new areas of human life.

In contrast, the original Internet, and some of the most well-known CI projects rely on what can be labeled as a deliberative reputational meritocracy. In these CI environments, a majority of individuals make minor contributions, while a small core does much of the work. These active contributors serve as leaders of the community and make the most important decisions.

In the online setting, these meritocracies originate from open source software communities (Castells, 2010). One example is the Linux operative system, where any change to the code of the central kernel can break the entire project. The founder Linus Torvalds and his "closest group" will decide which of the submitted modules are included in the upgrade of the software (Kittur & Kraut, 2008). Although Linus does not have a legal or technical authority, he has a persuasive authority. Anyone is legally free to do as they please, but the community is still built around a hierarchy of meritocratic respect, mutual recognition, and some kind of peer review

system (Benkler, 2006). Here, the skills count regardless of age, sex, race, position, or qualification. Everyone is given an opportunity in a decentralized system where recognition is based on what you do and not who you were (Levy, 2001). Likewise, in one study of the Apache server project, a core group of 15 developers contributed 88 percent of the new lines of code, but did only 66 percent of the bug fixes, which was a less interdependent task. This finding indicated that low coordination tasks such as bug fixes was done by many different contributors, while high coordination tasks such as strategic planning was done only by a small group. The leader group set direction and provided a structure to which others could contribute (Kittur & Kraut, 2008).

While the reputation systems in social media are part of an individualistic, accumulative, and competitive paradigm, some of the large CI projects, like Wikipedia, build on a peer production community that honors hard work. Here, individuals with many different backgrounds will interact and build a common identity through their shared passion. Advancement in meritocracies are based on performance, rather than wealth or social background. In these reputation systems, achievements are displayed on personal profile pages. One example is the gaming community Foldit, which provide a multitude of statistics on the gaming performances. The login information on the site encourages users to register so they can get credit for the volunteer work ("You are an anonymous user and will not get any credit for your contributions. Sign in now!").

In the same way as in research communities, part of the motivation is about gaining recognition by peers (Himanen, 2001). Both Foldit and Wikipedia illustrate that even very young persons, like teenagers, can do important work in these communities. The main distinction between different subgroups in the community will typically be between newcomers and old-timers. In Wikipedia, there will be thousands of informal leaders who work on separate articles, depending on who does most of the work. Over time, some of the most active contributors can choose to move into strategic roles in the community (Kittur & Kraut, 2008).

Wikipedia is also interesting because of how deliberation is an important part of how persons are evaluated in the community. This includes the process of selecting individuals to become administrators or Wikipedians in the community. When a person is nominated to become an administrator, the evaluative deliberation will last for seven days. Anyone can ask the candidate questions, but no person can ask more than

two questions each. This process let other editors get to know the candidate, and explore the candidate's motivation to become an administrator and if they understand their new role. An uninvolved third person, a bureaucrat, decides whether there is consensus to approve the request. The final judgement is not based exclusively on the percentage of support, but in practice one will usually need more than 75 percent of the votes because most candidates with less than 65 percent support are not approved ("Wikipedia:Administrators," 2020).

In contrast to the dominant trend of using reputation scores, Wikipedia still highlights "quality, not quantity." Because edits can vary in size and quality, edit counts are not an important part of the assessment ("Wikipedia:Edit count," 2020). However, all contributors will acquire a track record because it is easy to identify all previous actions in the environment. These actions will also be part of the assessment. Most active contributors will already have developed an informal reputation based on the work they have done. Some contributors even make their own personal profile pages that display the work they have done. Still, it is the deliberation and voting by members in the community that decides who is promoted to the most important roles in the community.

12.3   Evaluating the Collective Work

12.3.1   Shared Coordination

The digitization of information does not only open new ways of evaluating persons, but it also influences how we monitor and coordinate the collective work in a range of different ways. Some degree of coordination is necessary in all kinds of group work and evaluations an important success factor. From a theoretical perspective, intelligent evaluations will build on some type of metacommunication or metadiscourse, as it utilizes our human ability to talk about how we talk. This ability requires language and is likely to have played an important role in human evolution. Some even consider the ability to communicate about our own communication as a basic condition for successful human communication (Bateson, 1972). Explicit, shared coordination is important when small groups engage in complex tasks that require a high degree of synchronization between members. Studies from an offline setting also find that explicit metatalk is important in regulating small group work in professional settings (Baltzersen, 2013). It provides feedback loops that enable groups to evaluate their ongoing work. Here, collaborative

problem solving differs from human swarm problem solving in its emphasis on such processes, like the need for joint coordination (see Chapter 8). If we look to specific CI practices, virtual innovation teams illustrate how metacommunication can be performed in efficient ways. For example, a top solver claims part of the solution is to ask for a clarification of the problem:

