Partakable Technology

Nardine Osman
Artificial Intelligence Research Institute (IIIA-CSIC)
nardine@iiia.csic.es

Abstract

This paper proposes a shift in how technology is currently being developed by giving people, the users, control over their technology. We argue that users should have a say in the behaviour of the technologies that mediate their online interactions and control their private data. We propose 'partakable technologies', technologies where users can come together to discuss and agree on its features and functionalities. To achieve this, we base our proposal on a number of existing technologies in the fields of agreement technologies, natural language processing, normative systems, and formal verification. As an IJCAI early career spotlight paper, the paper provides an overview of the author's expertise in these different areas.

1 Motivation

There is a growing dissent against the power and control that technology is having over us. Recent concerns that have gained much attention lately are privacy and security concerns is social networks, especially with the Cambridge Analytica scandal which illustrated how our data is being used to manipulate elections worldwide.[1][2][3]

This paper is concerned with the issue of giving us, the users, control over the technologies that mediate our interactions. Today, most of our interactions are moving online, yet the systems mediating these interactions are governed by centralised monopolies that control how we interact as well as how our private data is used. We argue that as users, we need to have a say in how these technologies work. We propose 'partakable technologies' as technologies where we, the users, take part in deciding the technology's features and functionality.

[1]https://www.theguardian.com/technology/2017/oct/26/cambridge-analytica-used-data-from-facebook-and-politico-to-help-trump

[2]https://www.theguardian.com/uk-news/2018/mar/27/brexit-groups-had-common-plan-to-avoid-election-spending-laws-says-wylie

[3]https://www.theguardian.com/uk-news/2018/apr/04/cambridge-analytica-used-violent-video-to-try-to-influence-nigerian-election

To achieve our vision, we propose to build on top of a number of existing mechanisms from various lines of research in AI: from agreement technologies and natural language processing, to normative systems and formal verification. The following section (Section 2) presents our proposed roadmap for partakable technology and the different AI mechanisms that we can build upon. We then conclude with an overview of the main challenges of our proposal and some final words in Section 3.

We note that as this paper is an IJCAI early career spotlight paper, the roadmap is strongly influenced by the author's research experience and expertise. The closing section (Section 3), however, illustrates how additional lines of research may be integrated into our vision.

2 Roadmap

We base our proposal for partakable technology on norms and normative systems [Osman and Sierra, 2018]. Norms are the rules that govern behaviour in groups and societies. In multiagent systems, the study of norms gained tremendous attention due to the critical issue of coordinating agent behaviour and actions. We adopt the idea of using norms to control, or mediate, behaviour. This is motivated by the fact that: (1) software is usually engineered based on similar 'rules of behaviour' (like what action may be performed, by whom, under what conditions, etc.), and (2) the use of norms permits users to discuss their interactions without the need for any technical knowledge about the software mediating these interactions. For example, a user in an online community can state that he favours a rule that "prohibits users from sharing content with others unless they have been specifically granted the permission to do so", or suggest a new rule that "allows users to filter which posts they see, based on topic, author, etc.". However, the user may not be capable of writing the code that implements these desired rules.

As such, we argue that people should be able to discuss and agree on the norms governing their online communities in natural language; and the system should automatically adapt itself to implement the agreed upon norms. We suggest the discussion and agreement phase to build on the field of agreement technologies that supports reaching collaborative agreements, and the automatic adaptation of the system to be based on three main fields of research, namely: (1) normative systems that can enforce the agreed upon norms, (2) natural language processing that can help automatically translate the norms agreed upon from a natural language into the language of the normative system, and (3) formal verification mechanisms that can confirm that the final system does indeed implement the norms agreed upon. In what follows, we discuss these different fields of research, while giving special attention to the author's contribution to these fields.

2.1 Agreement Technologies

Our main objective is to give the novice person the means to be able to discuss and agree, along with others, on the rules (or norms) that should govern their online interactions. To achieve our goal, we suggest to make use of agreement technologies. Agreement technologies [Pacheco *et al.*, 2017] have emerged as an imperative field in multiagent systems with the aim of helping individuals collaboratively reach a decision. The field is based on a number of models and mechanisms, such as argumentation and negotiation mechanisms, computational social choice, and trust and reputation models.

