

Outline

Moral Parliament Tool for Allocating Healthcare Resources

Authors: Hayley Clatterbuck, Bob Fischer, Arvo Muñoz Morán, Derek Shiller

Note: a beta version of the Moral Parliament tool for healthcare allocations can be accessed at <https://bioethics-version.parliamentary-project.pages.dev/projects> . A previous version of the tool with more functionality can be found at <https://parliament.rethinkpriorities.org/> .

Abstract

The Moral Parliament Tool provides a framework for modeling participants' general moral views and simulating democratic deliberation among them. We show how the tool can be used to support democratic decisions about how to allocate healthcare resources. We apply the tool to three different bioethics case studies, showing how a hypothetical moral parliament composed of delegates of diverse worldviews would decide to distribute resources. This approach provides a more generalizable alternative to existing methods for participatory algorithm design.

1. Introduction

Decisions about the best way to allocate limited resources, like vaccines and kidneys, depend on both empirical and moral assumptions. Policymakers can rely on public health experts for guidance about the allocation schemes that will most efficiently satisfy our aims. However, the aims themselves are products of values about which there may not be consensus. To respect the diverse values of stakeholders, achieve legitimacy, and gain trust, many bioethicists argue that policymakers should involve stakeholders in healthcare allocation decisions via direct democratic deliberation (Mitton, *et al.* 2009; Abelson, *et al.* 2013; Degeling, *et al.* 2015). This requires establishing transparent and fair deliberative processes that enable stakeholders' involvement and oversight (Gruskin and Daniels 2008).

We present an algorithmic method, the Moral Parliament Tool, as a component of one such process. This tool allows us to model stakeholders' moral values and represent them as participants in democratic decision-making processes that deliver recommendations about how to allocate resources. It makes several novel contributions to existing participatory algorithm design efforts. By modeling users' general moral outlooks or worldviews, it is generalizable to many decision domains and allows for the integration of expert opinion about effectiveness. It incorporates several different democratic processes, including forms of voting and bargaining, allowing us to model not just stakeholders' first-order moral values but also their preferences regarding how to make decisions in the face of disagreement. Lastly, it includes pedagogical features that increase transparency, making it easier for users to understand why particular inputs lead to particular allocations.

2. Background and similar approaches

Decisions about how to allocate important healthcare resources are often made by public health professionals or regulatory bodies with little direct input from the public. Advocates of citizen participation in decision-making argue that it promotes citizens' democratic rights, leads to better decision-making, and enhances legitimacy and public trust (OECD 2022). Existing frameworks for democratic deliberation typically include assembling a representative group of participants in a "mini public." After being given additional information about the issue at hand, participants discuss in a structured way; then, they make recommendations (on, e.g., a policy proposal or suggested budget).

These approaches have several limitations (see Dullaghan 2019 for a review). First, it can be expensive to gather and compensate a group of citizens and experts for an extended period of deliberation. Second, it can be difficult to identify and recruit a representative sample of participants. Third, the process can fail in many ways—including, e.g., within-group deliberations being dominated by demographics that already have outsized voices in decision-making. Fourth, the outputs of focused deliberation may be so specific to particular policies or budgets that they fail to generalize. Lastly, though citizen participation is intended to temper the influence of experts, this could be taken too far. If participants are not (and perhaps cannot be) sufficiently informed about the intricacies of the problem under discussion, then their decisions may produce outcomes that are suboptimal even by their own lights.

Digital tools offer promising alternatives to in-person democratic deliberation. These tools can be much less costly and allow for many more voices to be included. While some of these tools seek to replicate the dynamics of in-person deliberation on online platforms (Horton 2018), a more radical approach is to automate the process of deliberation itself. Algorithmic methods for democratic deliberation seek to model individuals' values and preferences, "letting our models of multiple people's moral values vote over the relevant alternatives" in an automated "deliberative" process (Conitzer, *et al.* 2017).

In participatory algorithm design processes, participants' inputs are used to construct or calibrate an algorithm that will determine policy, like food allocation from a central foodbank (Lee, *et al.* 2019); actions taken by an autonomous vehicle (Noothigattu, *et al.*, 2018); or a method for coordinating kidney donations (Freedman, *et al.* 2020). Typically, participants' values are inferred by presenting them with pairs of outcomes that differ in morally relevant ways and asking which they prefer.¹ A discrete choice model is then used to infer how much weight a user places on various moral values (Awad, *et al.*

¹ Lee, *et al.* (2019) also allow participants to define their own models in the form of explicit rules for how various aspects of value should be weighted and used. Participants can choose whether they are represented by their learned or stated model. Freedman, *et al.* (2020) introduce explicit participant input by polling people about which factors are appropriate for allocating kidneys (e.g., age of recipient) and which are not (e.g., race of recipient).

2020).² For example, Freedman, *et al.* present participants with pairs of hypothetical individuals who vary with respect to important features (e.g., age, health status, etc.) and ask who ought to be prioritized to receive a kidney. They then aggregate participants' responses and score how much each feature contributed to priority judgments (e.g., strong preference for young over old recipients). Finally, these scores are used as weights in kidney exchange models.

Because the participatory algorithm design process is designed to generate an algorithm that makes decisions in a particular domain, most existing work models participants' values as constrained to that domain. Most often, they model participants' direct preferences over the actions that are the algorithm's target (e.g., giving a kidney to a young vs. elderly person, hitting 5 pedestrians over 1). This approach has at least two limitations. First, models of participants' judgments about specific applied domains will not easily generalize, requiring us to construct new models for each decision. Second, a person's judgments about specific actions depend on both moral values and empirical matters about which participants may lack relevant expertise. By fitting algorithms to these judgments, we risk making allocations dependent on mistaken empirical judgments, ones that participants themselves would not endorse given more information.

