Finding Our Moral Values: Guidelines for Value System Aggregation

Víctor Abia Alonso[1], Marc Serramia[2], and Eduardo Alonso[1]

[1] City St George's, University of London, United Kingdom
[2] IIIA - CSIC, Barcelona, Spain

Abstract. Ensuring AI aligns with our moral values is an important problem. To achieve AI alignment though, we first need to know what values we want AI to align with. However, obtaining a model of society's values is far from a simple problem. This problem, called value inference, typically needs to perform *value system aggregation* which consists in combining moral value models of several individuals to obtain one representing everybody. So far, only one such method has been proposed. This paper discusses why research in value system aggregation is necessary and outlines avenues to implement it depending on the value alignment problem at hand. In particular, we discuss how value system aggregation can be useful for policymaking applications.

Keywords: Value alignment · Aggregation · Participatory Budgets

1 Introduction

Ensuring AI aligns with our moral values is one of the key challenges for AI research in the coming years. However, obtaining the value system (i.e. the moral values and value preferences) we want AI to align with is an important preliminary step often overlooked. The process of obtaining a societal value system is not only useful for AI alignment but for other areas like governance and policymaking. Institutions like the European Commission highlight the importance of moral values to better understand citizens, co-create, and communicate policy [28]. Thus, in the case of policymaking, a model of citizen values would be useful both to inform governments and to use it in AI tools for policymaking.

Researchers have only very recently started to look into value inference, with Liscio et al. [17] presenting an approach based on three steps: value identification, value system estimation, and value system aggregation. Value identification is concerned with identifying the moral values that are relevant for a context (e.g. if we want to discuss traffic planning the value of security is relevant whereas the value of tradition is not relevant). Value system estimation is the process of eliciting the value system (that is the set of moral values and their preferences) of an individual citizen. Finally, value system aggregation consists in combining the values of all citizens into a common representation of all society. While there have been multiple proposals in value identification (e.g. [18, 37]) and in value system estimation (e.g. [1, 33]), there has been only one proposed approach to value system aggregation [16]. Furthermore, the authors of this method did not study the ethical properties or describe the use cases for which it is better suited.

This paper aims to explore ideas for possible paths to value system aggregation. In more detail, this paper provides a taxonomy of value system aggregation approaches based on four key dimensions: judgment structure, ethical paradigms of aggregation, generality, and deliberation. Furthermore, we provide an exploration of practical use cases illustrating how different aggregation approaches can address value alignment challenges in policymaking.

2 Motivation: Participatory budgets

Participatory democracy serves as an alternative to representative democracy allowing citizens to get involved in day-to-day policymaking. Although there are many forms of participatory methods, like parliamentary petitioning [23, 24], in this section we focus on participatory budgets (PBs).

PBs allow residents to make proposals to spend a pre-defined budget. Then, citizens vote on the proposals they like most and a resource allocation process selects the proposals which get funded (usually considering votes only). The *Participatory Budget World Atlas* counts more than 10,000 participatory budgeting processes worldwide in 2021. Despite their popularity, PBs suffer from relatively low participation, for example, the Paris PB process (considered one of the most successful) in its 2023 iteration attracted 137.622 voters, which represents about 6.5 % of residents [22]. Furthermore, empirical studies have found low participation leads to bias in the result benefitting wealthier citizens [27]. Consequently, results in PB processes may fail to reflect the priorities of all the population. Given that the approved projects will affect all citizens, it would be desirable to consider votes while also consider the preferences of all citizens (thus correcting the biases of the voter base).

In practice, **value-system aggregation** can be used to find a representation of the society's values. These values can then be considered in the resource allocation process to compensate for the voter bias due to low participation. Serramia et al. [31] show how to find a compromise PB allocation considering both the votes of participants and societal values. Importantly, the authors show with real world data that this compromise solution highly satisfies both participant and non-participant citizens. This shows the usefulness of societal values in AI-aided policymaking applications to increase democratic quality. Hereafter we will use participatory budget allocation as the base example for the different approaches to value system aggregation we discuss.

