Computer-aided Reasoning about Collective Decision Making

Doctoral Consortium

Arthur Boixel
Institute for Logic, Language and Computation
Amsterdam, The Netherlands
a.boixel@uva.nl

ABSTRACT

Many contributions to the field of computational social choice employ a variety of techniques developed in computer science to study the properties of preference aggregation mechanisms. However, there has been little research on how the availability of such techniques can help us *reason* about collective decision making. In this positional paper I discuss how computer-aided methods can be used to—automatically—reason about the outcomes of a collective decision making process. I also lay down several research directions one could explore to further develop the field in this direction.

KEYWORDS

COMSOC; Automated Reasoning; Combinatorial Optimisation

ACM Reference Format:
Arthur Boixel. 2020. Computer-aided Reasoning about Collective Decision Making. In *Proc. of the 19th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2020), Auckland, New Zealand, May 9–13, 2020,* IFAAMAS, 3 pages.

1 INTRODUCTION

Computational social choice exploits the idea that any mechanism we may use to take collective decisions is ultimately an algorithm. Thus, many techniques developed in computer science greatly contributed to the field. Computational complexity is a beautiful example as it allows for an informed discussion about such mechanisms: for example by studying the hardness of computing an outcome [8, 11] or by acting as a barrier against manipulation [6]. Logic and knowledge representation have been used to study the compact representation of preferences [7] and to develop voting in combinatorial domains [13]. Many more examples are reviewed in the Handbook of Computational Social Choice [2].

In their seminal paper, Tang and Lin [16] showed how to obtain new proofs of several famous impossibility theorems using computer-aided methods. Specifically, they make use of SAT solvers to automatically prove the base cases of such theorems. Since then, this approach has been successfully applied in other contexts. Geist and Endriss [9] for example, automated the search for new impossibility theorems in the context of ranking sets of objects. In a related vein, Brandt et al. [3] found optimal bounds for the no-show paradox. Using bounded model checking, Kirsten and Cailloux [12] developed a method to automatically find a counterexample showing that a given rule fails to satisfy a given axiom. Although there seems to be a growing interest in the use of computer-assisted reasoning techniques [10], many of their potential applications remain largely unexplored; applications that I wish to examine during the remainder of my thesis.

Proc. of the 19th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2020), B. An, N. Yorke-Smith, A. El Fallah Seghrouchni, G. Sukthankar (eds.), May 9–13, 2020, Auckland, New Zealand.

2 JUSTIFYING COLLECTIVE DECISIONS

As a first step in this direction, I will now present how the availability of computer-aided methods can help us to reason about the possible outcomes of a collective making decision process. This section is based on the notion of *justification* developed in a recent paper [1] co-written with Endriss in which we extend the work of Cailloux and Endriss [5].

Example 1. The employees of a company need to elect their new president. The statutes of the company prescribe the use of a specific voting rule for this purpose. However, no one remembers why this rule was selected nor how it works. The employees, while being perfectly able to verify the correctness of the outcome, would like to do more. They hope to find a *justification* of the outcome in terms of normative principles—on which everyone can agree—that happen to be able to explain why it is the “best” possible compromise. △

Example 2. Imagine a group of collaborators who wants to reach a consensus on which business strategy to adopt for their company. They engage into a deliberative protocol, discussing the ins and outs of the possible strategies. Their individual preferences are likely to change during this process and they would like to, every now and then, be able to *justify* why a specific strategy would be the “best” possible one given their current preferences. △

Example 3. The members of a research group want to settle on a way to take decisions in the years ahead. After their colleague—a social choice theorist, worshiping the myriad of impossibility results marring the field—had taken away all hope of finding the perfect voting rule, they eventually agree on a method of decision making: “from first principles”. Whenever a decision needs to be taken, the “best” outcome will be the one they are able to *justify* using their preferred corpus of normative principles. △

Thus, one may then ask: How to *justify* that a target outcome is the “best” possible compromise in a given situation? Such justifications may be used to defend an already taken decision (Example 1), to support a group intent on arriving at a good decision through deliberation (Example 2), or to serve as a stand-in when no specific voting rule seems acceptable (Example 3).