The limitations of specificity motivate more general methods for automated deliberation. Models of participants' general moral views could potentially be used as representatives across a much wider range of applications. By separating moral values from empirical beliefs, we can make space for expert judgments about the latter and potentially involve a wider group of participants who need less background information about the domain in question. For example, instead of convening separate groups to deliberate about vaccine distribution and maternal health investment proposals, we could re-use the same participants' moral views for each, calling on different groups of experts to connect those moral views to policy judgments in each domain.

Modeling domain-general moral views and using them to draw conclusions about concrete policy proposals presents its own set of challenges (Allen, *et al.*, 2005). We discuss some of them below. However, we stress that the question is not whether to model stakeholders' domain-specific views or domain-general views. Rather, the question is whether there are good reasons to model stakeholders' domain-general views *in addition* to their domain-specific views. All algorithmic approaches to democratic decision-making are attempts to address some mix of practical and principled difficulties associated with realizing democratic ideals. Each has various pros and cons that should be understood in detail. Then, they can be inputs to a broader decision-making process that is enriched by a more comprehensive understanding of stakeholder preferences.

² Approaches differ with respect to when and how individuals' responses are aggregated. Lee, *et al.* (2019) and use pairwise comparisons to build models of each individual and then these distinct models vote based on the Borda rule. Noothigattu, *et al.* (2018) builds models of each individual, which it then combines into a single model of the aggregate. Then, the collective uses a voting rule (such as Borda) that satisfies certain desirable features. Freedman, *et al.* (2020) does not model each individual's values but starts by aggregating users' judgments and uses the frequency with which participants favored an option to estimate the overall value of that option, via a Bradley-Terry model. The Moral Parliament Tool is most similar to the process used by Lee, *et al.* in that we arrive at allocations via deliberation among distinct models of individuals. However, we include more voting and non-voting methods in addition to Borda.

3. Moral Parliament Tool

The Moral Parliament Tool provides a framework for modeling participants' general moral views and simulating democratic deliberation among them. Deliberations take place over a set of projects, and the tool outputs a recommended allocation of resources across those projects. There are three steps to this democratic algorithmic process:

- I. Eliciting and modeling participant worldviews
- II. Determining the policy recommendations of each worldview
- III. Simulating democratic deliberation among participants

I. Eliciting and modeling worldviews

A worldview comprises a set of moral and decision-theoretic commitments that have implications for the choiceworthiness of various aims. Normative dimensions of a worldview may include:

- a. **Who matters:** which entities who might be affected by our actions matter morally, and to what extent to they matter? Examples: how should we weight impacts on children vs. adults? Should we take effects on other species into consideration?
- b. **What matters:** which features of an outcome are relevant to assessing its value? Examples: how much should we focus on suffering, autonomy, or other outcomes? Should we assess outcomes by their total welfare gains or does the distribution of outcomes matter?
- c. **How we should act:** what types of actions should we take to promote things that matter? Examples: how much should we prioritize avoiding bad outcomes versus trying to achieve good outcomes? Should we prefer systemic or surgical interventions?

Each dimension can take various values, reflecting different normative judgments. For example, a user's views about different age groups could be captured by numbers assigned to classes *Children*, *Adult*, and *Elderly* that reflect the relative "moral weight" given to each.

The Moral Parliament tool allows users to choose from a menu of pre-set worldviews (e.g., utilitarianism or libertarianism) or to configure their own idiosyncratic worldviews by specifying their normative commitments. The tool also allows for the construction of new normative dimensions, making it highly adaptable to new decision contexts.

II. Determining policy recommendations for each worldview

A central challenge for this methodology lies in deriving the policy implications of abstract worldviews. The overall score (or utility) that a worldview gives to a project will be a function of how well that project promotes the things the worldview cares about. This depends on two factors. First, the magnitude (or scale) of the project's effects is how much good the project would achieve if one were to

fully value all of the things the project does.³ Second, we ask what proportion of the overall effect is valued by the worldview.

Importantly, judgments about scale and how much a project promotes different kinds of value (and hence project utilities) are not derived from users' direct assessments of particular projects but are instead supplied exogenously. The user specifies their ends, but judgments about which projects would be means to those ends are derived from expert knowledge. The normative dimensions framework provides important structure to guide experts in connecting abstract moral worldviews to evaluations of concrete proposals.

For illustration, a project that distributes 50 million vaccines will have a larger scale than one that distributes 1 million and therefore ought to receive a higher score from worldviews that approve of vaccine distributions. If *Age of Beneficiary* is a relevant dimension, we specify what proportion of the project's benefits accrue to beneficiaries of various ages. If the vaccine project primarily targets youth, it will be highly valued by a worldview that puts a high weight on children's health, while a worldview that favors the elderly will value a much smaller proportion of the project's effects.

From the project scores, we can derive the overall utility of a proposed allocation of resources across projects, which reflects a judgment about how much value an allocation would achieve by the lights of that worldview. If we assume that projects do not have any diminishing returns, then the value of an allocation will be the sum of the project scores weighted by the amount of resources given to each project. The tool allows users to add diminishing returns, where more steeply diminishing returns tend to favor more diverse allocations of resources.