3 Background: Value Systems

Context. In applied ethics [2,26], reasoning the goodness of actions and decisions with values is dependent on the *context*. This prevails in the AI value alignment literature [18, 21, 32] and thus, for value-guided decision and policy making. For example, if we want to increase welfare in a society with older citizens we might want to invest more in healthcare than in education, whereas in a younger society the reverse might be best. In practice, we can represent contexts in a computer-readable way using any logical language.

Values. A *value* is a deeply held belief that guides decision-making by defining what is important and desirable [29]. We formally characterize values as utility functions that assess the desirability of actions. Given an action a and a context c, each value v is defined by two judgment functions: $v^+(a,c)$ evaluates how good or bad is performing a in the context, while $v^-(a,c)$ evaluates how good or bad is avoiding to perform a in that context. The functions output a number in $[-1,1]$. For instance, in a traffic-accident context the value *benevolence* rates the action HELP positively ($v^+(help, accident) = 1$) while avoiding to perform this action is negatively judged ($v^-(help, accident) = -1$).

Value systems. When making value-aligned decisions we will usually consider more than one value. Naturally, with several values, we may have preferences between them. We formalise these values and preferences as a *value system* which is a pair $VS = \langle V, \succcurlyeq \rangle$ where V is a set of values and $\succcurlyeq$ preferences among them.

Value system aggregation. Given the set of all possible value systems $\mathcal{VS}$, a *value-system aggregation function* is a mapping $F : \mathcal{VS}^n \rightarrow \mathcal{VS}$ that receives n citizen value systems and returns the consensus one. Importantly, although preference aggregation is a well-studied problem in the social choice literature, it cannot be readily applied to value system aggregation, this is for two reasons. First, typical preference aggregation treats options as atomic, whereas in value system aggregation must consider the structure of judgement functions. Second, while preference aggregation emphasises classical social-choice axioms (e.g. monotonicity, independence of irrelevant alternatives), value-system aggregation prioritises ethical properties and application-specific design criteria.

4 Personal vs. Universal Judgements

A fundamental consideration when designing a value system aggregation approach is whether all agents share a common understanding of the values (i.e. their judgement functions). If interpretations vary, the aggregation process must reconcile these differences by establishing a consensus interpretation for each value along with the collective preferences. Conversely, if universal judgements are assumed, the aggregation process is simplified in this regard, as the structure of values remains consistent throughout. Next, we discuss each case in more detail and explain its use cases.

Personal judgements. People usually have different understandings of a value, for example, many people associate the value of security with gun control, while others associate security to being able to defend themselves with guns. If we allow personal judgements, we allow that the value systems to aggregate may contain different understandings of the same value. In this case, the aggregation of the value systems has to be able to aggregate these judgements apart from the preferences.

Use Case 1. Aggregation approaches considering personal judgements are useful in cases where the agents have small differences in their understanding of how values judge actions but do not hold opposite views. Otherwise, the aggregated value judgements may not be shared by anybody. In cases with opposed views it may be better to reconcile them before aggregation.

Universal judgements. This option assumes every value system to be aggregated shares the same judgement functions for its values. For example, if we are making decisions in the traffic context, the value of security will almost universally be linked to the probability of having an accident, there is no room for different interpretations, thus in this case we could assume a universal understanding. Note that, universal judgements eases aggregation (as only preferences have to be aggregated) but may flatten moral nuance.

Use Case 2. Domains in which there is no room for interpretation of values or policy domains with polarised interpretations (pro- vs anti-gun) where differences are so large that in order for the aggregation to make sense we would have to first make all people agree on a common definition.

	Universal judgements	**Individual judgements**
Representation	May not capture all perspectives.	Custom interpretations; diverse perspectives.
Comparability	Preferences directly comparable.	Preferences not directly comparable.
Value Identification	Hard. People may not share the assumed universal judgement.	Easy. Everyone can use their own judgment.
Value Estimation	Easy. Estimate preferences only.	Hard. Estimate preferences and judgments.
Consensus Output	Clear, interpretable consensus value system.	Value judgements may lose meaning.