> *When we started this recent challenge we asked to have a conference with the client. We asked very pointed, detailed and technical questions, so that we could understand exactly what it was that they were after, because the challenge was not written in great detail. They were sort of vague and you're like saying, "Well if you don't want to answer this question that's fine, but if you tell us what's your bottom line, what is it that you want to get out of this, what is your product, or what is your goal, it's going to be a lot better, because if we understand that then we will be able to provide you with that solution." So in terms of this one, once we met and spoke with the client, it became pretty clear to me. I was like, well, I didn't know what the solution is but I was pretty sure I could figure it out.*

This is an example of metacommunication in the initial phase of the problem-solving process. A discussion with the seeker in this phase can help the group better understand the problem, and thus increase the likelihood of solving it. In general, many studies point to the importance of discussing and establishing good group norms early in the problem-solving process, even if the meeting is short. Some studies report that it helps handing out a written description of the rules of the discussion and read aloud the rules (Fishkin, 2018: 176; Grönlund, Herne, & Setälä, 2015).

Other CI researchers also highlight that collective work in large organizations need to build on metacommunication or reflective communication (Mulgan, 2018; Schuler, 2015). These processes are often connected to different types of feedback, such as second-loop learning (Argyris & Schön, 1997) or triple-loop learning (Tosey et al., 2012), which raises the awareness of how organizations learn. The notion of triple-loop learning assumes that intelligence operate at multiple interconnected levels; the first loop uses existing models to process data and perform existing work efficiently, the second loop generates new relationships or new procedures, while the third loop creates new ways of thinking. Participants reflect on how they think about the "rules," not only on whether rules should be changed (Mulgan, 2018: 156, 237).

However, when the group size increases, it becomes increasingly difficult to coordinate and get an overview of everything that's going on. For

example, if only one person can talk at a time, it can be frustrating to be in a large group because it can take much time to let everyone be heard. When coordination requirements increase, this may reduce motivation. As Brooks Law states in the domain of software projects: "Adding manpower to a late software project makes it later." When the group size reaches a certain level, shared responsibility or control become a problem. The members will have to use relatively more time on procedural issues rather than substance or actually doing the task (Kittur & Kraut, 2008).

With the support of digital technology, direct coordination is possible in larger groups than what was previously possible. However, the exact threshold of the maximum group size is uncertain. A common rule of thumb for face-to-face groups is that the optimal group size is somewhere between five and ten people. If there are fewer, there is not enough benefit from diverse points of view. If the group size is above ten, coordination will take too much time. Even when groups make this extra effort, the difficulties of working together may outweigh the benefits of having more people (Malone, 2018: 184–186). Still, different digital tools can provide a better overview of the group work and make it easier for larger groups to work together in real time compared with what is possible through verbal discussions in an offline setting. Different tools can provide both qualitative and quantitative feedback that give a better overview of the ongoing work.

Several of the CI projects that encompass cognitive or informational diversity when they scale up in size face challenges with overview that needs to be tackled. In the Polymath 1 project that lasted 37 days, 27 individuals made 800 mathematical comments, in total they wrote 170,000 words. As the projects evolved, it became increasingly difficult and time-consuming for newcomers to join the project because of the amount of information they had to read.

Because the blog that was used in the Polymath project is a time-centric tool, new comments to a blog post were automatically listed below previous contributions. This chronological organization of the contributions made it gradually more difficult to get an overview of all the perspectives. Because the discussions were organized into several different blog posts with attached comments. The discussion is also becomes messier when both relevant and irrelevant comments are included.

One alternative strategy is to design statistical tools that can provide a simplified overview of the collective work (e.g., deliberation metrics). They can support coordination by making it easier for groups to evaluate their own work. Some CI research examine different types of metadiscourse

tools, like to-do lists and chatbots that can prompt different types of group evaluations (Young Ji Kim, Gupta, Glikson, Woolley, & Malone, 2018).