III. Deliberation among worldviews

Once we have worldview models that represent each participant and have derived the scores that each worldview assigns to each project, we can simulate deliberations among the participants. The Moral Parliament Tool includes a suite of different deliberative procedures, including:

- a. Voting: approval, Borda, ranked choice
- b. Bargaining: Nash bargaining, proportional allocation (Moral Marketplace)
- c. Social choice functions: maximin, Maximize Expected Choiceworthiness, My Favorite Theory

Each method outputs a recommended allocation of resources across projects. When available, the Tool also reports relevant information, such as which worldview is least satisfied or which worldviews approved of the final decision.

³ There are several ways we could define scale. Here, we define the scale of a project as its expected value. In previous iterations of the parliament, we defined scale as the expected value of the project conditional on its being successful, and we specified its probabilities of success, failure, and backfire. We then included a normative dimension of risk, where different worldviews put different weights on success, failure, and backfire (e.g. a risk averse worldview puts significant decision weight on worst-case outcomes). Including attitudes toward risk will be especially important in cases where projects have very different risk profiles that are not well-captured by their expected value. We judged risk to be less important in the case studies we use here, since few projects we consider here carry large risks of failure or backfire.

4. The Parliament at work: Three examples

To illustrate, we construct models of six hypothetical participants and apply their worldviews to three different sets of bioethics-related resource allocation problems.

Normative dimensions

We characterize worldviews and projects by the following normative dimensions:

Normative dimension	Worldview weighting	Project score
Egalitarianism vs. prioritarianism	Relative priority assigned to benefiting the worse off versus distributing benefits equally	Proportion of benefits that accrue to worse-off versus better-off individuals
Time of effect	Relative weight assigned to present versus future people	Proportion of benefits that accrue to present or future people
Incremental vs. systemic change	Preference for acting via systemic changes to social structures versus surgical, incremental changes	Relative strength of systematic versus incremental changes effected by project
Age of beneficiaries	Relative priority assigned to benefiting children, adults, and the elderly	Proportion of benefits that accrue to children, adults, and the elderly
Saving vs. improving lives	Importance assigned to saving lives versus improving welfare of living people	Proportion of project benefits that involve saving versus improving lives
Autonomy vs. general well-being	Relative priority placed on people's abilities to make autonomous choices versus effects on general well-being	Relative strengths of project's effects on autonomy versus general well-being

Worldviews

We have constructed six hypothetical participants whose worldviews can be characterized as follows (scores for each dimension are provided in Appendix A):

1. **Social Engineer:** Seeks to transform healthcare systems to maximize overall health of future societies
2. **Rule of Rescue:** Prioritizes helping people who are currently in the most immediate danger

3. **Capabilities:** Focuses on ensuring that every individual has the freedoms and resources required to flourish
4. **Utilitarian:** Aims to maximize the overall amount of wellbeing and minimize overall suffering
5. **Libertarian:** Prioritizes autonomy and letting individuals make their own healthcare decisions
6. **Redress past inequalities:** Seeks to distribute resources in ways that make up for past healthcare injustices and increase the welfare of disadvantaged groups

Projects

We convene these individuals into virtual parliaments that will decide how to allocate resources across candidate projects. (Normative dimension scores for each project are provided in Appendix B.) To illustrate the generality of our parliament, we consider three distinct bioethical distribution problems.

Problem 1: Infectious disease

A local government has received a federal block grant to combat COVID-19. It is considering five projects:

- Enhancing air filtration in schools
- Mask mandate enforcement in healthcare facilities and public transit
- Distributing antivirals to hospitals to treat people with severe cases
- Funding research to discover a novel, improved vaccine for the disease
- Vaccination outreach campaign to increase use of current vaccines

Problem 2: Maternal healthcare

Gruskin and Daniels (2008, 1574) present the case of a national government that wants to improve maternal health and considers the following five projects:

- Education about family planning services for married women
- Emergency OB facilities in urban areas, targeting underserved populations
- Training and placement of attendants in health centers in rural areas
- Advocacy to change laws allowing girls to marry at young ages
- Outreach to increase enrollment of girls in school

Problem 3: Global health

A philanthropic organization is looking to allocate funds to global health charities operating in a target developing country. It considers charities that do the following:

- Investment to build fundamental health infrastructure
- Malaria treatment for infected adults
- Malaria prevention for young children
- Initial research on a new malaria vaccine
- Direct cash payments to families

Constructing a moral parliament

To use the Moral Parliament tool, the first step is to select a set of projects to be considered. Next, users select the set of worldviews that will be represented in the parliament, altering or constructing worldviews as needed. At this stage, users can see the utilities that a worldview assigns to the candidate projects. Again, these utilities are derived from the weights the worldview assigns to each normative dimension and how much value a project provides according to those dimensions.⁴

For illustration, consider the set of five COVID projects. The Rule of Rescue worldview finds the highest utility in providing antivirals to hospitals, which makes sense given its prioritization of people who are currently at the highest risk of dying:

Project	Worldview Valuation
Antivirals	31.45
Mask Mandates	8.39
Vaccine Outreach	3.80
Air Filtration	0.22
Vaccine Research	0.20

Figure 1: Project scores assigned by the Rule of Rescue worldview.

The Libertarian favors projects that promote people's abilities to make their own healthcare decisions. Accordingly, they favor voluntary vaccine outreach and strongly disfavor mask mandates:

Project	Worldview Valuation
Vaccine Outreach	18.71
Air Filtration	6.41
Antivirals	5.02
Vaccine Research	3.90
Mask Mandates	1.32

Figure 2: Project scores assigned by the Libertarian worldview

The next step is to populate the parliament with representatives of each individual. If multiple people share a worldview, this can be done by adding extra delegates of that worldview.⁵

⁴ For the three illustrations we present here, we assigned each project the same scale, which means that each project has the same expected value (were you to fully value everything it achieves). This is unrealistic, but it allows us to focus on normative disagreements rather than empirical uncertainty about the cost effectiveness of various projects.