Table 1: Comparison of Universal and Individual Value Approaches.

Compromise strategies. On the one hand, if the big differences among agents prevent a meaningful consensus, aggregation could focus not on a single output, but rather on first clustering similar value interpretations and then output several possible consensus that preserve the individual interpretations of values. On the other hand, to make shared value judgements more nuanced, value identification could adopt a more exhaustive approach—distinguishing between different interpretations of a value as distinct values themselves. For instance, following on the previous example on guns and the value of security, rather than treating "security" as a single value with multiple interpretations, one might identify "personal security" and "public security" as separate, clearly defined values. This way people in favour of gun regulation would prefer "public security" over "personal security", while the inverse would be true for those in favour of gun ownership. This emphasis on granular value identification illustrates how value system aggregation should inform value estimation, supporting the view that the interaction between value inference steps remains an under-explored yet crucial area of research [17]. Additionally, deliberation between agents may lead to meta-agreements on the concepts being used, which, as discussed in Section 7, aligns with the universal judgement approach.

5 Aggregation Ethical Paradigms

Beyond value judgement choices, the aggregation process may follow different aggregation principles (e.g. fairness). Distance-based aggregation is a widely used aggregation framework [3] that selects the consensus that minimises total distance to all inputs. Thus, given a set of candidate options C, a set of agents $1, \dots, n$, preference orders $\succeq_i$ for each agent i over C, and a distance function d between preference orders, a consensus order can be found using the general distance-based aggregation formula 1:

$$\succeq_{\text{agg}} = \arg\min_{\succeq} \sum_{i=1}^{n} d(\succeq, \succeq_i) \quad (1) \qquad vs_{\text{agg}} = \arg\min_{vs \in \mathcal{VS}} \Big(\sum_{i=1}^{n} d(vs, vs_i)^p \Big)^{1/p} \quad (2)$$

A classic example in social choice is the Kemeny rule which minimizes Kendall's tau distance [14]. Distance-based aggregation have been studied by the work of Gonzalez-Pachón et al. [9] which employs the p-metric distance function to categorize guiding ethical principles; this can be generalised to value system aggregation. Thus, given agents $1, \dots, n$, a value system vs_i for each agent i, a distance function d between value systems, and let $\mathcal{VS}$ be the set of all possible value systems, we can then define distance-based value system aggregation incorporating the ethical parameter $p \in [1, +\infty)$ as in formula 2.

Note this equation defines a family of aggregation functions depending on distance d and parameter p, which allows to consider the trade-offs between overall utility, fairness, and the influence of extreme positions. Table 2 provides guidelines for choosing p.

Utilitarian Aggregation ($p = 1$): With $p = 1$, equation 2 corresponds to minimising the *sum* of distances. This gives equal weight to every agent, delivering a median value system. Minority positions, however, exert negligible influence.

Use Case 3. A citywide participatory budget that has achieved high participation and is representative of the population. Utilitarian aggregation translates the majority's broadly shared priorities into a consensus value system without over-correcting for small extremist factions.

Rawlsian Fairness ($p = \infty$): Setting $p = \infty$ minimises the *maximum* distance to any individual, thus ensuring maximum representativity and fairness. The approach may conflict with majority preferences when opinions are polarised.