In argument mapping, metadiscourse tools, termed deliberation analytics, aim to provide better overview of all the contributions in a large group. The system mines the traces of the group's activity and generate customized metrics that can give both the participants and moderators a better overview of how the map is evolving. Fisrt, the topology of the argumentation map (e.g., breadth and depth of the branches of the map) provides information about the maturity of the deliberations. This is a better proxy than metric algorithms like word frequency statistics. Second, the metrics notifies participants about issues they may want to resolve based on their previous interests, which both include their viewing activity and the content of their contributions. They can then choose to either rate the comment or add a new post. Users with similar topical interests are also clustered together in an attempt to motivate them to collaborate on a branch in the tree structure. Third, dysfunctional communication can be identified through a social network analysis of the interactions in the deliberation map, if there are tendencies toward groupthink (Fujita et al., 2017; Klein, 2012, 2017). This example shows how digital tools can provide support for formative assessments, making it easier to monitor the ongoing collective work.

Wikipedia illustrates how the evaluation of the collective work can utilize crowdsourcing methods. Every article has a talk page that enables a written metadiscourse of the collective work. The modularization of each article allows for a myriad of "content-focused" discussions on different topics. Participants can choose to only discuss the content and not write about it. These discussions can be regarded as open conversations about our own culture that anyone can join.

However, studies show that most of the work on a Wikipedia article will usually be coordinated by a small number of contributors. They typically solve the complex, interdependent tasks, for example on how to structure or organize the article so it becomes more cohesive. Explicit coordination is primarily valuable in work on articles when there are few contributors. It is usually more important in the early life cycle of an article when the direction is more uncertain and open-ended. There are also many simple, stand-alone tasks, which require little coordination. This can be tasks like fixing grammar, combating vandalism or creating links. (Kittur & Kraut, 2008).

In addition, there are a huge number of other special pages dedicated to discussing Wikipedia policies and technical issues. These wiki pages were

not part of the original design of Wikipedia, but they have gradually emerged in line with community needs. Although many of these pages are only of interest to the Wikipedia community itself, they comprise the evaluation policy that ensures the sustainability of the online community (Nielsen, 2011: 52–53; Rijshouwer, 2019). Wikipedia is a particularly interesting case because it shows that citizens cannot only be knowledge producers, but also successful evaluators of their own collective work.

12.3.2   The Need for Coordinators

Another strategy in evaluating collective work is to utilize some type of centralized control, like appointing a leader or establishing a small core group that coordinates the larger group. Most of the CI projects have a coordinator. The titles vary, being a moderator, facilitator, or copilot, but they all aim to organize the problem-solving process in an effective way.

Most intermediaries in online innovation contests also use a facilitator to support both the seekers and the solvers. In Topcoder, every project is assigned a copilot who works with the seeker. The copilot manages the logistics, answer technical questions, and help the seeker in producing a realistic project plan. This involves giving an accurate description of the challenge, making sure that all deliverables are received and that the review process is done in a proper way (Topcoder, 2019b). The copilot is an elite member, and needs to have won a minimum of three challenges (Topcoder, 2019c). He earns money if projects are on time, and the outcomes are delivered with high quality (e.g., $40 for one challenge and $600 for another challenge) (Jefts, 2018). In *Innocentive*, the PhD-educated facilitators are primarily a support for the seekers, helping them to formulate the problem in an appropriate way while the facilitator in IdeaConnection is supposed to support the solver team in their work (see also Chapter 8). In this type of collaborative problem solving, the facilitator will help the team to do the work within the deadline and not stray off the topic.

In other CI projects, the moderators act more like project leaders. For example, in the Polymath project, Gowers, the founder, has usually been responsible for organizing the academic discussions. Successful Polymath projects have required a project leader to moderate and guide the discussion, and generally to keep the momentum going (Michelucci & Dickinson, 2016). In the first Polymath project, Gowers acted as a moderator, but there were few problems with internet "trolls" or people persistently posted distracting comments. Nor were well-intentioned but

unhelpful comments a significant problem (Gowers & Nielsen, 2009). However, being a moderator can be very time consuming, and only a few people have done it so far. This is why there is usually only one Polymath project a year (Michelucci & Dickinson, 2016).

Furthermore, the coordinator will need to solve conflicts between contributors. They help solve disagreements on what content should be in the specific articles. For instance, a typical conflict in Wikipedia can be that writers follow personal preference instead of adhering to community norms. The moderators who guide and help new contributors are called Wikipedians (J. Beck, Neupane, & Carroll, 2018). They have expertise about the community norms and do a lot of the maintenance work, which is vital for the sustainability of the encyclopedia. They do not necessarily write articles, but spend much time editing the content and turning it into a more coherent resource. They serve as "protectors" of the encyclopedia in the sense that they cope with vandalism or others who do not follow the norms of the community. It is also important to ensure that users follow citation rules and copyright rules. It is these persons who transform the encyclopedia into being something more than a broad collection of individual contributions (Algan et al., 2013; Benkler, 2006).