⁵ The Parliament tool can be used to represent individuals' moral uncertainty by adding multiple delegates to represent a single person. For example, if a participant gives equal credence to the capabilities and utilitarian

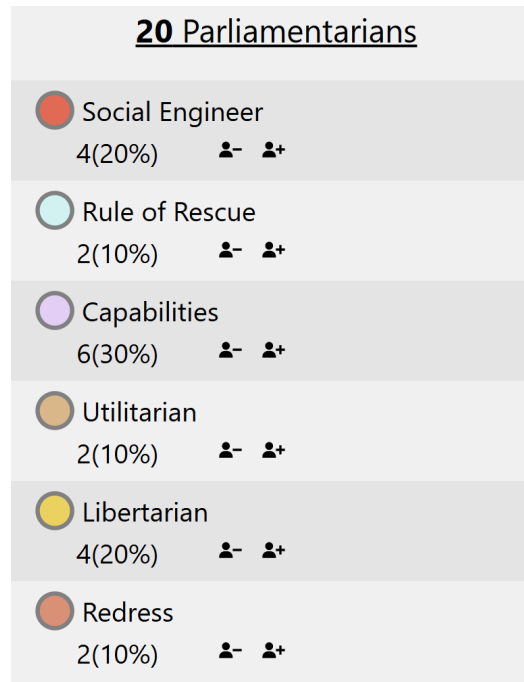


Figure 3: Composition of a hypothetical moral parliament.

Allocation strategies

The final step is to choose a deliberative method, an Allocation Strategy, that the parliament will use to arrive at a final allocation.⁶ The Moral Parliament tool includes eight different methods (three social choice functions, three voting methods, and two bargaining methods) that a parliament can use to arrive at a final allocation, given the preferences of participants:

Allocation Strategy	Description
My Favorite Theory	Selects the favored allocation of the most popular worldview in the parliament.
Maximin	Rawlsian social welfare function. Selects the allocation with the highest utility for the least-satisfied worldview.
Maximize Expected Choiceworthiness	Selects allocation that maximizes the sum of the utilities across worldviews, weighted by worldviews' representation in the parliament.
Approval Voting	Parliamentarians approve of allocations that surpass a threshold of utility. The allocation with the most approval votes wins.

approaches, we could represent her by giving one delegate to each, making sure that the other participants are also represented by the same number of delegates.

⁶ Users can choose whether to include decreasing marginal returns. Some Allocation Strategies also have additional optional settings. For example, users can select among different tie-breaking strategies for approval voting.

Borda Voting	Worldviews give higher scores to projects they rank more highly. The project with the largest combined score, weighted by worldviews' representation in the parliament, wins. This method selects a single project, not an allocation.
Ranked Choice Voting	Worldviews rank candidate projects. If no candidate receives a majority of first-choice votes, the lowest-scoring candidate is eliminated and the votes are redistributed to voters' next choices, continuing until one candidate achieves a majority.
Moral Marketplace	Each worldview is given a share of the budget, proportional to their representation in the parliament, to spend as it likes.
Nash Bargaining	Moral Marketplace, with the ability to move to alternative allocations that improve upon a worldviews' base allocation.

5. The Parliament at work: selected results

We will start by considering a parliament in which each worldview is represented equally (e.g., deliberations among 6 participants, each represented by a different worldview).

COVID project deliberation

The results of the parliamentary deliberations depend on the allocation strategy used. Here, voting methods favor vaccine outreach and research. Vaccine outreach wins the Borda vote, as every worldview ranks it at least fairly highly. Approval voting favors allocating the budget to just these two projects.⁷

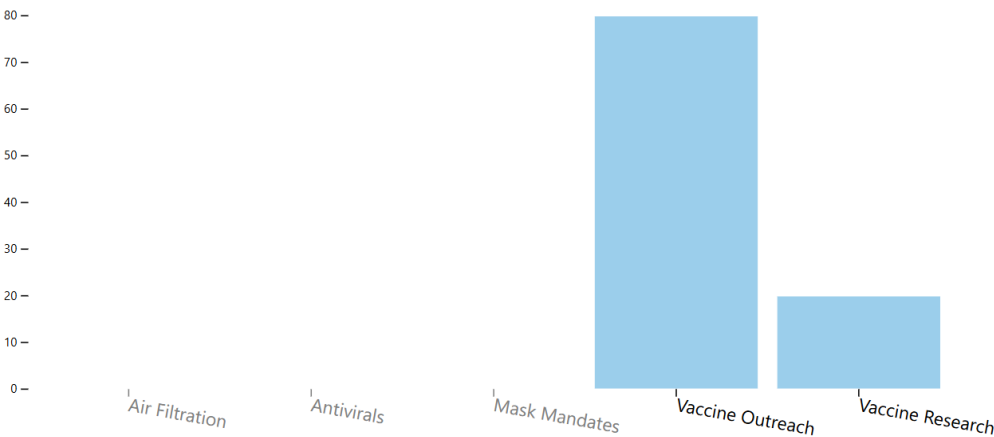


Figure 4: The resource allocation with the highest approval rate when equal worldview is equally represented.

⁷ It is worth noting that the winning allocation was only approved of by half of the voters. If we set marginal returns to decrease, then approval voting recommends giving most of the budget to vaccine projects and a smaller share to the other three projects.

Bargaining methods favor more diverse allocations because each worldview is given control over $\frac{1}{n}$ of the budget and they favor different projects. Since none of the worldviews assign much utility to mask mandates, it does not receive any funding (though it would if projects have decreasing marginal returns).

