Metrics for Reflection in Distributed Information Processing

Asimina Mertzani* and Jeremy Pitt

*corresponding author: asimina.mertzani20@imperial.ac.uk

Abstract

The performance of a Distributed Information Processing (DIP) unit in collectively answering a question using the Regulatory Theory Social Influence (RTSI) can be evaluated against eight criteria, which are four conflicting pairs of systemic drivers. To maintain the balance between these drivers, the DIP must reflect and deliberate upon its own performance – in effect, using RTSI to answer questions asked by and about itself. In this position paper, we consider the drivers as dimensions and propose a set of possible metrics for 'measuring' the performance of a DIP. However, we also discuss some issues in the careless application or interpretation of metrics which might adversely affect the reflective process. We conclude with a discussion of what might be the fundamental conflicting pair of drivers in metrics-based reflection for DIP: quality vs. dignity.

Introduction

A Distributed Information Processing (DIP) unit (Nowak et al., 2019) is a collection of socially-networked but otherwise distinct, autonomous and heterogeneous components, whose task is to process information (signals) received individually and use it to produce a collective output, for example an answer to a question (see Figure 1). This output might be, for example, which candidate to appoint to a role (if the DIP is a social system and the components are people in a committee), an average value of a sensed signal (if the DIP is a cyber-physical system and the components are sensors in a sensor network), or the relative fairness of a policy for resource allocation (if the DIP is a socio-technical system and the components are both people and computer processes).

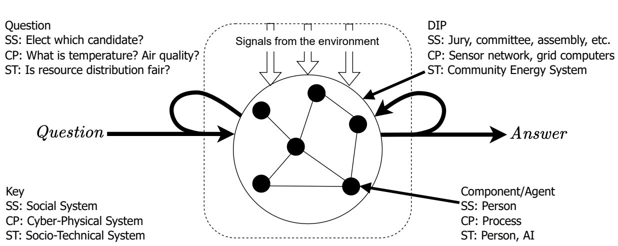

Figure 1: A Distributed Information Processing (DIP) Unit

The Regulatory Theory of Social Influence (RTSI) (Nowak et al., 2019) is a psychological theory that examines the role of social influence and social networks in distributed information processing in social systems, and proposes three distinctive features. These are, firstly, that social influence is bidirectional: just as sources seek targets to influence, targets also seek sources by whom to be influenced. Secondly, that as well as exchanging opinions, agents also exchange information processing rules (i.e. they can learn and improve). Thirdly, the theory presupposes that each agent is trying to maximise individual cognitive efficiency as well as producing the "best" (or "good enough") outcome for the collective: therefore, if the cost of communication is less than the cost of cognitive effort, some agents will delegate information processing to agents perceived as "better" at the task.

An algorithmic model of answer-generation based on RTSI has been operationalised in a self-organising multi-agent system for deciding the fairness of a resource allocation from a common-pool (Mertzani et al., 2022). This model implements *reactive RTSI*, as the DIP reacts to a question being asked, and each agent in the system changes its attributes (e.g. their attitudes to neighbours in their social network, and their own processing rules) in reaction to a comparison between their own answer and the answer generated by the system (i.e. the DIP).

The operational model of RTSI in a cyber-physical DIP also considered that the performance of the DIP over time (i.e. in response to a series of questions) could be evaluated according to eight criteria, which were actually four pairs of conflicting systemic drivers: accuracy vs. economy, stability vs. flexibility, inclusivity vs. expertise, and coherence vs. diversity (cf. (Dryzek and Pickering, 2017)). In each pair, each element is pushing in opposite directions, and for effective operation the DIP needs to maintain a balance between each of these four driver pairs, as depicted in Figure 2.

To determine whether or not this balance is being maintained, the DIP needs information about and judgement on itself, i.e., the DIP must engage in reflection (Landauer and Bellman, 2016) and reflective governance (Dryzek and Pickering, 2017) to answer questions asked by and about itself.