Argument schemes and critical questions (SchCQ) [Walton, 1996] can be used to enable the discussion about which norms should be adopted. Argumentation mechanisms [Bench-Capon and Dunne, 2007; Leite and Martins, 2011; Modgil and Prakken, 2013] can then be used to help assess the strength of arguments, whereas negotiation mechanisms can help people arrive to mutual agreements [Jonge and Sierra, 2017; Baarslag *et al.*, 2016].

However, in any setting where agreements need to be reached collaboratively, trust between users (or people's reputation) plays a crucial role in helping one assess the reliability of others' opinions and arguments. Trust and reputation has always been an important issue in open systems, and is crucial in online discussions where participants may not know each other personally or have interacted much in the past. In what follows, we briefly present some of the existing work on trust or reputation.

The trust model developed by Osman *et al.* [2013] helps find suitable agents to collaborate with in a given interaction in distributed open systems. It calculates the expectation about an agent's future performance in a given context by assessing both the agent's willingness and capability through the semantic comparison of the current context in question with the agent's performance in past similar experiences.

In [Osman *et al.*, 2014a], a trust algorithm has been developed that helps assess the reliability of advice and advisers. The trustworthiness of advice is calculated based on the expectation of an advice's outcome, again by learning from an analysis of similar past experiences using tools such as semantic matching and action empowerment. A trustworthy advisor is then an adviser who gives good advice.

In [Gutierrez *et al.*, 2015], a trust-based model is developed to support collaborative assessments in online classrooms. The model essentially aggregates peer assessments using a weighted average, where the weight of each assessment is the trustworthiness of the peer with respect to the teacher. This trustworthiness is based on the similarity of the peer and tutor's marking profiles. In other words, the more similar a peer's assessments are to the tutor's, the more trustworthy will his future assessments be. In [Gutierrez *et al.*, 2016], a similar model is developed, however, instead of representing trust and assessments as ordinal numbers (as in the model in [Gutierrez *et al.*, 2015]), they are represented as probability distributions.

In [Osman and Sierra, 2016], a reputation model is developed to assess the reputation of researchers and their research work. The model relies on peer opinions (researchers reviewing each others work), and it introduces the idea of reviewers of the same paper assessing each others reviews (which is referred to as judgements). As such, the reputation of authors becomes dependent on the reputation of their papers. The reputation of papers depends on the reviews they receive, where the weight of each review depends on the reputation of the reviewer. The reputation of reviewers depends on the reputation of their reviews, which in turn depends on the judgements that the reviews receive.

In [Osman *et al.*, 2010], a model is developed which allows one to infer her opinion about unfamiliar entities (or nodes) in a structural graph based on her view of related entities. The proposed mechanism focuses on the "part of" relation to illustrate how reputation may flow (or propagate) from one entity to another. For example, one's about a given brand can help her form an initial opinion about one of the brands latest products that she has never tried before.

In [Osman *et al.*, 2014b], opinions are extracted from behavioural information, such as the results of football games, and they are uses to assess reputation and predict behaviour accordingly (such as assessing the reputation of football teams and predicting the results of future matches).

Going back to our proposal of partakable technology, we suggest future work to focus on using a combination of agreement technologies to help support people reach an agreement on their technologies' norms. For example, social voting may be incorporated with argumentation [Leite and Martins, 2011], trust may be combined with argumentation [Bonatti *et al.*, 2014], and voting algorithms may be adapted to incorporate trust measures (the weights of votes) [Endriss, 2014; Brandt *et al.*, 2016]. Some existing applications that combine agreement technologies for reaching collaborative agreements are WeCurate [Confalonieri *et al.*, 2015; Amgoud *et al.*, 2012b] and WeShare [Amgoud *et al.*, 2012a]. They both use a basic negotiation-based approach, enhanced with bilateral arguments and voting mechanisms, for collaborative decision making. They are socio-technical systems that support co-browsing across multiple devices and enable groups of users to collaboratively curate artistic work or buy a gift together.