In many online communities, conflicts arise because of poorly defined policies. This may even involve conflicts between moderators. This is why communities like Wikipedia have procedures and policies on most activities, including how to resolve or manage online conflicts (J. Beck et al., 2018).

The role of the moderator in argument mapping is also very important since several hundreds of participants may be involved. They organize the debate and cultivate the discussions by deciding which claims are acceptable and which need to be improved. They guide participants and monitor debates for duplicate claims, "fake" contributions, or abusive content. As the debate grows, moderators will also sometimes have to reorganize the entire debate (J. Beck et al., 2018).

Furthermore, moderators ensure that new posts are correctly structured and that authors follow the map conventions. Sometimes posts will first be given a "pending" status, and become available after they have been checked by moderators. This ensures that the map is well structured. One study found that two moderators were able to handle nearly 200 active contributors, with most posts (~85 percent) requiring no or only minimal moderator support. In some argument maps, moderators can automatically be notified about a conflict, or where users have rated posts without reading the arguments. Because moderators represent a potential extra

cost, researchers are also exploring how one can crowdsource moderation work into a series of easy-to-do micro-tasks that every participant can do (Klein, 2017).

Although the role of the moderators is not to evaluate the merits of a post, some studies suggest that it is a challenge to take a "neutral" stance in the debate. Beck et al. (2018) identified adversarial beliefs and values as a common source of conflict between moderators. Some of these conflicts were not productive and undermined collaboration. In some cases, librarians were used to successfully strengthen the competence and position of the moderators. In general, these different CI projects indicate a need for coordinators because very few projects can rely only on self-organization.

12.4 Institutionalizing Critical Discourse

12.4.1 *The Nomothetai*

The use of evaluations is not something new in society. In ancient Athens, the citizens managed not only to invent democracy, but gradually they were able to improve these institutions by strengthening the critical and deliberative discourse. After the Peloponnesian War with Sparta, the Athenians briefly lost their democracy, but managed to reinstate it (in 402–401 BC). The citizens had experienced that a demagogue can win the votes in the Assembly regardless of the citizens' interest. Therefore, they established the nomothetai, a new institution that was devised to avoid this from happening again. In this new system, proposed changes in law, which was passed by the Assembly, could not become a law unless it was also approved by the nomothetai (Fishkin, 2018: 52–53, 203).

Nomothetai were probably recruited from the panel of 6,000 jurors who had sworn the Heliastic Oath. They were ordinary citizens picked by lot for a given day from among those who showed up. Their function was to examine proposals more closely than the Assembly could be capable of doing. The number of nomothetai varied according to the importance of the legislation proposed – probably at least 501, but for more important matters even 1,001 citizen jurors or more. A meeting lasted only a single day, and it is likely that the nomothetai could deal with more than one legislative proposal in the same meeting (Hansen, 1999: 167–169).

The laws were passed by a procedure analogous to a trial, hearing the arguments for and against the proposal. Legislation is assumed to be a revision or change of the law currently in force. A new legislative proposal is therefore regarded as an accusation against the existing laws. The author

of the proposal for change will first speak as the accuser of the existing laws. Afterwards, the five advocates chosen by the Assembly defend the existing laws. When both have spoken, the nomothetai decide by show of hands. If the majority supports the proposal for change, it becomes the law in force (Hansen, 1999: 166–169). Some claim that this type of critical discussion and questioning is the very essence of democracy and is the most important precondition for the overall growth of knowledge and development of a prosperous society (J. F. Mueller, 2018).

At a societal level, the nomothetai served the purpose of being a security or democratic brake that could restore order and potential ill effects of voting in the Assembly (Fishkin, 2018: 52–53, 203). A new multistage institution was introduced that could have a critical and evaluative function regarding decisions that other democratic institutions had made. As a result, legislation became less casual, and it reduced inconsistencies in the legal code. From this perspective, it is the political norms and the design of institutions that afford democratic conversation and critical discourse; democracy has less to do with social norms and informal conversations on a micro level (Schudson, 1997: 305).