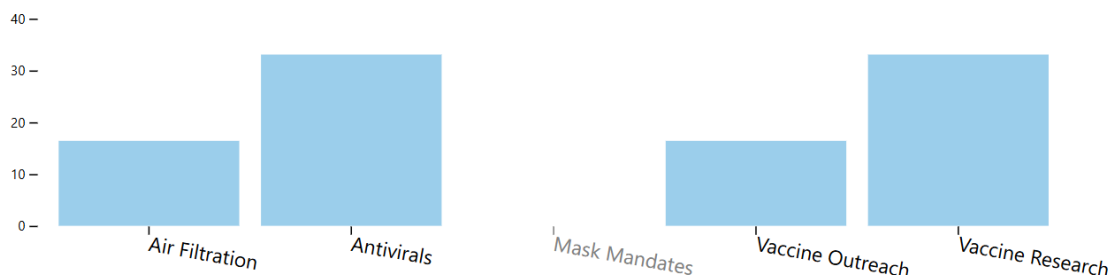


Figure 5: The resource allocation resulting from a “moral marketplace”, where each worldview gets a proportional share of the budget to spend as they like.

Social welfare functions have mixed results. Maximizing Expected Choiceworthiness picks the allocation with the highest weighted average utility. Here, it recommends giving everything to vaccine research since the worldviews that like it (Social Engineer, Utilitarian) so strongly favor it. Maximin selects the allocation that maximizes the payoff to the worst-off worldview. Here, it recommends a very even allocation across projects, which no worldview finds overly objectionable. Because all worldviews are equally represented, My Favorite Theory doesn’t yield a result.

Changing the makeup of the parliament has largely predictable effects. For example, if we add additional Social Engineers and Utilitarians, nearly all methods now favor vaccine research. Changes in parliament composition interact with the choice of allocation method, e.g. bargaining methods are more resistant to change, since minority worldviews still control some portion of the budget.

Maternal health deliberation

There are some strong similarities among maternal health projects that affect the outcome of the parliament. Campaigns against child marriage and for girls in school tend to perform very similarly, and the former tends to be slightly higher-rated. Likewise for emergency OBs and rural attendants. As a result, when there are no diminishing returns, bargaining and social choice methods tend to give funding to the slightly preferred project in each pair. For example, Moral Marketplace yields the following:

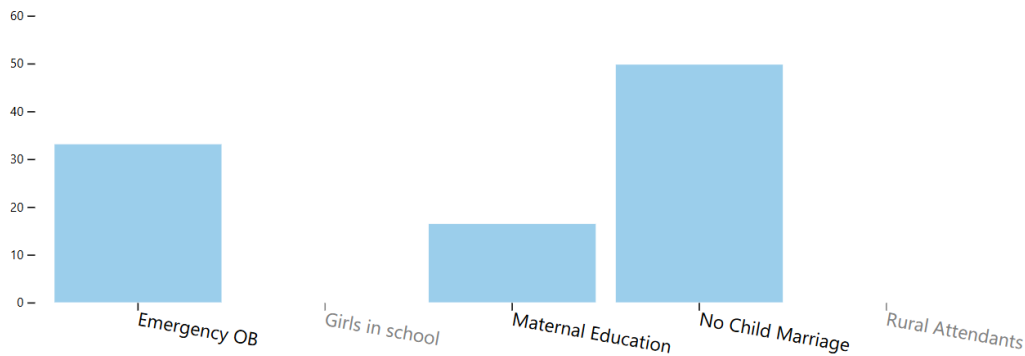


Figure 6: The resource allocation resulting from a “moral marketplace”, where each worldview gets a proportional share of the budget to spend as they like.

When projects have diminishing marginal returns, most methods yield very even allocations across all projects, reflecting the fact that many projects are assigned very similar utilities.

Voting methods pick out either emergency OB or child marriage projects, as they are the top choice of multiple worldviews.

Global health deliberation

While there is significant disagreement across worldview rankings of global health projects, malaria projects tend to be fairly well-liked. When the parliament votes, malaria treatment for adults slightly edges out malaria prevention for children since adult survival rates are somewhat lower and adults suffer greater autonomy losses. Bargaining approaches also allocate most of the money to malaria causes, but some money is given to direct payments (the Libertarian’s favorite) and health infrastructure (the Social Engineer’s favorite).

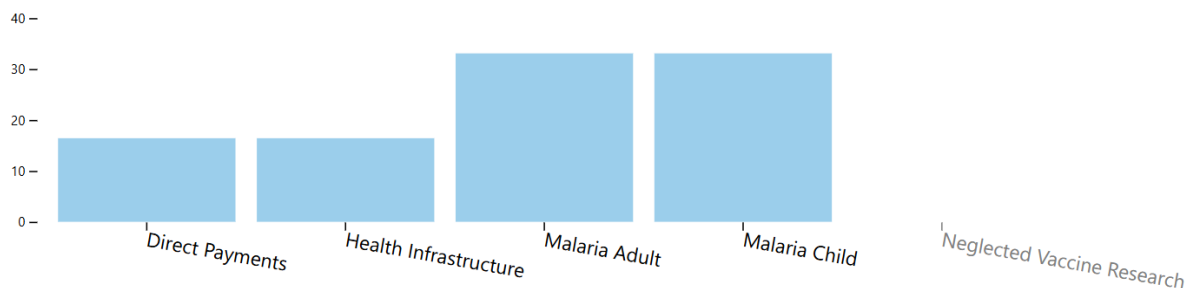


Figure 7: The resource allocation resulting from a “moral marketplace”, where each worldview gets a proportional share of the budget to spend as they like.