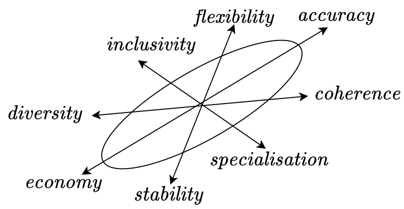

Figure 2: The drivers of the DIP

Therefore, the DIP needs a set of metrics to "measure" its performance in each of the eight "dimensions", and it needs a process to deliberate upon these measurements. This process could also be implemented using RTSI: this in effect makes the DIP a cybernetic system (Ashby, 1952) with a reactive part and a reflective part, as illustrated in Figure 3.

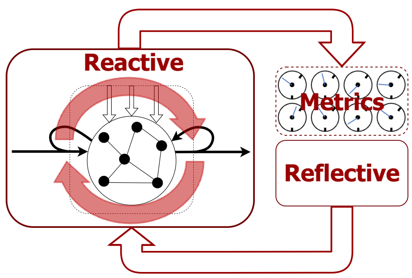

Figure 3: The DIP as a cybernetic system

However, as a stepping stone to *reflective RTSI*, this position paper considers the issue of defining the metrics which will provide the input to the RTSI algorithm. We consider the drivers as dimensions and propose a set of possible metrics for 'measuring' the performance of a DIP. However, we also discuss some issues in the careless application or interpretation of metrics which might adversely affect the reflective process. From this discussion, we conclude by considering what might be the fundamental conflicting pair of drivers in metrics-based reflection for DIP: quality vs. dignity.

Evaluation, Metrication and Reflection

Previous work (Mertzani et al., 2022) focused on the implementation of the *reactive* RTSI which is dealing with the reaction of the agents to the changes in the system attributes. However, it has been demonstrated that complex dynamic institutions are prone to get to catastrophic trajectories, if they deny or fail to change (Dryzek and Pickering, 2017). To manage to maintain the balance between the conflicting drivers and remain sustainable, such dynamic systems need to develop the ability to respond effectively to the signals from their environment. Therefore, they not only need a process for reacting to the changes but also a mechanism for reflecting these changes upon themselves.

As a result, to leverage the agents to do reflection and deliberation, we need to include in the system design informative metrics that provide feedback to the agents regarding the eight dimensions of interest. Besides that, according to Condorcet's ideals (Lopez, 2020), it is impossible for a group of heterogeneous individuals to coordinate without universal metrics that measure the dimensions of interest. As a result, many different metrics for 'measuring' the performance of a multi-agent system have been proposed. Some of them measure quantitative elements of the system while some other focus on the qualitative attributes (Ciobanu, 2006; Krol and Zelmozer, 2008). Those metrics are trying to provide information regarding dimensions of the system such as communication, autonomy or flexibility and are measured using elements such as number of connections between the agents in the network, connectivity, clustering-coefficient, and reachability (Ali et al., 2017).

Many multi-agent systems which use electronic institutions, as a set of mutually-agreed, conventional rules, have to determine (for themselves) whether or not the institution is *fit for purpose*, i.e. that it is satisfying its 'mission statement'. For example, if the mission of an institution is intended to achieve a fair and sustainable distribution of some common pool resources (cf. (Ostrom, 1990)), then it needs to be evaluated to decide if that is the case or not. In social systems, institutional evaluation is often considered to be a matter of *procedural justice*; inspired by this, a framework of procedural justice for electronic institutions was proposed, using metrics to measure the performance according to principles of participation, transparency and balance (Pitt et al., 2013).

However, based on the observations in (Mertzani et al., 2022) and the aim to optimise the performance of the DIP with respect to the eight evaluation criteria (Figure 2), we need metrics to measure the performance with respect to accuracy, economy, flexibility, stability, coherence, diversity, inclusivity and specialisation. Starting with stability, this metric is trying to measure whether agents manage to operate efficiently regardless of the circumstances. Therefore, we will follow the definition proposed by Ashby (1952), who posits that stability is a field property and that to measure stability in a dynamic system we actually need to observe the trajectories of the field of interest in space. On the other hand, the DIP needs to know whether there are many different clusters of opinion in the population and whether they diverge significantly or not. Therefore, a metric for coherence in the form of opinion similarity and divergence between different groups is required (Nowak et al., 2019).