2.2 Natural Language Processing

As the norms that people agree on are expected to be in natural language (or a controlled natural language), we must be able to translate these norms into an executable language, or the software that will mediate interactions and behaviour following these norms. As such, we must build an automatic translator for translating the norms (or rules of interaction) from natural language into an executable language. This is a rather challenging task, though important research has been carried out in that direction. [Wyner and Governatori, 2013] illustrates and discusses the use of existing state-of-the-art techniques in the automatic translation of regulatory rules in natural language into a machine readable formal representation. [Shiffman *et al.*, 2010] translates a complete set of paediatric guideline recommendations into a controlled language (Attempto Controlled English, ACE). [Wyner *et al.*, 2010] adopts and applies a controlled natural language to constrain the domain of discourse in an on-line discussion forum for e-government policy-making. The controlled natural language helps eliminate ambiguity, and allows a logical representation of statements. Each of the policy statements is then automatically translated into first-order logic. [Wyner and Peters, 2011] presents a linguistically-oriented, rule-based approach, for extracting conditional and deontic rules from regulations specified in natural language. Finally, [Athan *et al.*, 2013] presents approaches for the logical representation of regulations.

One approach that the author is currently investigating is to base the translation on exploring the connection between natural language features and the formal ones. For example, we are trying to label a statement whether it is an obligation, a permission, or a prohibition based on analysing its verb (such as checking whether the verb contains 'must', 'shall', 'ought to', 'may', etc.). Or trying to figure out who a norm addresses by looking for the subject in the sentence. Or trying to figure out the conditions by searching for conditional conjunctions, such as 'if', 'when', etc.

2.3 Normative Systems

The literature provides a variety of solutions that deal with specifying and regulating interactions in multiagent systems based on the concept of following social norms [Shoham and Tennenholtz, 1995], such as having contracts and commitments [Dignum *et al.*, 2002], organisational approaches [Horling and Lesser, 2004], electronic institutions [d'Inverno *et al.*, 2012], distributed dialogues [Robertson, 2004], and so on.

Two specifically interesting approaches are electronic institutions [d'Inverno *et al.*, 2012; Osman, 2018] and the lightweight coordination calculus [Robertson, 2004]. In [d'Inverno *et al.*, 2012], it is argued that open multi-agent systems can be effectively designed and implemented as agent mediated electronic institutions where heterogeneous (human and software) agents can participate, playing different roles and interacting by means of speech acts. An institution is defined by a set of roles that agents participating in the institution will play, a common language to allow heterogenous agents to exchange knowledge, the valid interactions that agents may have structured in conversations, and the consequences of agents' actions within an institution, captured by obligations that agents acquire and fulfil. Electronic institutions have gained a lot of attention in the multiagent system field, and they have been applied to various domains, from online learning, to social networks [Osman, 2018].

The lightweight coordination calculus (LCC) [Robertson, 2004] is a process calculus, based on logic programming, that provides means of achieving coordination in distributed systems by enforcing social norms. The process calculus specifies what actions agents can perform, when they can perform such actions, under what conditions these actions may be carried out, and so on. However, unlike electronic institutions, there are no 'governors' that ensure that agents abide by norms. Of course, like all the approaches above, these rules are associated with roles rather than physical agents; and agents can play more than one role in more than one interaction. This provides an abstraction for the interaction model from the individual agents that might engage in such an interaction. But what is particularly interesting about LCC from our proposal's perspective is that LCC is an executable process calculus. In other words, the same executable code can be fed into a model checker for verification without the need to model the interaction, as is traditionally the case with model checking. The following section elaborates further on this topic and its practical implications.

2.4 Formal Verification

As we are automatically translating people's requirements from natural language into an executable language (or the software that mediates interactions), it is crucial to verify that the overall behaviour of the software satisfies the requirements put forward by the users. To develop the verification mechanism, we suggest to build on previous work in model checking multiagent systems [Wooldridge *et al.*, 2002; Lomuscio *et al.*, 2009; Bordini *et al.*, 2003].

A particularly interesting model checker, however, is that of [Osman *et al.*, 2006a; 2006b]. First, it verifies a combination of interaction models with specific agent requirements (those that the agent wishes to make public, such as the seller stating that it can only accept payments through PayPal). The argument for doing this is that it is usually much more interesting to verify dynamic properties of multiagent interactions that take into consideration both the interaction model and the agents involved in that interaction, as opposed to static temporal properties of the interaction only (such as deadlocks). However, agents do not usually make their specification public (to be accessed by a model checker in open systems), and as such, only information that the agents wish to be made public can be used (such as the requirement that only PayPal payments can be accepted). In summary, this allows for the specification of a much more interesting set of properties that can be checked, which depends on the interaction/agents combination [Osman and Robertson, 2007].