Use Case 4. This aggregation method is suited for decision processes where it is important that minorities are given especial care. For example, funding accessibility upgrades (e.g. converting stairs to ramps) only benefits a minority of the

	Utilitarian ($p = 1$)	**Rawlsian** ($p = \infty$)	**Intermediate** ($p > 1$)
Risk of Bias	Ignores minorities	Ignores majority	No risk
Suitability with LDI	Best. The result will be representative	Good as there are no major outliers.	Good as there are no major outliers.
Suitability with HDI	Bad. Minorities will not be represented	Best. The solutions is the fairest to outliers.	Good. Fair solution for everybody.
Applicability	Readily applicable	Readily applicable	Finding p is hard

Table 2: Comparison of ethical paradigms with regards to Risk of Bias, Suitability with Low Dispersion Input (i.e. similar input value systems), Suitability with High Dispersion Input (i.e. cases with high polarisation), and Applicability.

population, hence decisions of this sort are bound to have low support. However, these changes greatly affect people with restricted mobility, if we want to be fair we have to give them heightened impact in the decision process. Rawlsian aggregation allows minority group's values to have as much impact in policy-making processes as possible.

Intermediate Paradigms ($p > 1$**):** Values of p between 1 and ∞ form a continuum that gradually amplifies outlier influence of minorities while preserving overall representativeness. Choosing p therefore becomes a policy lever for calibrated pluralism.

Use Case 5. As previously mentioned, high-income citizens are more likely to participate in participatory budgets than low-income citizens. Selecting an intermediate p can partially boost the under-represented low-income value systems, correcting turnout bias without completely eclipsing the majority.

6 General vs Tailored aggregation

Another important distinction among value-system aggregation methods is whether they are of *general* (i.e. usable in any domain) or *tailored* to one specific decision setting. The only published algorithm so far, the l_p*-regression* rule of Lera-Leri et al. [16], is a general purpose aggregation rule devised without a target application. This approach is ready to be applied straight away and in any scenario.

Use Case 6. A small municipality can estimate a city-wide value system with a general method, obtaining a good-enough input for participatory budget portals, without investing in upfront costs for a tailored aggregation.

Because value-system aggregation extends preference aggregation (Sec. 3), many contexts call for *tailored* aggregation methods which are engineered for a single policy domain and embed explicit moral or application-specific axioms. As indicated in Table 3, this precision comes at higher research, data, and computational cost, yet promises better results where generic rules fall short.

	General Approaches	Tailored Approaches
Availability	Readily available [16].	Not existent yet
Resources	Minimal economic and computational resources needed.	Investment needed. May not be computationally tractable.
Optimality	A "good enough" value system for most applications.	The resulting value system is optimal for the task at hand.

Table 3: Comparison of general and tailored aggregation approaches.

Example: Minimal Decision Divergence. Value system aggregation looks at finding a consensus value system but does not guarantee that the decisions made from that value system are also consensus decisions. To resolve this, we can tailor the aggregation process for this goal. We call **Minimal Decision Divergence (MDD)** the principle of aggregation methods whose aim is to ensure that the decisions made from the consensus aggregated value system are as similar as possible to the decisions each individual value system would lead to. A possible approach to achieve this is to define a distance-based aggregation function (see Equations 1, and 2) where the distance between two value systems is proportional to their decision divergence (i.e. how dissimilar the decisions produced by the two value systems are). In the case of participatory budgets, each citizen value system would lead to different proposals getting funded. Thus, our aim in aggregation is to find a consensus value system whose solution to the participatory budget allocation is as similar to the individual citizen solutions as possible.

Use Case 7. Tailored value system aggregation approaches are useful when the consequences of the decisions made from the aggregated value system are highly relevant, like in policymaking. However, tailored approaches need more economic and computational resources, so when these are not available, general approaches will represent sub-optimal but adequate solutions.

7 Alternatives to aggregation

Beyond formal aggregation methods which employ mathematical calculations, collective value systems can emerge through *group decision-making* (GDM) processes, notably *multi-expert decision-making* (MEDM) [5, 11]. MEDM has been applied in AI governance [4], policy design [13], and forecasting [34], which highlights their real-world applicability.