Moreover, to avoid group-think and clique emergence, the agents should be able to observe whether the sources are distributed and whether different opinions are taken into consideration over time. To optimise the system performance with respect to the conflicting drivers, the DIP should delegate the most of the work to the experts (Mertzani et al., 2022) but also reassure that all agents engage in the collective process. Specifically, the distribution of work is expected to follow the Price Law, meaning that half of the work is done by the square root of the population (experts of the population on that cognitive task) while the other half is delegated to the remaining agents. To measure performance with respect to these dimensions of the system, metrics for specialisation and diversity, in the form of expertise and influence should be included in the systemic design, as depicted in Table 1. Furthermore, the DIP not only needs to be stable but also has to be flexible and adapt to the different circumstances. The ability to change and the ability to stay the same are clearly conflicting tendencies.

Dimensions	Metrics	Purpose
Accuracy	Performance	Accuracy with respect to expense
Economy	Expense	Processing Cost
Coherence	Cohesivity	Network clusters' similarity and divergence
Diversity	Influence	Distribution of sources
Flexibility	Flexibility	Adaptation to the different conditions
Stability	Stability	Effective functionality
Inclusivity	Participation	Equal participation
Specialisation	Expertise	Emergence of expertise

Table 1: Multi-agent dimensions mapping to metrics

A Basket of Metrics

This section proposes a metric for each of the dimensions of the multi-agent system, as summarised in Table 1.

A Metric for Performance

In this context, the metric for performance is concerned with measuring the accuracy of the collective opinion of the DIP on the fairness of the lived experience, in relationship with the expense of resources for achieving that accuracy, given the relative significance of getting a correct answer to this question (i.e. the fairness of the resource distribution process). This is a subjective judgement that describes the amount of resources it is worth spending on getting an acceptably accurate answer, taking into account how important it is to get the "correct" answer to that question.

This metric aims to inform agents about the amount of resources spent to get an answer and whether they spent too much on getting the answer to something not really important, or whether they spent too little that they couldn't get an acceptably accurate answer on something significant, or whether they cannot achieve an acceptable accuracy no matter what.

To that end, the performance metric $\mathcal{Q}_t$ could be defined as the accuracy of the collective answer acc_t divided by the sum of resource expenditure $sumres_t$ as a proportion of total resources $allres_t$, given a relative significance σ in epoch t:

$$\mathcal{Q}_t = \sigma(\frac{acc_t}{sumres_t/allres_t}) \tag{1}$$

One and a Half Metric for Influence

The Influence metric is concerned with the number of agents seeking to delegate to sources instead of performing the task themselves, and the number of agents being sources influencing others. This would allow the observation of whether trust is being put in the experts, which on the one hand can be a sign of optimisation for resource conservation by delegation of cognitive processing to agents that are perceived as more knowledgeable, but on the other hand might lead to group-think or exclusion.

Therefore, influence $\mathcal{I}_t$ could be defined as the number of agents seeking for sources $targets_t$ divided by the number of the agents being asked $sources_t$ in epoch t, given by:

$$\mathcal{I}_t = \frac{targets_t}{sources_t} \tag{2}$$

However, to observe whether some agents affected more agents we can define an auxiliary (Price) metric, that of $\mathcal{I}_t^{(max)}$, which identifies the proportion of sources in the DIP influenced half the total number of agents.

A Metric for Cohesivity

Cohesivity is concerned with the deviation of processing rules within community. Using this metric, we intend to provide feedback to the agents regarding the interconnection between clusters from a network perspective (i.e. agents belonging to sets of agents with many interconnections as presented by Emmons et al. (2016)) and clusters from a societal perspective (i.e. agents using the same processing rules).