Second, and more important for our proposed roadmap, the model checker introduces interaction time verification for multiagent systems. This presents the agents with the opportunity of performing the verification themselves when the conditions for verification are met. This is highly useful for allowing agents to automatically decide which interaction model and which group of agents is suitable to join. This is also usually very difficult to obtain with other model checkers, especially when they follow a global model checking technique that requires the entire statespace to be generated before verification can happen. In fact, efficiency is a common problem of model checking, as it is common for model checkers to hit the state space explosion problem. However, as the model checker of [Osman *et al.*, 2006a; 2006b] follows a Prolog-based approach, the problem of searching the statespace is automatically solved with Prolog's backtracking mechanism. This essentially implements of a local model checking approach that checks for satisfaction without the need for a prebuilt statespace.

This lightweight model checker (that accepts as input an executable language, the LCC language of Section 2.3), implemented in less than 200 lines of Prolog code, is what makes interaction time verification possible. As our proposal expects people to agree on norms, and the system to automatically adapt to those norms, interaction time verification becomes crucial. One approach would be to adapt the input languages that the model checker of [Osman *et al.*, 2006a; 2006b] currently accepts into the language of the normative system that will be implemented (or chosen).

3 Challenges and Final Words

This paper presented a roadmap for partakable technology. The main objective is giving people control over the technologies that mediate their everyday interactions, such as social networks, collaboration tools, online learning systems, to name a few.

The proposal is built on top of a number of well established lines of research, mostly relying on agreement technologies, natural language processing, normative systems, and formal verification. This ensures the feasibility of our proposal. Nevertheless, we foresee the main challenge of our proposal to be having all the previous languages and mechanisms *working together*. As illustrated earlier, there already exist many languages and mechanisms for each of the research lines presented, though these mechanisms have been designed under strict assumptions that limit their impact on or usefulness in real life scenarios. The challenge will be to bring in all these languages and mechanisms to work together to realise our vision. We take inspiration here from the SIMPLE language [de Jonge and Sierra, 2015], a controlled natural language that is also an executable language, which illustrates that one simple language may be possible to cover the language used by the people (in this case a *controlled* natural language) as well as be used as the executable language of the software.

As illustrated in the introduction, the proposed roadmap is strongly influenced by the author's research history. Nevertheless, a number of additional research lines can help contribute to this proposal. In machine learning, pattern recognition techniques can help figure out when things are wrong and the norms need to be revised (for example, when collaboration decreases). Norm synthesis techniques can help learn and suggest to users the best norms for their community. Learning the consequences of norms can help contribute to the argumentation process of Section 2.1. Alternative techniques can also be useful here. For example, sentiment analysis can help pick up the dis/satisfaction of the community, and suggest whether the norms needs to be revisited accordingly. Simulations or analogical reasoning can help understand the consequences of a given norm. As such, we acknowledge that the lines of research presented in this paper form only part of the technologies that can help realise the proposed vision of partakable technology.

Last, but not least, we also acknowledge that the suggested work cannot succeed with a purely technological approach. As such, this work must be carried out in a close-knit collaboration with a multidisciplinary team, including experts in philosophy and legal studies. For example, the input from both philosophy and law can help us better understand the ethical and legal implications of allowing people to decide their technologies' features and functionalities, and how to avoid potential abuse.

Acknowledgments

This work is supported by the CIMBVAL Spanish project (funded by MINECO, project # TIN2017-89758-R) and the AppPhil Catalan project (funded by RecerCaixa 2017).

References

[Amgoud *et al.*, 2012a] Leila Amgoud, Roberto Confalonieri, Dave De Jonge, Mark d'Inverno, Katina Hazelden, Nardine Osman, Henri Prade, Carles Sierra, and Matthew Yee-King. Sharing online cultural experiences: An argument-based approach. In *Proc. of MDAI'12*, pages 282–293. Springer, 2012.

[Amgoud *et al.*, 2012b] Leila Amgoud, Roberto Confalonieri, Dave De Jonge, Mark d'Inverno, Katina Hazelden, Nardine Osman, Henri Prade, Carles Sierra, and Matthew Yee-King. Wecurate: Designing for synchronised browsing and social negotiation. In *Proc. of AT'12*, pages 168–179. CEUR-WS.org, 2012.