Deliberation-based approaches seek consensus through iterative dialogue, as illustrated by assemblies [6, 10, 36]. They can surface hidden arguments and foster mutual understanding, but outcomes vary with group dynamics; biases such as *groupthink* [12] and *group polarisation* [35] may arise, and large-scale mobilisation is not expected. Many real systems mix the two logics. Sometimes even participatory budgets feature debates along with a voting phase [30]. The *Delphi method* [7] exemplifies a hybrid: anonymous, multi-round expert questionnaires iteratively refine judgments, mitigating social-bias pitfalls while ending in an aggregated result. Social-choice studies further show that even incomplete deliberation can yield *meta-agreements* on core concepts, easing subsequent aggregation [8, 15, 19, 20, 25]. Deliberation may thus shrink interpretation gaps highlighted in Section 4.

Use Case 8. Value inference (and therefore, value system aggregation) normally needs rich, labelled data [17]. A resource-constrained municipality could instead convene a citizen panel to deliberate and draft a shared value system, avoiding expensive surveys.

	Aggregation-Based	**Deliberation-Based**
Imposition	Aggregation may impose an outcome.	Encourages voluntary compromises.
Reliability	Satisfies formal social choice properties	Prone to social biases (groupthink, polarization...)
Consistent outcomes	Yes. Same inputs yield same results.	No. Context and discussion affect results.
Time Consuming	Fast. One-shot decision-making.	Needs rounds of refinement.
Necessary input	Requires individual value systems obtained beforehand	There is no required input beforehand.
Participation	No recruiting necessary.	Recruiting citizens is necessary.

Table 4: Comparison of aggregation-based and deliberation-based methods.

8 Conclusions and future work

Understanding citizen values is of great utility for policymaking, however the process to obtain a model of citizen values, and in particular *value system aggregation* needs more research. This position paper has explored the possible lines of research to develop better value system aggregation methods with the aim of using them for value-aligned decision making. The sole published algorithm, l_p*-regression* [16], is a personal-judgement, general approach that spans ethical paradigms via a parameter. This means this approach would be useful to obtain a model of citizen values assuming we have tons of citizen data (not only their value preferences but how their understanding each value), we would have to know which ethical paradigm to use, and despite this apparent complexity, the aggregation is not tailored for any specific application. This setting is far from the usual, as governments usually have limited data on their citizens' value preferences, and may need approaches that are more tailored to sensible policy making applications. Future work will develop such specialised aggregators, beginning with the Minimal Decision Divergence criterion (Section 6).

Acknowledgments. Research funded by Ajuntament de Barcelona through Fundació Solidaritat UB, project code 25S03332-001. Marc Serramia is funded by project MMT24-IIIA-01. The funding for this contract comes from the European Union's Recovery and Resilience Facility-Next Generation, in the framework of the General Invitation of the Spanish Government's public business entity Red.es to participate in talent attraction and retention programmes within Investment 4 of Component 19 of the Recovery, Transformation and Resilience Plan.