To that aim, cohesivity $\mathcal{C}_t$ is defined as the average "in-group" divergence of processing rules across the different clusters, $\bar{d_{in}}$, divided by the average "out-group" divergence of processing rules between each cluster and agents from other clusters ,$\bar{d_{out}}$, and is given by:

$$\mathcal{C}_t = \frac{\bar{d_{in}}}{\bar{d_{out}}} \tag{3}$$

Specifically, $\bar{d_{in}}$ computes the divergence between the processing rules of the agents in a cluster (e.g. divergence of processing rules between agents of cluster X, between agents of cluster Y, and so on) and is averaged across clusters. Metric $\bar{d_{out}}$ measures the divergence of the processing rules of the agents between the agents within a cluster and all the other agents (e.g. the divergence of the processing rules of agents in cluster X with the processing rules of all the agents not belonging to cluster X, the divergence between agents that belong to cluster Y with those that not belonging to cluster Y, and so on) and is averaged across clusters.

To keep the population coherent and avoid fragmentation and conflict, it is desirable to maintain low $\bar{d_{in}}$ and low $\bar{d_{out}}$, which would indicate low divergence or in other words some form of alignment or consensus. However, to allow innovation and change overtime, it is desirable to maintain diversity, therefore $\bar{d_{in}}$ should be slightly lower that $\bar{d_{out}}$, so that there is some form of agreement within the agents of a clusters and some relative variance between agents belonging to different clusters.

A Metric for Expense

In this context, there are two forms of expense in the system, the one being the expense in communication (to get the information from the social network) and the other being the expense in cognitive processes (to find the information). Although in large-scaled networks communication might be very expensive, we assume that in small and medium-scaled networks the cost of performing a cognitive task, $cost_{cogn}$, is greater than the cost of communicating, $cost_{com}$.

The total expense $\mathcal{E}_t$ of the DIP could then be defined as the sum of the number of agents delegating the task $targets$ multiplied by the cost of communication $cost_{com}$ and number of agents performing the cognitive task $sources$ multiplied by the cost of cognition $cost_{cogn}$ given by:

$$\mathcal{E}_t = targets * cost_{com} + sources * cost_{cogn} \tag{4}$$

Two Metrics for Expertise

Expertise refers to the sources of influence which might either attempt to persuade and modify the beliefs of others (Afshar and Asadpour, 2010), or be identified by others as the ones that can assist them to accomplish the collective goal (Cacioppo and Petty, 1984). From a functional perspective the DIP need to be able to identify the expertise, to delegate the task to the ones that can get the correct answer, in order to improve the quality of the final outcome by managing to get an accurate opinion and to achieve resource conservation through delegation. Additionally, expertise should change over time to adapt to dynamic environments where the knowledge is varying over time.

Therefore, we can identify two possible metrics of expertise, the short-term expertise E^s, which reflects the expertise of the last x epochs, where x is a short-term period with respect to the agents' life expectancy and total number of epochs, and the long-term expertise E^l, which reflects the expertise of the last $n * x$ epochs, where $n > 3$.

In particular, short-term expertise E^s measures the number of agents being sources the $m\%$ of the last x epochs, where $m > 50\%$, while the long-term expertise E^l measures the number of agents being sources the $m\%$ of the last $n * x$ epochs, where $m > 50\%$.The desired situation is that E^s remains high, which indicates that expertise emerges and remains the same in short-term, while the E^l remains low, which indicates that expertise changes in long-term.

A Metric for Participation

This metric tries to capture the quality and the quantity of participation. It is concerned with how many agents participate in the collective decision-making process by proposing an opinion (quantity of participation), how many of them always delegate the task to another, how many of them always do the task (which might indicate that they refuse to listen to other opinions or that they aim to direct the collective choice), and how many of them do not reflect the collective decision to themselves.