In a classical voting scenario, a voting rule is selected and determines the outcome. The normative principles (called *axioms*) characterising the rule may be seen as justifying the outcome. But selecting a rule is a complex task that requires expertise and the plethora of rules available makes it hard to settle on a specific one.

No rule is perfect and will only satisfy some of the axioms we might reasonably care about. But maybe we can justify outcomes independently of how the decision will be or has been taken.

In order for a justification to be understood and accepted by a group of agents, it must use arguments referring to societal norms everyone can agree on. It should also provide the agents with a step-by-step explanation—referring to concrete *instances* of these principles—for why selecting the target outcome is the only way to satisfy these normative principles. Thus, the notion of justification we seek to develop has both an explanatory and a normative component. Finally, one should be able to automatically search for such justifications in practice. But finding an explanation grounded in a suitable normative basis is a computationally demanding task. To make it practically feasible, we exploit an analogy between the notions of (*i*) explanation for an observed inconsistency and (*ii*) the notion of *minimal unsatisfiable subset* familiar from the field of automated reasoning [4]. This allows us to build on the enormous progress in solving intricate combinatorial optimisation problems made by the AI and OR research communities in recent years [14].

The notion of justification we develop is general and provides us with theoretical guarantees. For example, it might be the case that different justifications exist for the same outcome: this is an important feature for application scenarios such as the one described in Example 1. In a given situation, it is possible that different outcomes can be justified using different normative principles; a possibility that seems to be necessary to develop the kind of application sketched in Example 2. On the other hand, justifying two contradictory outcomes using the same normative basis is impossible; an impossibility enabling the kind of application mentioned in Example 3. From a practical point of view, we showed how to operationalise our approach by (*i*) using a simple algorithm and (*ii*) encoding normative principles in the widely used constraint programming language MiniZinc [15]. This proof-of-concept implementation allowed us to search for justifications in small voting scenarios and gave us insights regarding the explanatory power of several well-known axioms from social choice theory.

3 FURTHER CHALLENGES

In this final section I will present the most interesting research directions I wish to explore for the remainder of my thesis. I will start by highlighting the challenges one will have to tackle in order to further develop the notion of a justification of a collective decision. Finally, I will outline a research agenda for exploring how computer-aided reasoning can contribute to computational social choice more generally.

Justifications. The work I presented in the previous section paves the way towards facilitating an informed discussion about the possible outcomes of a collective decision making process. However, while being a promising research direction, it also raises a lot of challenges of both a theoretical and a practical nature.

First, the notion of justification we developed is independent from the logical language used to encode the normative principles considered. This logical language, in turn, determines what an instance of an axiom is and how hard it is to extract one. While we have successfully managed to operationalise our approach by using a simple language, others might be better suited for our purpose. Thus, understanding where the hardness of this problem lies through a rigorous study of its the computational complexity should (*i*) give us more insights on which language would be the most appropriate and (*ii*) help us design efficient algorithms to improve the scalability of our approach. Variants of this problem, such as restricting our attention to specific types of justifications or principles, might also be easier to solve in practice.

Second comes the question of what makes for a “good” justification. There are several intuitive directions one could explore in search of an answer: its length, its shape, the type of principles it uses, etc. Here, running real-life experiments might be helpful: one could for example let people rate the convincingness of a justification. Hopefully, such an experiment would also give us insights into how best to present a justification to a non-expert audience.

Finally, it would be interesting to investigate possible applications of this notion, such as those described in Section 2 or new ones outside voting theory.

Computer-aided reasoning. Most recent developments towards using computer-aided methods to reason about collective decision making heavily rely on well-known techniques, such as SAT solving. While these approaches have proven to be useful, others such as using SMT solvers or automated theorem provers—that allow to reason within specific theories—remain largely unexplored. Mainly because the application of these techniques do require some expertise both in social choice theory—to come up with interesting questions—and in the use of these tools—to search for answers in practice. Thus, I reckon that a close collaboration between both communities, alongside with the development of specific tools helping social choice theorists to easily investigate any idea, will lead to fruitful contributions to the field. This is the direction I intend to follow for the remainder of my thesis.