References

1. Araque, O., Gatti, L., Kalimeri, K.: Moralstrength: Exploiting a moral lexicon and embedding similarity for moral foundations prediction. Knowledge-Based Systems **191**, 105184 (2020). https://doi.org/https://doi.org/10.1016/j.knosys.2019.105184, https://www.sciencedirect.com/science/article/pii/S095070511930526X
2. Beauchamp, T.L., Childress, J.F.: Principles of Biomedical Ethics. Oxford University Press, New York, 8th edn. (2019)
3. Brandt, F., Conitzer, V., Endriss, U., Lang, J., Procaccia, A.D.: Handbook of computational social choice. Cambridge University Press, Cambridge, UK (2016)
4. Brown, E.: Collaborative frameworks for ethical ai governance. Journal of AI Ethics **5**(2), 123–145 (2020). https://doi.org/10.1007/s00146-020-00999-9
5. Butler, C.L., Rothstein, A.: On conflict and consensus: A handbook on formal consensus decisionmaking. Food Not Bombs, Cambridge, MA (1991)
6. Convention Citoyenne pour le Climat: Final report. https://www.conventioncitoyennepourleclimat.fr/en/ (2020), accessed: 2025-02-10
7. Dalkey, N., Helmer, O.: An experimental application of the delphi method to the use of experts. Management Science **9**(3), 458–467 (1963)
8. Dryzek, J.S., List, C.: Social choice theory and deliberative democracy: A reconciliation. British Journal of Political Science **33**(1), 1–28 (2003)
9. González-Pachón, J., Romero, C.: Bentham, marx and rawls ethical principles: In search for a compromise. Omega **62**, 47–51 (2016). https://doi.org/https://doi.org/10.1016/j.omega.2015.08.008
10. Graeber, D.: The Democracy Project: A History, a Crisis, a Movement. Spiegel & Grau, New York, US (2013)
11. Huynh, V.N., Nakamori, Y.: Multi-expert decision-making with linguistic information: A probabilistic-based model. In: Proceedings of the Proceedings of the 38th Annual Hawaii International Conference on System Sciences (HICSS'05) - Track 3 - Volume 03. p. 91.3. HICSS '05, IEEE Computer Society, USA (2005). https://doi.org/10.1109/HICSS.2005.448, https://doi.org/10.1109/HICSS.2005.448
12. Janis, I.L.: Groupthink: Psychological Studies of Policy Decisions and Fiascoes. Houghton Mifflin, Boston, 2nd edn. (1982)
13. Jones, M.: Integrating expert opinions in public policy development. Policy Studies Review **32**(1), 89–110 (2015). https://doi.org/10.1111/psr.12075
14. Kendall, M.G.: A new measure of rank correlation. Biometrika **30**(1/2), 81–93 (1938), http://www.jstor.org/stable/2332226
15. Knight, J., Johnson, J.: Aggregation and deliberation: On the possibility of democratic legitimacy. Political Theory **22**(2), 277–296 (1994)
16. Lera-Leri, R., Bistaffa, F., Serramia, M., Lopez-Sanchez, M., Rodriguez-Aguilar, J.: Towards pluralistic value alignment: Aggregating value systems through lp-regression. In: Int. Conf. on Autonomous Agents and Multiagent Systems. p. 780–788. AAMAS '22, IFAAMAS, Richland, SC (2022)
17. Liscio, E., Lera-Leri, R., Bistaffa, F., Dobbe, R.I., Jonker, C.M., Lopez-Sanchez, M., Rodriguez-Aguilar, J.A., Murukannaiah, P.K.: Value inference in sociotechnical systems. In: Int. Conf. on Autonomous Agents and Multiagent Systems. p. 1774–1780. AAMAS '23, IFAAMAS, Richland, SC (2023)
18. Liscio, E., van der Meer, M., Siebert, L.C., Jonker, C.M., Murukannaiah, P.K.: What values should an agent align with? Autonomous Agents and Multi-Agent Systems **36**(1), 23 (2022). https://doi.org/10.1007/s10458-022-09550-0, https://doi.org/10.1007/s10458-022-09550-0