[Athan *et al.*, 2013] Tara Athan, Harold Boley, Guido Governatori, Monica Palmirani, Adrian Paschke, and Adam Wyner. OASIS LegalRuleML. In *Proc. of ICAIL'13*, pages 3–12. ACM, 2013.

[Baarslag *et al.*, 2016] Tim Baarslag, Mark J. C. Hendrikx, Koen V. Hindriks, and Catholijn M. Jonker. Learning about the opponent in automated bilateral negotiation: a comprehensive survey of opponent modeling techniques. *Autonomous Agents and Multi-Agent Systems*, 30(5):849–898, 2016.

[Bench-Capon and Dunne, 2007] Trevor Bench-Capon and Paul Dunne. Argumentation in artificial intelligence. *Artif. Intell.*, 171:619–641, 2007.

[Bonatti *et al.*, 2014] Piero Bonatti, Eugenio Oliveira, Jordi Sabater-Mir, Carles Sierra, and Francesca Toni. On the integration of trust with negotiation, argumentation and semantics. *Knowl. Eng. Rev.*, 29:31–50, 2014.

[Bordini *et al.*, 2003] Rafael H Bordini, Michael Fisher, Carmen Pardavila, Willem Visser, and Michael Wooldridge. Model checking multi-agent programs with casp. In *Proc. of CAV '03*, pages 110–113. Springer, 2003.

[Brandt *et al.*, 2016] Felix Brandt, Vincent Conitzer, Ulle Endriss, Jérôme Lang, and Ariel D. Procaccia, editors. *Handbook of Computational Social Choice*. Cambridge University Press, 2016.

[Confalonieri *et al.*, 2015] Roberto Confalonieri, Matthew Yee-King, Katina Hazelden, Mark d'Inverno, Dave De Jonge, Nardine Osman, Carles Sierra, Leila Amgoud, and Henri Prade. Engineering multiuser museum interactives for shared cultural experiences. *Eng. Appl. Artif. Intell.*, 46:180–195, 2015.

[de Jonge and Sierra, 2015] Dave de Jonge and Carles Sierra. SIMPLE: a language for the specification of protocols, similar to natural language. In *Proc. of COIN '15*, pages 49–64, 2015.

[Dignum *et al.*, 2002] Virginia Dignum, John-Jules Meyer, and Hans Weigand. Towards an organizational model for agent societies using contracts. In *Proc. of AAMAS '02*, pages 694–695. ACM, 2002.

[d'Inverno *et al.*, 2012] Mark d'Inverno, Michael Luck, Pablo Noriega, Juan A. Rodriguez-Aguilar, and Carles Sierra. Communicating open systems. *Artif. Intell.*, 186:38–94, 2012.

[Endriss, 2014] Ulle Endriss. Social choice theory as a foundation for multiagent systems. In *Proc. of MATES'14*, pages 1–6. Springer, 2014.

[Gutierrez *et al.*, 2015] Patricia Gutierrez, Nardine Osman, Carme Roig, and Carles Sierra. Trust-based community assessment. *Pattern Recognit. Lett.*, 67:49–58, 2015.

[Gutierrez *et al.*, 2016] Patricia Gutierrez, Nardine Osman, Carme Roig, and Carles Sierra. Personalised automated assessments. In *Proc. of AAMAS'16*, pages 1115–1123. ACM, 2016.

[Horling and Lesser, 2004] Bryan Horling and Victor Lesser. A survey of multi-agent organizational paradigms. *Knowl. Eng. Rev.*, 19:281–316, 2004.

[Jonge and Sierra, 2017] Dave De Jonge and Carles Sierra. D-brane: a diplomacy playing agent for automated negotiations research. *Appl. Intell.*, 47(1):158–177, 2017.

[Leite and Martins, 2011] João Leite and João Martins. Social abstract argumentation. In *Proc. of IJCAI'11*, pages 2287–2292. AAAI Press, 2011.

[Lomuscio *et al.*, 2009] Alessio Lomuscio, Hongyang Qu, and Franco Raimondi. MCMAS: A model checker for the verification of multi-agent systems. In *Proc. of CAV '09*, pages 682–688. Springer, 2009.