6. Discussion

The Moral Parliament Tool supports a wide range of inferences from a small set of inputs. The same set of worldview models was used to deliberate about projects in different domains, without requiring any domain-specific judgments from the participants. It can simulate how a parliament’s decisions would

change given different parliamentary methods. It can also simulate how changes to the composition of the parliament would change the results. This will be particularly helpful when modeling populations (instead of specifically modeled individuals) and when trying to understand how changes in public opinion would change democratic deliberations.

The results of the Parliament are specific to project sets and the composition of the parliament. Nevertheless, we can identify a few general trends. First, deliberative strategies matter a lot. Some methods, like maximizing expected choiceworthiness, tend to go all-in on top projects and can be strongly influenced by a small subset of worldviews. Others, like Nash bargaining or approval voting, favor diversifying widely across projects and cause areas and reflect the judgments of a broader set of delegates. When projects have diminishing marginal returns, all methods favor more equal allocations across projects.

The ability of our Moral Parliament Tool to faithfully simulate the moral deliberations of a group will depend on how well we can elicit and model participant worldviews, infer the policy implications of worldviews, and model deliberative methods. We'll consider these in order.

There are several methods for eliciting and modeling users' worldviews. First, we could present users with a list of worldviews corresponding to well-established moral theories.⁸ This requires a familiarity with moral theories that many participants may lack. We also expect that this method will be too coarse-grained, failing to capture nuanced and idiosyncratic worldviews. Second, participants could construct their own worldviews by specifying their positions for each normative dimension. An interesting alternative would be an expert- or LLM-assisted method that matches people to worldviews based on conversation (Sorensen, *et al.* 2025).

Among participatory algorithm design frameworks, preference modeling on pairwise comparisons is the most common method for modeling values. Participants are presented with pairs of options (A vs. B) that differ in normative dimensions. From their pattern of preferences across options that vary with respect to normative dimensions, we can infer a subject's views about each dimension.⁹ For example, the Moral Machine asks people to decide which action an autonomous vehicle should take, where each action will hit some number of pedestrians (Awad, *et al.* 2018). Scenarios vary on 22 dimensions, including the number of pedestrians killed, whether the pedestrian is illegally crossing, the age of the pedestrian, etc. After a person makes dozens of A vs. B comparisons, we can infer, for instance, whether they favor young vs. old pedestrians and how strong of a driver the age variable is in their decisions.

We are uncertain which of these elicitation methods will work best. Lee, *et al.* (2019) found that models learned through preference modeling were more accurate and were favored by subjects than models that were explicitly specified by subjects themselves. However, preference modeling may be

⁸ For example, we can model the Kantian worldview as: only giving weight to existing humans; more concerned with avoiding harms than beneficence; primarily focused on justice and rights; and so on. In contrast, a utilitarian: gives significant weight to other animals; is focused on suffering; and doesn't distinguish between harms and beneficence.

⁹ Lee, *et al.* (2019) modeled subjects' values on the basis of pairwise comparisons and also had subjects explicitly specify their values. They compared how well each of these models performed when predicting subjects' preferences over a new set of pairwise comparisons. They found that the learned models were more accurate and were favored by subjects.

significantly more difficult for general worldviews. More work is needed to design comparisons that elicit useful worldview information.

Another concern is that our methodology may not be able to represent some moral worldviews, especially deontological ones. Currently, we treat normative dimensions as additive contributors to a project's overall utility. However, many worldviews posit absolute prohibitions on certain kinds of actions (e.g., rights violations) that shouldn't be modeled as simple utility subtractions. We think that there are promising frameworks for modeling deontological theories in a utility framework (Lazar 2017, Lazar & Graham 2021). Further, there are other ways of formulating normative dimensions within the tool that could capture non-additive kinds of values.

The second task is to infer how worldviews would assess a given project or allocation of resources. A key innovation of the Parliament Tool is its systematization. Experts do not have to consider particular worldviews and projects one by one; they merely need to assess how well a project satisfies each normative dimension. This process will be straightforward for some dimensions (e.g., age of beneficiary) but potentially quite difficult for others, especially those whose interpretations may be normatively laden. For example, the egalitarian vs. prioritarian dimension specifies what proportion of a project's benefits accrue to the worst off, but there are different senses in which someone could be worse off (e.g., historically marginalized, poorest, in worst health). The challenge is to specify normative dimensions that are precise enough to evaluate projects, general enough to capture broad worldviews, and have relatively objective standards of application. The reliability of our methodology could be tested by surveying bioethics experts to see how much convergence there is in their judgments about which policy recommendations follow from particular worldview models.

The third key task is to faithfully model various deliberative procedures. This step rests on the firmest ground, given the extensive literature about voting, bargaining, and social choice theory; for an extensive discussion of methods used in the Parliament tool, see Clatterbuck, *et al.* (2024). The Tool uses established formal methods to aggregate individual utility functions into group decision-making. What is less well-known is which of these methods will best achieve the aims of participatory deliberation. It would be helpful to survey participants to evaluate which of these methods they find the most fair, which they deem to confer the most democratic legitimacy, and which gives the most acceptable results (evaluated by participant rankings of the blinded recommendations of different methods).

Finally, though we have emphasized the advantages of modeling people's general worldviews, domain-general and abstraction from concrete details present challenges. There is widespread disagreement about how best to interpret common moral principles, such as fairness or beneficence, so accurately modeling people's moral views based on their stated values may be difficult. Normative dimensions may be context-dependent. For example, someone might assign different moral weights to children and adults across different contexts, e.g. with adults mattering more with respect to political rights and children more in health contexts. We could treat normative dimension judgments as simplifications or reasons to restrict the domain of application of a person's parliament model. Alternatively, we could include additional normative dimensions to capture this contextual variation.