Since change is desired in dynamic institutions, agents need to change strategy (ask network/use own opinion) and their beliefs and processing rules; they have to ask different agents from the network in order to listen to different opinions and they need to be willing to participate because they understand the importance of the collective processes.

Therefore, having knowledge of the values of that metric, agents will be able to understand whether the collective output reflects the opinion of the majority and whether they need to modify the collective process so that it encourages more agents to participate or excludes the ones that do not follow good practices.

A Metric for Stability

The problem with determining stability, as proposed by Ashby (1952), is that it is a property of a field and not of a material body. Therefore he offers one definition of stability: "given the field of a state-determined system and a region in the field, the region is *stable* if the lines of behaviour from all points in the region stay within the region". In that sense, we are dealing not with points in a space but trajectories within that space. However, Ashby also noted that a change of some parameter can also change the field, on which stability depends. Therefore, DIP units are more likely to be *quasi-stable* (Welburn, 2013; Pitt, 2017), and so undergo periods in their trajectories where they are stable, experience a period of instability, and then return to a stable state, but with their control variables having different values.

A Metric for Flexibility

The counterpoint to stability is flexibility, and there are corresponding problems with trying to measure flexibility. In particular, flexibility has both a functional and a normative aspect: not only is a DIP able to change (functional) but also it *should be* capable of change (normative). In this way, it might be possible to distinguish between a DIP that has "stagnated" due to, for example, path dependency (Collier and Collier, 1991) (where present actions are constrained or eliminated as a consequence of past decisions), and a quasi-stable DIP which is experiencing a period of relative "calm".

Further work is required to define informative metrics for the stability-flexibility driver pair. However, we next consider some issues in the application and interpretation of metrics, especially in systems involving components with some awareness, intelligence or capacity for reflection.

Risks

In this section, we present some common issues in the interpretation of metrics, highlighting the difficulties of "getting it right" and the adverse consequences of "getting it wrong".

Getting it Right

Metrics one way or another will affect decisions (Hauser and Katz, 1998), they provide important information and ease understanding of problems and in most cases constitute criteria for decision-making (Patterson and Miller, 2012). Although metrics play a fundamental role in systemic self-improvement, wrong choices of metrics, misinterpretation and misuse of them are some of the common challenges encountered in cyber-physical institutions.

Initially, metrics are hard to define since they should be linked with the system characteristics. Choosing a metric from another system is not guaranteed to provide the same information about the present system. Also, metrics should be defined so that they really capture the information that they are supposed to describe. Since metrics affect decisions and actions, it is also important to design them in a way that they capture not only a single parameter but also the side information related with that parameter, which can be accomplished by defining auxiliary or multiple metrics. For instance, if something is affected by *a*, *b*, *c* and *d*, the choice of a metric that captures only *a* and *b* might lead to actions that solve *a* and *b*, but amplify *c* and *d*.

Another issue in metric interpretation is the fact that metrics can be self-referential. The definition of a metric might require the knowledge of the outcome of another metric or the combination of the knowledge of some other metrics. Therefore, metrics are interconnected, and consequently, the analysis of their results in order to take actions should take into consideration the values of all the different metrics.

Moreover, while the first step towards getting some understanding over the system is to identify the appropriate metrics to describe the present, the next step towards achieving organisational goals, such as sustainability or balanced tensions between different incentives, is to add some *meta*-metrics that describe the rate of change of the system and provide visibility over the intertemporal evolution of the corresponding observations. However, defining *meta*-metrics is challenging, while analysing and understanding them is even trickier especially if you are an internal observer (e.g. an agent). As a result, in many cases systems fail to adapt and maintain sustainability because the metrics that they use reflect only short-term effects.

While metrics are undoubtedly important, organisations many times become victims of those metrics and that is because they end up being obsessed with metrics instead of focusing on identifying the right metrics that provide them the desired information. As a result, individuals spend too much time and effort in finding ways to measure performance that they end up in a situation in which they disregard important things. Therefore, it is very challenging to identify the minimum required metrics that provide clearly the desired information, and avoid being a victim of these metrics.