19. List, C., Luskin, R.C., Fishkin, J.S., McLean, I.: Deliberation, single-peakedness, and the possibility of meaningful democracy: Evidence from deliberative polls. The Journal of Politics **75**(1), 80–95 (2013)
20. Miller, D.: Deliberative democracy and social choice. Political Studies **40**(S1), 54–67 (1992)
21. Noriega, P., Plaza, E.: On autonomy, governance, and values: An agv approach to value engineering. In: Osman, N., Steels, L. (eds.) Value Engineering in Artificial Intelligence. pp. 165–179. Springer Nature Switzerland, Cham (2024)
22. Paris, D.: Budget participatif 2023 resultats. https://www.paris.fr/pages/budget-participatif-2023-114-laureats-devoiles-25161 (2023), accessed February 2024
23. Petitions: UK Government and Parliament. https://petition.parliament.uk/ (2015), accessed 02/2024
24. des pétitions, P.: Assemblée Nationale. https://petitions.assemblee-nationale.fr (2024), accessed 02/2024
25. Rafiee Rad, S., Roy, S.: Deliberation and single-peakedness: A computational study. Journal of Artificial Intelligence Research **70**, 1101–1130 (2021)
26. Ross, W.D.: The Right and the Good. Oxford University Press, Oxford (1930)
27. Saguin, K.: Why the poor do not benefit from community-driven development: Lessons from participatory budgeting. World Development **112**, 220–232 (2018). https://doi.org/https://doi.org/10.1016/j.worlddev.2018.08.009, https://www.sciencedirect.com/science/article/pii/S0305750X18303097
28. Scharfbillig, M., Smillie, L., Mair, D., Sienkiewicz, M., Keimer, J., Pinho Dos Santos, R., Vinagreiro Alves, H., Vecchione, E., Scheunemann, L.: Values and identities - a policymaker's guide. Scientific analysis or review KJ-NA-30800-EN-N (online),KJ-NA-30800-EN-C (print),KJ-NB-30800-EN-Q, European Comission, Luxembourg (Luxembourg) (2021). https://doi.org/10.2760/349527(online),10.2760/022780(print),10.2760/059689
29. Schwartz, S.H.: An overview of the schwartz theory of basic values. Online readings in Psychology and Culture **2**(1), 2307–0919 (2012)
30. Serramia, M., Ganzer-Ripoll, J., López-Sánchez, M., Rodríguez-Aguilar, J.A., Criado, N., Parsons, S., Escobar, P., Fernández, M.: Citizen support aggregation methods for participatory platforms. In: Sabater-Mir, J., Torra, V., Aguiló, I., Hidalgo, M.G. (eds.) Artificial Intelligence Research and Development - Proceedings of the 22nd International Conference of the Catalan Association for Artificial Intelligence, CCIA 2019, Mallorca, Spain, 23-25 October 2019. Frontiers in Artificial Intelligence and Applications, vol. 319, pp. 9–18. IOS Press, Amsterdam (2019). https://doi.org/10.3233/FAIA190102, https://doi.org/10.3233/FAIA190102
31. Serramia, M., Lopez-Sanchez, M., Rodriguez-Aguilar, J.A., Moretti, S.: Value alignment in participatory budgeting. In: Proceedings of the 23rd International Conference on Autonomous Agents and Multiagent Systems. p. 1692–1700. AAMAS '24, International Foundation for Autonomous Agents and Multiagent Systems, Richland, SC (2024)
32. Serramia, M., Rodriguez-Soto, M., Lopez-Sanchez, M., Rodriguez-Aguilar, J.A., Bistaffa, F., Boddington, P., Wooldridge, M., Ansotegui, C.: Encoding ethics to compute value-aligned norms. Minds and Machines **33**(4), 761–790 (2023). https://doi.org/10.1007/s11023-023-09649-7, https://doi.org/10.1007/s11023-023-09649-7
33. Siebert, L.C., Liscio, E., Murukannaiah, P.K., Kaptein, L., Spruit, S., Van Den Hoven, J., Jonker, C.: Estimating value preferences in a hybrid participa-

tory system. In: HHAI2022: Augmenting Human Intellect, pp. 114–127. IOS Press, Ansterdam, Netherlands (2022)
34. Smith, J.: Advanced Techniques in Forecasting. Academic Press, New York, NY (2010)
35. Sunstein, C.R.: Deliberation and polarization. Political Philosophy **3**(1), 1–24 (2002). https://doi.org/10.1002/9780470693711.ch1
36. The Citizens' Assembly: The eighth amendment of the constitution. https://2016-2018.citizensassembly.ie/en/The-Eighth-Amendment-of-the-Constitution/ (2017), accessed: 2025-02-10
37. Wilson, S.R., Shen, Y., Mihalcea, R.: Building and validating hierarchical lexicons with a case study on personal values. In: Staab, S., Koltsova, O., Ignatov, D.I. (eds.) Social Informatics. pp. 455–470. Springer International Publishing, Cham (2018)