12.4.2   The Citizen's Assembly in Ireland

The 2017 World Values Survey shows a worrying shift in attitudes toward democracy. More young people, in both Europe and the US, are skeptical of democracy as a governing model. There is more political apathy than previous generations. People are less interest in joining political parties and experience that the political elites have become more detached from the people (Foa & Mounk, 2016, 2017; Micu, 2018). People also have less faith in public institutions. For example, in 1964, 76 percent of Americans had faith in the government to do what is right, but in 2015, only 19 percent were of the same opinion (Micu, 2018). In addition to the lack of faith in political institutions, there is increasing concern about the dysfunctional engagement in the online setting. This situation calls for new ways of involving citizens that reduce polarization and strengthen consensus-building processes.

As the nomothetai institutionalized new types of critical discourse after a crisis in Athens, there is a need for new democratic institutions today that can perform intelligent evaluations. The Citizen's Council in Ostbelgien, Deliberative Polling in Mongolia, and the Better Reykjavík platform are all interesting new examples. However, the paradigmatic case is the Citizens' Assembly in Ireland.

It is the first country in the world to hold three national mini-publics in quick succession. A representative sample of 99 citizens from the population are invited to meet and discuss important constitutional questions or complex political issues over a longer period. They are selected through random sampling from the whole population in the country. These assemblies have been established by Parliament, and they also make their final recommendations to the same institution. In the Irish context, deliberative democracy is being implemented as part of the wider political system in a systematic manner because of these citizens' assemblies (Farrell, Suiter, & Harris, 2019).

Most of the issues that the assemblies have worked with have later been set out for mass voting through a national referendum. Both the first Convention on the Constitution (2012–2014) and the Irish Citizens' Assembly (2016–2018) were essential in supporting national referendums for constitutional change, legalization of same-sex marriage in 2015, and removing the constitutional ban on abortion in 2018 (Farrell et al., 2019). This Citizens' Assembly was assigned to deliberate on five issues: abortion, the aging population, fixed-term parliaments, organization of referenda, and climate change policies. There was international pressure to change policies concerning abortion and climate change. Opinion polls also showed strong support for a liberalization of Ireland's abortion laws (Farrell et al., 2019). On highly contested political issues, a simple aggregation of votes through a referendum might not be the best option because many citizens will want to debate the issue. In the case of same-sex marriage and abortion, the Assembly helped break a political deadlock and were important in establishing public acceptance for change (Devaney, Torney, Brereton, & Coleman, 2020).

Input Legitimacy

If we look closer at the success behind the Citizens' Assembly in Ireland (2016–2018), several features ensured the legitimacy of the process. First, input legitimacy is crucial to ensure trust both in the political system and among the wider public. The government established the Assembly and invited citizens were randomly selected from the wider population. In total, 99 citizens and 99 substitutes were selected. Unlike the first Assembly, no politicians participated because they wanted to distance themselves from the controversial abortion issue.

The members were stratified according to sex, age, social class, and region. However, some reported too many participants coming from urban areas, which limited the discussion of climate challenges in rural areas. This

illustrates the importance of inviting a large enough group to ensure sufficient representation.

Another challenge was the large turnover, with more than 150 individuals needing to be recruited during the 15-month period. Since the members did not get an honorarium, this had a negative effect on attendance (Devaney et al., 2020).

Throughput Legitimacy

Regarding throughput legitimacy, it is important that the discussions in the Citizens' Assembly are organized in a fair way. There were monthly weekend meetings. During these meetings, members used much of the time to discuss issues in groups of seven to eight persons. The group members were rotated, so every individual had to discuss issues with many different persons. The participants reported about challenging discussions with individuals who held other viewpoints (Devaney et al., 2020; Farrell et al., 2019).

Diverse groups are important because attitudinal change is more likely to happen when being exposed to views that are different from your own. Like in Deliberative Polling, a facilitator helped the group stay focused, be respectful towards each other, and ensure that every member had an equal opportunity to speak. These sessions were also closed, with no cameras or recording, so individuals could feel safe to state their opinion (Devaney et al., 2020; Farrell et al., 2019). The participants report being very satisfied with the format of the meetings, especially that the group rules ensured fairness, civility, and equality in participation. In addition, the participants were given some individual reflection time. A secretariat was also established to coordinate the process in a proper way (Devaney et al., 2020).