And although identifying the appropriate metric is one thing, finding the appropriate way to use it is another, equally challenging, thing. The fact that a set of metrics is defined is not enough to guarantee sustainable self-improvement. The agents of the system need to have access to these metrics, the ability to interpret them and the willingness to adapt their behaviours and policies based on the feedback on those metrics.

Finally, to mitigate all the possible issues in metric definition, interpretation and application, metrics should also be changed over time even if the policy of the system is not modified. And this is because, first and foremost, if a wrong metric is chosen then by modifying this metric we might get the information we initially were aiming for, and also if a metric finally ends up constituting a target, modifying that metric might decrease this effect. Additionally, in dynamic institutions of dynamic populations, the change of the metrics is required to capture changes in the DIP, changes in the knowledge, and the changes in needs and practices.

Getting it Wrong

Everything that can be measured can be managed, goes one mantra. However, there are a number of potential issues with "measurement", including the tyranny of metrics, Goodhart's Law, vanity metrics, interpretation of metrics, faux leagues, and social credit systems. In this subsection, we briefly consider each of these issues in turn.

The tyranny of metrics (Muller, 2019) is the observation that a fixation on metrics in order to evaluate performance can distort and diminish the performance itself. Examples range from the trivial to the deadly serious: from footballers averse to riskier passes (but which might create more goal-scoring opportunities) in order to maximise their pass-completion percentage, through lecturers who avoid teaching challenging and difficult material to get higher student ratings (but produce lower student comprehension and competence), and onto surgeons who refuse to take on riskier and more complicated but potentially life-saving procedures to maintain their patient operation survival rank.

Moreover, there is Goodhart's Law (Goodhart, 1975), which formalises the experience that when a measure is used as a target it ceases to carry any meaning; and the quasi-quantum effect on human behaviour, where awareness of being measured affects task performance, as does being consulted (the eponymous Hawthorne Effect; rather like Isaac Asimov's fictional psychohistory, which supposed that the subjects of psychohistorical analysis were unaware of being analysed). Goodhart was concerned with monetary supply in economics, and Asimov was writing science fiction, but the effect can be observed in many domains. For example, h-index, as a measure of scientific productivity and achievement, was quite effective, until academics found out that appointments and promotions were influenced by their h-index. Behaviour changed: for example, CVs claiming citation count rather than scientific contribution as an ambition; the emergence of citation clubs formed of mutually self-referencing cliques, and undermining of the peer review process (essential to the scientific method) by reviewers accepting an article on the condition that it cited their work.

Another way a metric becomes meaningless is when it becomes a so-called vanity metric (Ries, 2011), i.e. a number that appears to be impressive but is relevant only to those whose are impressed by a number, and is not indicative of true performance. Arguably, this is now the fate of h-index: a high h-index is only impressive to those impressed by their own high h-index. Consequently, of precisely what a metric is an indicative measurement may morph over time. Again, the example of h-index is instructive: it is no longer a measure of scientific productivity or performance, but an indicator of network centrality and longevity (Sarigöl et al., 2014), i.e. some people are cited because they are popular and have been around a long time (the network effect applies here too: some papers are cited simply because they have been cited).

Therefore, not only can a metric measure something of no interest, but it can miss something in which there is interest. For example, in previous work, we were interested in measuring "civic participation" in a self-organising multi-agent system (Pitt and Ober, 2018), defined as each having a duty to share equally the making, adjudicating and enforcing of rules. Outliers for excess participation could be detected, but it was not clear whether or not this was due (again) to network centrality or refusal to release power; on the other hand under-participation could be a consequence of network isolation or a refusal to take responsibility, but could *not* be detected as an outlier by the statistical technique being used to analyse the measurements of participation.