Setting aside concerns about empirical errors, people's judgments about particular actions might differ from what their moral views would recommend. We rely on expert judgments to connect worldviews to judgments about projects, thereby inferring what projects a person *should* support, given her values. However, it's possible that this may be quite different from what people actually *would*

support. To the extent that there is disagreement between the two, it is an open question which judgments should serve as input to democratic deliberation.

Conclusion

The Moral Parliament Tool represents a promising approach to enhance democratic deliberation in healthcare resource allocation decisions. By modeling stakeholders' general moral worldviews rather than their domain-specific judgments, the tool offers several advantages over existing participatory methods. It allows for greater generalizability across decision domains, creates space for expert empirical judgments while still representing diverse stakeholder values, and can simulate multiple democratic processes to reflect not only first-order moral values but also metanormative preferences about decision-making under disagreement.

The Moral Parliament Tool is not intended to replace human deliberation entirely, but rather to complement and enhance existing democratic processes. It offers an efficient way to model the implications of diverse moral commitments and to gain insight into how different deliberative procedures might resolve moral disagreements. Future work should focus on developing and testing effective elicitation methods, exploring ways to represent non-consequentialist moral views, and evaluating which deliberative procedures best achieve the aims of participatory democracy.

Acknowledgements

We'd like to thank Christopher Woodard and participants at the Critical Perspectives on Democratic Deliberation in Health and Science Policy conference for helpful comments. We'd also like to thank David Moss, Adam Binksmith, Willem Slegers, Zachary Mazlish, Gustav Alexandrie, John Firth, Michael Plant, Ella McIntosh, Will McAuliffe, Jamie Elsey and Eric Chen for the helpful feedback provided on the tool.

References

- Abelson, J., Blacksher, E. A., Li, K. K., Boesveld, S. E., & Goold, S. D. (2013). Public deliberation in health policy and bioethics: mapping an emerging, interdisciplinary field. *Journal of Deliberative Democracy*, 9(1).
- Allen, C., Smit, I., & Wallach, W. (2005). Artificial morality: Top-down, bottom-up, and hybrid approaches. *Ethics and information technology*, 7, 149-155.
- Awad, E., Dsouza, S., Kim, R., Schulz, J., Henrich, J., Shariff, A., ... & Rahwan, I. (2018). The moral machine experiment. *Nature*, 563(7729), 59-64.
- Cabannes, Y. (2004). Participatory budgeting: a significant contribution to participatory democracy. *Environment and urbanization*, 16(1), 27-46.

Clatterbuck, H., Muñoz Morán, A., and Shiller, D. Metanormative methods.

https://docs.google.com/document/d/1pOzOpVxGVSoGW6n4h-BoFqrfzQQAoVj8hzk_VQf8dfA/

V. Conitzer, W. Sinnott-Armstrong, J.S. Borg, Y. Deng, M. Kramer, Moral decision making frameworks for artificial intelligence, in: AAAI, 2017, pp.4831–4835, Blue Sky track.

Degeling, C., Carter, S. M., & Rychetnik, L. (2015). Which public and why deliberate?—A scoping review of public deliberation in public health and health policy research. *Social Science & Medicine*, 131, 114-121.

Dullaghan, N. (2019). Deliberation may improve decision-making. *Rethink Priorities*.

Freedman, R., Borg, J. S., Sinnott-Armstrong, W., Dickerson, J. P., & Conitzer, V. (2020). Adapting a kidney exchange algorithm to align with human values. *Artificial Intelligence*, 283, 103261.

Horton, C. (2018). The simple but ingenious system Taiwan uses to crowdsource its laws. *MIT Technology Review*.

Lazar, S. (2017). Deontological decision theory and agent-centered options. *Ethics*, 127(3), 579-609.

Lazar, S., & Graham, P. A. (2021). Deontological decision theory and lesser-evil options. *Synthese*, 198(7), 6889-6916.

Lee, M. K., Kusbit, D., Kahng, A., Kim, J. T., Yuan, X., Chan, A., ... & Procaccia, A. D. (2019). WeBuildAI: Participatory framework for algorithmic governance. *Proceedings of the ACM on human-computer interaction*, 3(CSCW), 1-35.

Noothigattu, R., Gaikwad, S., Awad, E., Dsouza, S., Rahwan, I., Ravikumar, P., & Procaccia, A. (2018, April). A voting-based system for ethical decision making. In *Proceedings of the AAAI Conference on Artificial Intelligence* (Vol. 32, No. 1).

Mitton, C., Smith, N., Peacock, S., Evoy, B., & Abelson, J. (2009). Public participation in health care priority setting: a scoping review. *Health policy*, 91(3), 219-228.

OECD (2022), OECD Guidelines for Citizen Participation Processes, OECD Public Governance Reviews, OECD Publishing, Paris, <https://doi.org/10.1787/f765caf6-en>.

Sorensen, T., Mishra, P., Patel, R., Tessler, M.H., Bakker, M.A., Evans, G., Gabriel, I., Goodman, N., & Rieser, V. (2025). Value Profiles for Encoding Human Variation. <https://arxiv.org/abs/2503.15484>

Appendix A: Worldview scores

Weight given to various aspects of a normative dimension, with 1 reflecting full weight.