[Modgil and Prakken, 2013] Sanjay Modgil and Henry Prakken. A general account of argumentation and preferences. *Artif. Intell.*, 195:361–397, 2013.

[Osman and Robertson, 2007] Nardine Osman and David Robertson. Dynamic verification of trust in distributed open systems. In *Proc. of IJCAI'07*, pages 1440–1445, 2007.

[Osman and Sierra, 2016] Nardine Osman and Carles Sierra. Reputation in the academic world. In *Proc. of Trust Workshop @ AAMAS 2016)*, pages 1–17. CEUR-WS.org, 2016.

[Osman and Sierra, 2018] Nardine Osman and Carles Sierra. Agreement computing and self-evolution. In *Anchoring Institutions. Democracy and Regulations in a Global and Half-automated World*. Springer Verlag, 2018.

[Osman *et al.*, 2006a] Nardine Osman, David Robertson, and Christopher Walton. Dynamic model checking for multi-agent systems. In *DALT IV*, pages 43–60. Springer, 2006.

[Osman *et al.*, 2006b] Nardine Osman, David Robertson, and Christopher Walton. Run-time model checking of interaction and deontic models for multi-agent systems. In *Proc. of AAMAS '06*, pages 238–240. ACM, 2006.

[Osman *et al.*, 2010] Nardine Osman, Carles Sierra, and Jordi Sabater-Mir. Propagation of opinions in structural graphs. In *Proc. of ECAI'10*, pages 595–600. IOS Press, 2010.

[Osman *et al.*, 2013] Nardine Osman, Carles Sierra, Fiona McNeill, Juan Pane, and John K. Debenham. Trust and matching algorithms for selecting suitable agents. *ACM Trans. Intell. Syst. Technol.*, 5(1):16:1–16:39, 2013.

[Osman *et al.*, 2014a] Nardine Osman, Patricia Gutierrez, and Carles Sierra. Trustworthy advice. In *Proc. of ECAI'14*, pages 1073–1074. IOS Press, 2014.

[Osman *et al.*, 2014b] Nardine Osman, Alessandro Provetti, Valerio Riggi, and Carles Sierra. MORE: merged opinions reputation model. In *Proc. of EUMAS'14*, pages 67–81. Springer, 2014.

[Osman, 2018] Nardine Osman, editor. *Electronic Institutions and their Applications*. Springer, 2018.

[Pacheco *et al.*, 2017] Natalia C. Pacheco, Carlos Carrascosa, Nardine Osman, and Vicente J. Inglada. *Multi-Agent Systems and Agreement Technologies: 14th European Conference, EUMAS 2016, and 4th International Conference, AT 2016*. Springer, 2017.

[Robertson, 2004] David Robertson. A lightweight method for coordination of agent oriented web services. In *Proc. of AAAI Spring Symposium on Semantic Web Services*, 2004.

[Shiffman *et al.*, 2010] Richard N. Shiffman, George Michel, Michael Krauthammer, Norbert E. Fuchs, Kaarel Kaljurand, and Tobias Kuhn. Writing clinical practice guidelines in controlled natural language. In *Proc. of CNL'09*, pages 265–280. Springer-Verlag, 2010.

[Shoham and Tennenholtz, 1995] Yoav Shoham and Moshe Tennenholtz. On social laws for artificial agent societies: off-line design. *Artif. Intell.*, 73:231–252, 1995.

[Walton, 1996] Doug Walton. *Argument Schemes for Presumptive Reasoning*. Lawrence Erlbaum Assoc., 1996.

[Wooldridge *et al.*, 2002] Michael Wooldridge, Michael Fisher, Marc-Philippe Huget, and Simon Parsons. Model checking multi-agent systems with MABLE. In *Proc. of AAMAS'02*, pages 952–959. ACM, 2002.

[Wyner and Governatori, 2013] Adam Wyner and Guido Governatori. A study on translating regulatory rules from natural language to defeasible logic. In *RuleML 2013*, pages 16.1–16.8, 2013.

[Wyner and Peters, 2011] Adam Wyner and Wim Peters. On rule extraction from regulations. In *Proc. of JURIX'11*, pages 113–122. IOS Press, 2011.

[Wyner *et al.*, 2010] Adam Wyner, Tom van Engers, and Kiavash Bahreini. From policy-making statements to first-order logic. In *Proc. of EGOVIS'10*. Springer, 2010.