Before the meetings, briefing material was sent out that intended to be as objective as possible. During the meetings, legal, ethical, and medical experts had presentations, and when abortion was discussed, there were presentations by advocacy groups and personal testimonials by women (Farrell et al., 2019). The participants were satisfied with the presentations and praised those speakers who exemplified success stories from their own countries (Devaney et al., 2020).

In general, the participants experienced a significant level of learning by being part of the process. However, the group was given only two weeks to discuss a broad and complex question such as "How to make Ireland a leader in tackling climate change," and the participants report that this was too little time. If the topic had been more specific and involved a cost analysis, they suggest that the recommendations could have been more

realistic (Devaney et al., 2020). These statements suggest an awakening of a citizen responsibility and an increased understanding of the complex trade-offs that need to be made in politics.

Furthermore, the wider public was invited to interact with the Assembly. All expert presentations were live streamed to strengthen public involvement and transparency in the process. A strong media presence amplified public awareness, and a dedicated website provided public access to all expert content, papers, and public submissions. On the topic of climate change, the public sent in more than 1,000 submissions, including 150 group submissions (including from nongovernmental organizations, sectoral interests, and representative groups). This shows a strong public engagement.

Although the participants were given some time to read the submissions, it is more uncertain to what degree they are able to integrate these comments in the deliberation. In the evaluation of the process, the participants suggest that a summary of the submissions would have been better to read. Nor were the submissions part of the presentations. However, the size of the feedback from the wider public illustrates the potential in connecting the citizens' assemblies with the wider society. If this is done more systematically, it provides an opportunity to enhance environmental literacy in the wider society. Environmental literacy seeks to empower citizens to make responsible lifestyle decisions. By more strongly involving the maxi-public in the process, it seems possible to design political systems that can promote mass deliberation, strengthen the communication with citizens, and motivate more sustainable behavioral change (Devaney et al., 2020).

Output Legitimacy
Output legitimacy is dependent on how the final recommendations are used in the wider political system, such as a mandatory follow-up from dedicated parliamentary committees. To complement this bottom-up form of governance, participants find it necessary with clear top-down political engagement to create policy coherence. This involvement also ensures that citizen efforts are honored.

In the Ireland case, an all-party parliamentary committee was established after the Citizens' Assembly to respond to the recommendations, and it delivered a report one year later in 2019. The committee was not obliged to pursue the recommendations, but it still endorsed most of the recommendations, with the exception of the proposed increase in carbon tax. To a significant extent, this report shaped the development of the

government's Climate Action Plan, published the same year. The process illustrates how citizen involvement can be connected to the political decision-making system in new ways (Devaney et al., 2020).

It is also interesting that the random sample of citizens were able to move beyond self-interest to engage in collective decisions for the greater public good, which is a typical criticism against climate engagement (Devaney et al., 2020). Obviously, output legitimacy is threatened if the politicians do not accept recommendations they do not like or if they choose not to respond to the recommendations at all. Recommendations made by randomly selected citizens, who are neither elected nor experts, can raise accountability issues (Devaney et al., 2020).

Furthermore, in making the final report on climate change, the participants found it difficult to rank the 13 recommendations on climate change since they found all of them to be important. They suggested that their recommendations be complemented by further expertise, cost assessments, and evidence-based input, since they were not experts in the area (Devaney et al., 2020).

The Citizen's Assembly as an Intelligent Evaluation

These deliberative forums can provide valuable information about the citizen opinion on a political issue. These mini-publics serve the purpose of being "trusted information proxies" that can establish a more efficient communication between elected politicians and the public. Here, it is essential that the participants perceive the information dissemination, like the briefing material or presentations, as legitimate. If the whole process and the final vote is legitimate, it can help politicians make tough political decisions, for example on climate change, which require trade-offs and public support for action. From one perspective, the final voting on recommendations can even be regarded as an informed "micro-referendum" that can potentially lead to better decisions (Devaney et al., 2020).

Moreover, a Citizens' Assembly will provide insight into how people speak about the climate crisis, including their local concerns and shared values. This is different from an opinion poll or referendum, which aggregates opinions without deliberation. If the wider maxi-public is engaged, this process can shape public opinion, not just be used to develop a specific public policy. Deliberative processes are especially important in addressing complex public problems by involving citizens in the decision-making (Devaney et al., 2020).

A very important political goal in the climate crisis is to engage the public more strongly. Citizens' assemblies represent an interesting new

way of involving citizens in building consensual solutions on environmental problems. Environmental literacy is not only about the dissemination of correct values and beliefs, but it requires a dialogue with audiences of different persuasions, knowledge, and levels of engagement. Identifying citizens' opinions on climate change can help politicians engage in dialogues that are more constructive. Those individuals in Ireland who have participated in the Citizens' Assembly also think this institution should become a regular part of a democratic system (Devaney et al., 2020).