Another problem is raised when metrics are used as the basis for a comparative ranking between peers arranged in supposed "league tables". Rather than a pairwise comparison which provides meaningful ranking over time (which is what league tables provide in the context of sport) ranking, for example, universities according to some specious metrics (e.g. as in the UK's costly and time-consuming "Research Evaluation Framework") does not provide anything useful (it is not like universities are trying to qualify for post-season or promotion, or avoid relegation). All it creates is a lot of sound and fury signifying nothing, and a perpetual cycle of self-congratulation and recrimination as individual institutions rise and fall (and, oddly enough, the powerful take credit for the former and blame others for the latter).

Finally, socially-constructed metrics of 'credence' or 'trust' between peers, such as those used in our experiments for the emergence of expertise in knowledge aggregation, could perhaps, if these opinions were to be aggregated themselves, be used as some sort of social credit system. In a social credit system, rewards are given to those deemed worthy, while punishments are handed out to those deemed unworthy. This could have many unintended and pernicious consequences, such as a tyranny of merit (Sandel, 2021) or the suppression of dissent or disobedience (Burth Kurka et al., 2018), which needed to expose a gap between the intention and the application of rules, for example, or when a 'ruler' is simply applying the rules wrongly (or retrospectively changing the rules after breaking them himself: the pettiness of would-be tyrants through history).

Summary and Conclusions

In summary, we have considered Distributed Information Processing (DIP) units from the perspective of cybernetic systems, with a reactive part and a reflective part (Ashby, 1952). We have previously implemented an algorithm based on the Regulatory Theory of Social Influence (RTSI) (Nowak et al., 2019) to do the reactive part; if the reflective part is to do information processing about the DIP itself (Dryzek and Pickering, 2017), then we can use RTSI for this deliberation as well.

For this task we need metrics, and we propose a number of metrics for measuring a DIP against eight performance criteria (which are actually four pairs of conflicting systemic drivers). We need this information if we are to maintain the tension between each pair of drivers, but also for the inclusion in the DIP of any machine learning component, such as an oracle. Then, as planned for future work, this oracle could, given system X with profile of agents $a_1, a_2, \ldots, a_n$ and dimension values $v_1, v_2, \ldots, v_8$ (as per Table 1), recommend a reflective action plan.

However, we also cautioned against some of the limitations of purely metrics-based evaluation. These cautionary words are important: some "things" can be measured but do not matter, and some "things" that matter cannot necessarily be measured: in particular, (human) *values*. Two (human) values that profoundly matter but cannot easily be measured are *quality* and *dignity*, which might be the fundamental conflicting pair of drivers in metrics-based reflection for DIP. Quality is about producing acceptable answers with acceptable cost, and it may be that answers to such subjective concerns can be addressed by some system of interactional justice (Pitt, 2017).

Dignity, on the other hand, is the cornerstone value of Ober's theory of Basic Democracy (Ober, 2017). It has two aspects, one positive and one negative. The positive aspect is that civic dignity is maintained when citizens are treated as equal and worthy participants in democratic processes; the negative aspect is that civic dignity is undermined when citizens are tricked into making decisions that, had they been fully informed, they would not otherwise have made.

Both aspects are problematic. The notion of 'treatment' is a subjective one: for some, a representative democracy that allows one vote every few years in a rigged and gerrymandered process using an impoverished winner determination method (e.g. plurality, or 'first past the post') is hardly 'dignified'. Equally, mis- or disinformation through social influence can make it easy to mislead citizens who are willingly misled; and against conformation bias the deities themselves contend in vain. Therefore, DIP might require external observers: the equivalent of the democracy index or independent factcheckers to provide a metric for dignity.

References

Afshar, M. and Asadpour, M. (2010). Opinion formation by informed agents. *Journal of artificial societies and social simulation*, 13(4).

Ali, A., Aslam, M., Janjua, J., and Chaudhry, M. (2017). Methodology for performance evaluation of distributed multi agent system. *The Nucleus*, 54(2):75–82.

Ashby, W. (1952). *Design for a Brain*. London, UK: Chapman-Hall.

Burth Kurka, D., Pitt, J., Lewis, P., Patelli, A., and Ekárt, A. (2018). Disobedience as a mechanism of change. In *12th IEEE International Conference SASO*, pages 1–10.