REFERENCES

[1] Arthur Boixel and Ulle Endriss. 2020. Automated Justification of Collective Decisions via Constraint Solving. In *Proceedings of the 19th International Conference on Autonomous Agents and Multiagent Systems (AAMAS-2020)*. IFAAMAS.

[2] Felix Brandt, Vincent Conitzer, Ulle Endriss, Jérôme Lang, and Ariel D. Procaccia (Eds.). 2016. *Handbook of Computational Social Choice.* Cambridge University Press.

[3] Felix Brandt, Christian Geist, and Dominik Peters. 2017. Optimal bounds for the no-show paradox via SAT solving. *Mathematical Social Sciences* 90 (2017), 18–27.

[4] Hans Kleine Büning and Oliver Kullmann. 2009. Minimal Unsatisfiability and Autarkies. *Handbook of Satisfiability* 185 (2009), 339–401.

[5] Olivier Cailloux and Ulle Endriss. 2016. Arguing about Voting Rules. In *Proceedings of the 15th International Conference on Autonomous Agents and Multiagent Systems (AAMAS-2016)*. IFAAMAS. Also presented at COMSOC-2016.

[6] Vincent Conitzer and Toby Walsh. 2016. Barriers to Manipulationin Voting. In *Handbook of Computational Social Choice*, F. Brandt, V. Conitzer, U. Endriss, J. Lang, and A. D. Procaccia (Eds.). Cambridge University Press, Chapter 6.

[7] Sylvie Coste-Marquis, Jérôme Lang, Paolo Liberatore, and Pierre Marquis. 2004. Expressive Power and Succinctness of Propositional Languages for Preference Representation.. In *Proceedings of the 9th International Conference on Principles of Knowledge Representation and Reasoning (KR'04)*, Vol. 4. 203–212.

[8] Andrew Davenport and Jayant Kalagnanam. 2004. A computational study of the Kemeny rule for preference aggregation. In *Proceedings of the 19th National Conference on Artifical Intelligence (AAAI'04)*, Vol. 4. 697–702.

[9] Christian Geist and Ulle Endriss. 2011. Automated search for impossibility theorems in social choice theory: Ranking sets of objects. *Journal of Artificial Intelligence Research* 40 (2011), 143–174.

[10] Christian Geist and Dominik Peters. 2017. Computer-Aided Methods for Social Choice Theory. In *Trends in Computational Social Choice*, Ulle Endriss (Ed.). AI Access.

[11] Edith Hemaspaandra, Holger Spakowski, and Jörg Vogel. 2005. The complexity of Kemeny elections. *Theoretical Computer Science* 349, 3 (2005), 382–391.

[12] Michael Kirsten and Olivier Cailloux. 2018. Towards Automatic Argumentation about Voting Rules. In *Actes de la 4ème Conférence Nationale sur les Applications Pratiques de l'Intelligence Artificielle (APIA-2018)*.

[13] Jérôme Lang and Lirong Xia. 2016. Voting in Combinatorial Domains. In *Handbook of Computational Social Choice*, F. Brandt, V. Conitzer, U. Endriss, J. Lang, and A. D. Procaccia (Eds.). Cambridge University Press, Chapter 9.

[14] Michela Milano. 2018. Twenty Years of Constraint Programming (CP) Research. *Constraints* 23, 2 (2018), 155–157.

[15] Nicholas Nethercote, Peter J. Stuckey, Ralph Becket, Sebastian Brand, Gregory J. Duck, and Guido Tack. 2007. MiniZinc: Towards a Standard CP Modelling Language. In *Proceedings of the 13th International Conference on Principles and Practice of Constraint Programming (CP-2007)*. Springer.

[16] Pingzhong Tang and Fangzhen Lin. 2009. Computer-aided proofs of Arrow's and other impossibility theorems. *Artificial Intelligence* 173, 11 (2009), 1041–1053.