In 2020, a new Citizens' Assembly in Ireland has been established to advance gender equality by bringing forward proposals that challenge the remaining barriers and attitudes that facilitate gender discrimination. Because of COVID-19, the meetings have been held online, and all video presentations are publicly available on YouTube (Farrell, 2019). This third Citizens' Assembly will provide important knowledge on how a new modern "nomothetai" can be organized through online communication.

12.4.3   Knowledge Commons

A significant challenge when implementing citizens' assemblies is how to cope with manipulative misinformation in the public sphere (Devaney, 2020). Therefore, the collective problem-solving capacity of this type of democratic institution depends on the quality of the publicly available knowledge. In a polycentric democracy with a range of different smaller institutions, like mini-publics and engaged maxi-publics, it is important that objective knowledge is accessible to everyone (J. F. Mueller, 2018). A strong knowledge commons with dedicated open access policies is crucial in facilitating societal innovation.

Already one hundred years ago, John Dewey underlined that print was necessary to create a true public, but it could still only serve democratic conversations. Democracies put a lot of effort into writing to secure, verify, and make public decisions. This greatly enhances the capacity of public memory and makes democratic talk civil, since democratic conversations will be oriented towards the explicit and transferable communications found in print (Schudson, 1997: 305). If we look at the face-to-face conversations in the citizens' assemblies, they also end up with a vote and a final written knowledge product that can be further used.

Many of the examples in this book show examples of how our collective memory is being strengthened through the digitization of information. According to Mulgan (2018: 170), any CI system is reliant on a rich informational knowledge commons. Commons are shared resources that

are free for anyone to use, like water, forests, libraries, science, and also the Internet. In the offline setting, commons are usually vulnerable to over-exploitation of resources if property rights are absent. However, this is not a problem with digital information because it can be copied infinitely at a low cost. Today, the Creative Commons licenses institutionalize the flexibility of being able to modify knowledge products without needing to ask the original creator for permission. When products are instantly and easily modifiable, it is much easier to add value to the work through sustained collective efforts. A knowledge product can be adapted and modified by others into many different versions, whether this is an open textbook or a Wikipedia article.

An increasing number of policies today belong to the knowledge commons. As previously mentioned, this includes open access policies in science and open textbooks, both of which exemplify open knowledge production. Free textbooks can provide knowledge that is more easily accessible and can recruit more readers. Open data is another example of how more people can be given access to valuable information. Collective problem solving becomes more efficient when it draws on a body of common knowledge.

However, the largest video-sharing platforms today have commercial ownership. YouTube, the largest video platform in the world, shares video content openly, but they can also charge money for usage. These videos comprise the dominant cultural archive of our time, being an increasingly important provider of our cultural heritage. Until now, this platform has supplemented the role of state-based cultural archiving institutions like public libraries and museums, as well as media companies and broadcasters who want to involve users in new ways. However, because YouTube is a commercial enterprise and not a public one, they have no obligation to store these data in a way that serves society for the best. Until now, there has been little debate of the long-term implications of letting commercial spaces be responsible for some of the functions that public cultural institutions previously held. Today, libraries cannot re-archive material on YouTube, because of legal barriers such as copyright law and YouTube's Terms of Use. This issue becomes even more important when we know that an uneven or deliberately reshaped collective memory can have significant influence on people's conception of history and justice, and one therefore needs to ask who should be in control of it. For instance, people have used YouTube to publish eyewitness videos documenting conflict and human right abuses. Because YouTube frequently removes violent or otherwise "inappropriate" content, there is a risk that this sometimes may involve content of historical

or political significance. In one case, footage of the Syrian conflict was removed because it was flagged as violent, but after protests from human right groups, it was later restored. States are today increasingly asking whether they should rely on YouTube, Vimeo, or Facebook to keep public records (Burgess & Green, 2018: 137–139).

Digitization of information has opened up a major movement towards a strengthening of our collective human memory. In the long term, this can lead to more innovation, but it appears that collective learning at a national level and global learning depend on the prevalence of a strong knowledge commons.

