ICRES 2019: International Conference on Robot
Ethics and Standards, London, UK, 29-30 July 2019.
https://doi.org/10.13180/icres.2019.29-30.07.005

ARGUMENT AGGREGATION IN A DEONTIC LOGIC

N.S. GOVINDARAJULU and S. BRINGSJORD*

Rensselaer AI & Reasoning (RAIR) Lab
Department of Cognitive Science; Department of Computer Science
Rensselaer Polytechnic Institute (RPI)
Troy NY 12180 USA
E-mail: naveensundarg@gmail.com and selmer.bringsjord@gmail.com*
www.rpi.edu

We will present initial results in handling **argument adjudication** in a deontic logic. There are many impossibility results in aggregation, such as Arrow's theorem in social choice theory and similar results for judgement aggregation. There has however been no such analysis for argument adjudication over nested modal statements, which are necessary for modelling ethical principles and theories. We lay out the background and present brief initial results. The problem that we aim to solve is given below:

Problem

We have a set of n agents producing possibly varying statements that can be ethically charged $\{\phi_1, \ldots, \phi_n\}$. The statements are in formal system containing a deontic operator $\mathbf{O}(\gamma, \rho)$. Each of the agents also produce arguments in support of their statements $\{\alpha_1 \rightsquigarrow \phi_1, \ldots, \alpha_1 \rightsquigarrow \phi_n\}$. We need to compute a statement ϕ^* and an argument α^* that best represents the diverse statements and arguments given a background ethical theory Γ.

In judgment aggregation, we are tasked with aggregating a group's judgment on a set of propositional options. When we move to the case where the judgment is a probability value, there are more usable aggregations available.[1] However, note that the results apply only when the judgement is a probability value. Our approach is best labeled *argument adjudication*. Argument adjudication is not to be confused with argument *aggregation*,[2] which is based on the standard approach of treating arguments as abstract objects having none of the nuanced, internal structure analogous to what formal proofs have[3]this standard, abstract conception of

arguments is given in. Since our formal theory of argumentation is fundamentally different than what is seen in argument aggregation, the axioms that constrain argument adjudication are different than those operative in argument aggregation. Along the same line, while we happen to often represent proofs and arguments in graphs, prior work on so-called *graph aggregation*[4] is completely separate from our "magic," since in this other work graphs are treated abstractly and their internal inferential structure is ignored. Note that aggregation theorems for uncertainties using strength factors for modal statements, and other expressions, do not exist and have to be derived.

We will present an algorithm and adjudication and analyze its properties. For the formal system, we use a variant the **deontic cognitive event calculus** ($\mathcal{DCEC}$), a computational formal logic, a logic that has been used to model various ethical principles and processes. For example, this logic has been used previously in [5,6] to automate versions of the Doctrine of Double Effect ($\mathcal{DDE}$), *akrasia*,[7] and model virtue ethics.[8]

References

1. C. List, *Synthese* **187**, 179 (2011).
2. P. Dunne, P. Marquis and M. Wooldridge, Argument Aggregation: Basic Axioms and Complexity Results, in *Computational Models of Argument*, eds. B. Verheij, S. Szeider and S. Woltran (IOS Press, Amsterdam, The Netherlands, 2012) pp. 129–140. This book is Vol 245 in the series *Frontiers in Artificial Intelligence and Applications*.
3. P. Dung, *Artificial Intelligence* **77**, 321 (1995).
4. U. Endriss and U. Grandi (2016), This arXiv paper is a synthesis and expansion of prior papers by E&G from 2012 and 2014, detailed reference info on which is provided by authors in their first footnote.
5. N. S. Govindarajulu and S. Bringsjord, On Automating the Doctrine of Double Effect, in *Proceedings of the Twenty-Sixth International Joint Conference on Artificial Intelligence, IJCAI-17*, ed. C. Sierra (Melbourne, Australia, 2017). Preprint available at this url: `https://arxiv.org/abs/1703.08922`.
6. N. S. Govindarajulu, S. Bringsjord, R. Ghosh and M. Peveler, Beyond The Doctrine Of Double Effect: A Formal Model of True Self-Sacrifice, International Conference on Robot Ethics and Safety Standards, (2017).
7. S. Bringsjord, N. S. Govindarajulu, D. Thero and M. Si, Akratic Robots and the Computational Logic Thereof, in *Proceedings of ETHICS • 2014 (2014 IEEE Symposium on Ethics in Engineering, Science, and Technology)*, (Chicago, IL, 2014). IEEE Catalog Number: CFP14ETI-POD.
8. N. S. Govindarajulu, S. Bringsjord, R. Ghosh and V. Sarathy, Towards the engineering of virtuous machines, in *Proceedings of the 2nd AAAI/ACM Workshop on Artificial Intelligence, Ethics, and Society (AIES-19)*, 2019.