Cacioppo, J. T. and Petty, R. E. (1984). The elaboration likelihood model of persuasion. *Advances in consumer research*, 11:673–675.

Ciobanu, G. (2006). Coordination and self-organization in multi-agent systems. In *ISDA '06: Proceedings of the Sixth International Conference on Intelligent Systems Design and Applications - Volume 01*.

Collier, R. and Collier, D. (1991). *Shaping The Political Arena*. Princeton, NJ: Princeton University Press.

Dryzek, J. and Pickering, J. (2017). Deliberation as a catalyst for reflexive environmental governance. *Ecological Economics*, 131:353–360.

Emmons, S., Kobourov, S., Gallant, M., and Börner, K. (2016). Analysis of network clustering algorithms and cluster quality metrics at scale. *PloS one*, 11(7):e0159161–e0159161.

Goodhart, C. (1975). Problems of monetary management: The u.k. experience. *Papers in monetary economics 1975*, 1(1):1–20.

Hauser, J. and Katz, G. (1998). Metrics: you are what you measure! *European Management Journal*, 16(5):517–528.

Krol, D. and Zelmozer, M. (2008). Structural performance evaluation of multi-agent systems. *Journal of Universal Computer Science*, 14:1154–1178.

Landauer, C. and Bellman, K. (2016). Reflective systems need models at run time. In *Proceedings of the 11th International Workshop on Models@run.time*, pages 52–59.

Lopez, V. (2020). How the French Revolution created the metric system. *National Geographic*.

Mertzani, A., Pitt, J., Nowak, A., and Michalak, T. (submitted on March 2022). Expertise, social influence and knowledge aggregation in distributed information processing.

Muller, J. (2019). *The Tyranny of Metrics*. Princeton, NJ: Princeton University Press.

Nowak, A., Vallacher, R., Rychwalska, A., Roszczyńska-Kurasińska, M., Ziembowicz, K., Biesaga, M., and Kacprzyk, M. (2019). *Target in Control: Social Influence as Distributed Information Processing*. Cham, CH: Springer.

Ober, J. (2017). *Demopolis: Democracy before liberalism in theory and practice*. Cambridge: Cambridge University Press.

Ostrom, E. (1990). *Governing the commons: The evolution of institutions for collective action*. Cambridge, UK: Cambridge University Press.

Patterson, E. and Miller, J. (2012). *Macrocognition Metrics and Scenarios: Design and Evaluation for Real-World Teams*. London, UK: Ashgate Publishing Limited.

Pitt, J. (2017). Interactional justice and self-governance of open self-organising systems. In *11th IEEE International Conference SASO*, pages 31–40.

Pitt, J., Busquets, D., and Riveret, R. (2013). Procedural justice and 'fitness for purpose' of self-organising electronic institutions. In Boella, G., Elkind, E., Savarimuthu, T., Dignum, F., and Purvis, M., editors, *Principles and Practice of Multi-Agent Systems (PRIMA)*, volume 8291 of *LNCS*, pages 260–275. Berlin, Heidelberg: Springer.

Pitt, J. and Ober, J. (2018). Democracy by design: Basic democracy and the self-organisation of collective governance. In *12th IEEE International Conference SASO*, pages 20–29.

Ries, E. (2011). *The Lean Startup*. London, UK: Portfolio Penguin.

Sandel, M. (2021). *The Tyranny of Merit*. London, UK: Penguin.

Sarigöl, E., Pfitzner, R., Scholtes, I., Garas, A., and Schweitzer, F. (2014). Predicting scientific success based on coauthorship networks. *EPJ Data Science*, 3:article number 9.

Welburn, D. (2013). Rawls, the well-ordered society and intergenerational justice. *Politics*, 33(1):56–65.

Acknowledgements

We are especially grateful to the Workshop Organisers and anonymous referees, whose insightful feedback was extremely helpful in revising this paper.