12.5   Summary

We are moving into the age of intelligent evaluation. Traces of online activities provide unprecedented opportunities to examine our personal behavior, our collective work, and discuss our political systems. At a group level, metacommunication is essential for good communication and these mechanisms can now be scaled up in new and innovative ways. On the positive side, the massive amounts of data provide an opportunity to discuss, deliberate, and learn. The digitization of information provides many interesting new ways of evaluating collective work. Feedback loops can support groups in coordinating problem solving more efficiently. Although third-loop learning in organization can perhaps not be planned, there are now digital tools, like argument mapping, that make it possible to support such reflective communication in new ways. The examples also show that most groups still need a coordinator to help organize the collective work.

Furthermore, online reputation score systems make decisions about whom to trust. In the sharing economy, these systems build on crowd assessments. They have been used to regulate sales and transactions successfully. Services are rated and reviewed, and these evaluations make it easier to find the right experts for the right job. The main attraction is definitely the time efficiency and simplicity of just having to rely on one simple quantifiable score. Many online communities also operate with lists of "top contributor" or "top reviewers" to motivate contributions.

Traditionally, reputation has been considered an "intangible asset" that is very difficult to measure, but this view is now changing. The emerging reputation society builds on the premise that reputations actually can be measured through online reputation scores. However, there is a concern that algorithms and metrics produce unfair simplistic evaluation scores,

and we know little of potential negative individual effects of being a "low achiever." There is increasing concern around the negative effects of being evaluated all the time. In addition, the scoring systems can be manipulated, and it is possible to attack others' reputation. Furthermore, there is a growing awareness that the reputation system is overly centered on an individualistic, accumulative, and competitive paradigm. Nevertheless, reputation scores are being taken in use in new areas that have nothing to do with economic transactions.

In social media, evaluating other persons has become the new normality. It has led to increased psychological stress as people become more obsessed by constant social comparison. When living in the gaze of others, it is all about getting likes and followers in social media, risking the escalation of a self-interested narcissistic individualism. Zuboff (2019) warns that social media life is becoming a collectivist life in a human hive where individual autonomy is lost. Humans follow the group pressure of the herd and the computational certainty of the "smart" solutions that can replace deliberation and democracy. It is the youngest members of our societies that are most at risk as these destructive mechanisms foster them from early age.

The social media systems are designed to tempt persons to constantly rate each other. The like button also provides the most valuable behavioral surplus, as metacommunicative data are the most powerful predictors of human behavior. The more a user presses the like button, the more information Facebook receives about a person (Zuboff, 2019).

On the more positive side, deliberative reputational systems are assumed to potentially democratize and decentralize power mechanisms in society. The talk page on Wikipedia demonstrates how the crowd can engage in important critical discussions of the content in one of the most important knowledge sources of our time. Independent of social background, the most active members can gain respect and become leaders if they make important contributions. Quantitative ratings matter less. Peers evaluate each other through the transparent traces of their work, and together they formulate the community policy.

Furthermore, the Citizens' Assembly in Ireland shows how deliberative democracy can be utilized in new ways. Citizens are challenged to discuss highly contested political issues such as abortion and climate change. The deliberation period is usually at least one year and is key to transformative change. It shows how governments can get access to the opinions of an "informed public." Many of the activities are also shared openly with the general public, who are invited to engage with the Assembly. An important

goal is to strengthen political interest on the issue in the broader community. (Devaney et al., 2020). A strong knowledge commons can provide a better basis for making informed decisions in this type of participatory governance.

A more fundamental question concerns the number of evaluations we need in a democratic society. Because we can collect more data, on both persons and work processes, does this also imply that we should do more evaluations? Can it be intelligent not to evaluate? Currently, the simple quantitative ratings are colonizing our society. There are infinite opportunities to produce evaluative data, and surveillance capitalists have learned to profit from it. These evaluations are primarily based on machine learning and hidden algorithms, which raises a number of questions: Who should design the evaluation? Who should have access to the evaluation results? Who should perform the evaluation?

However, in a democracy built on evaluation, it is not the quantitative ratings that really matter, but rather the tough and unpleasant critical discourses between citizens with different views. John Dewey once claimed that politics should be treated as a scientific evaluation. After implementation of a policy, the effects need to be evaluated and if the results are unfavorable, policies must change. Citizens were essential in evaluating these policies through voting in periodic elections, public opinion polling, and by giving public comments on proposed regulations (Anderson, 2006). Today, new types of intelligent citizen evaluation are being invented, such as the Citizens' Assembly in Ireland, which can likely strengthen democratic institutions in the future.