Egalitarianism vs. Prioritarianism

Worldview	Bottom 1%	Next 9%	Next 40%	Next 50%
Social engineer	1	1	1	1
Rule of rescue	1	0.4	0.1	0.1
Capabilities	1	1	0.6	0.4
Utilitarian	1	1	1	1
Libertarian	1	1	1	1
Redress	1	0.9	0.3	0.1

Time of effect

Worldview	Present	Future
Social engineer	0.25	1
Rule of rescue	1	0
Capabilities	1	1
Utilitarian	1	1
Libertarian	1	0.25
Redress	1	0.5

Change type

Worldview	Incremental	Systemic
Social engineer	0.1	1
Rule of rescue	0.9	0.1
Capabilities	0.7	1
Utilitarian	1	1
Libertarian	1	0.5
Redress	0.5	1

Age of recipient

Worldview	Children	Adults	Elderly
Social engineer	0.42	1	0.13
Rule of rescue	1	0.4	0.6
Capabilities	1	0.9	0.1
Utilitarian	0.42	1	0.13
Libertarian	0.18	1	0.18
Redress	0.42	1	0.13

Saving vs. improving lives

Worldview	Save	Improve
Social engineer	1	1
Rule of rescue	1	0.1
Capabilities	0.7	1
Utilitarian	1	1
Libertarian	0.7	1
Redress	1	1

Autonomy vs. general wellbeing

Worldview	Autonomy	General Wellbeing
Social engineer	0.2	1
Rule of rescue	0	1
Capabilities	0.9	0.7
Utilitarian	0	1
Libertarian	1	0.2
Redress	0.8	0.6

Appendix B: Project scores

Proportions of project's benefits that accrue to each normative aspect. Scores add to 1.

Egalitarianism vs. Prioritarianism

Project	Bottom 1%	Next 9%	Next 40%	Next 50%
Air filtration	0.01	0.09	0.4	0.5
Mask mandates	0.2	0.3	0.3	0.2
Antivirals	0.5	0.3	0.1	0
Vaccine research	0.01	0.09	0.4	0.5
Vaccine outreach	0.1	0.2	0.3	0.4
Maternal education	0.005	0.045	0.42	0.53
Emergency OB	0.3	0.4	0.15	0.15
Rural attendants	0.1	0.2	0.3	0.4
No child marriage	0.4	0.4	0.1	0.1
Girls in school	0.2	0.3	0.4	0.1
Health infrastructure	0.01	0.09	0.4	0.5
Malaria adult	0.6	0.3	0.1	0
Malaria child	0.6	0.3	0.1	0
Vaccine Research	0.01	0.09	0.4	0.5
Direct payments	0.01	0.09	0.4	0.5

Time of effect

Project	Present	Future
Air filtration	0.5	0.5
Mask mandates	1	0
Antivirals	1	0
Vaccine research	0.1	0.9
Vaccine outreach	1	0
Maternal education	0.8	0.2

Emergency OB	0.9	0.1
Rural attendants	0.9	0.1
No child marriage	0.2	0.8
Girls in school	0.3	0.7
Health infrastructure	0.3	0.7
Malaria adult	0.9	0.1
Malaria child	0.9	0.1
Vaccine Research	0.2	0.8
Direct payments	0.6	0.4

Change type

Project	Incremental	Systemic
Air filtration	0.2	0.8
Mask mandates	1	0
Antivirals	1	0
Vaccine research	0.5	0.5
Vaccine outreach	1	0
Maternal education	0.5	0.5
Emergency OB	0.9	0.1
Rural attendants	0.9	0.1
No child marriage	0	1
Girls in school	0.1	0.9
Health infrastructure	0.1	0.9
Malaria adult	1	0
Malaria child	0.8	0.2
Vaccine Research	0.5	0.5
Direct payments	0.8	0.2

Age of recipient

Project	Children	Adults	Elderly
Air filtration	0.8	0.15	0.05
Mask mandates	0.1	0.4	0.5
Antivirals	0.1	0.2	0.7
Vaccine research	0.3	0.6	0.1
Vaccine outreach	0.3	0.6	0.1
Maternal education	0.2	0.8	0
Emergency OB	0.5	0.5	0
Rural attendants	0.5	0.5	0
No child marriage	0.9	0.1	0
Girls in school	0.7	0.3	0
Health infrastructure	0.3	0.6	0.1
Malaria adult	0.1	0.8	0.1
Malaria child	0.9	0.1	0
Vaccine Research	0.3	0.6	0.1
Direct payments	0.3	0.6	0.1
Baseline	0.3	0.6	0.1

Saving vs. improving lives

Project	Save	Improve
Air filtration	0.2	0.8
Mask mandates	0.7	0.3
Antivirals	0.9	0.1
Vaccine research	0.5	0.5
Vaccine outreach	0.5	0.5
Maternal education	0.1	0.9
Emergency OB	0.9	0.1
Rural attendants	0.9	0.1

No child marriage	0.1	0.9
Girls in school	0.1	0.9
Health infrastructure	0.5	0.5
Malaria adult	0.7	0.3
Malaria child	0.8	0.2
Vaccine research	0.7	0.3
Direct payments	0.2	0.8

Autonomy vs. general wellbeing

Project	Autonomy	General Wellbeing
Air filtration	0.5	0.5
Mask mandates	-0.2	1.2
Antivirals	0	1
Vaccine research	0.1	0.9
Vaccine outreach	0.2	0.8
Maternal education	0.7	0.3
Emergency OB	0	1
Rural attendants	0.1	0.9
No child marriage	0.8	0.2
Girls in school	0.9	0.1
Health infrastructure	0.1	0.9
Malaria adult	0.2	0.8
Malaria child	0	1
Vaccine Research	0.1	0.9
Direct payments	0.9	0.1