

Fair navigation planning: a humanitarian robot use case

Martim Brandão
martim.brandao@kcl.ac.uk
King's College London
London, UK

ABSTRACT

In this paper we investigate potential issues of fairness related to the motion of mobile robots. We focus on the particular use case of humanitarian mapping and disaster response. We start by showing that there is a fairness dimension to robot navigation, and use a walkthrough example to bring out design choices and issues that arise during the development of a fair system. We discuss indirect discrimination, fairness-efficiency trade-offs, the existence of counter-productive fairness definitions, privacy and other issues. Finally, we conclude with a discussion of the potential of our methodology as a concrete responsible innovation tool for eliciting ethical issues in the design of autonomous systems.

CCS CONCEPTS

• **Computing methodologies** → **Robotic planning**; • **Social and professional topics** → **User characteristics**.

KEYWORDS

humanitarian robots, robot navigation, algorithmic fairness

1 INTRODUCTION

In recent years there has been a proliferation of research concerned with the ethics of autonomous systems and artificial intelligence [7–9]. For example, recent studies have shown that machine learning algorithms in recidivism prediction [1, 5], gender classification [4], and other applications [2] can perform better for some people compared to others. This concern has led to greater pressure on developers to innovate responsibly, as well as to the development of a great number of guidelines and principles for ethical development [13]. Such guidelines and principles are helpful in providing a framework for researchers and practitioners. However, they are limited in terms of supporting to implement and satisfy the principles in practice. For example, it is often not clear how “fairness” or “beneficence” principles are relevant to a new technology in practice.

Examples of social inequalities produced by seemingly fairness-unrelated decision-making are numerous and extend well past recent machine learning developments. In the field of “environmental justice” [12], for example, researchers have argued that the implementation of some transportation policies, which could supposedly improve mobility and access to jobs, can indirectly reinforce inequalities of opportunities [7]. Other work has shown that waste

management sites are often concentrated on low-income, and high racial-minority-percentage locations [12]. Often such policies do not overtly target such populations, and inequalities of access or exposure to harm can happen because people are not uniformly distributed across space and are in fact usually distributed in ways that relate to economic, cultural and racial factors [7, 12]. Discrimination is also often embedded within housing markets and the organization of institutions [12], which can implicitly influence decision-making and decision outcomes.

These particular discussions of spatially-organized inequalities also relate strongly to mobile robotics. Similar issues of environmental discrimination can be found, as we will see, in reconnaissance robots, search-and-rescue robots and humanitarian robots, as they could provide different benefits for high- and low-populated areas, gentrified and young areas, etc.

In this paper, which summarizes a previous publication [3], we investigate the concept of fairness in a humanitarian mapping robot use case. We use this example to make several claims about fair design in robot navigation, such as:

- (1) Robot navigation paths can give rise to concerns of distributive fairness and indirect-discrimination;
- (2) Fairness-aware navigation planners will involve efficiency-fairness trade-offs;
- (3) Enforcing certain formalizations of fairness in path planning can be counter-productive;
- (4) Design should be an iterative process of understanding the fairness issues of the context at hand;
- (5) Mathematical formalization and technology-deployment simulations can be a good tool to draw out fairness-related design decisions with stakeholders in the early stages of design.

2 FAIRNESS IN ROBOT NAVIGATION

2.1 Walkthrough mapping-robot example

Imagine a robot that is deployed in the aftermath of a humanitarian disaster in order to find victims that need to receive support, and communicate their location to a response team. This could be a drone searching for earthquake victims. Let us call this a “rescue drone”. The drone departs from a base station and needs to go back to the same station for re-charging batteries after a maximum distance is covered. The drone does several of these trips, although we focus on a single trip in isolation.

Let us say this happens in a specific location—the city of Oxford, UK—and we use census data to guide the robot towards high population-density areas to increase efficiency. Population density is not uniformly distributed in space. Furthermore, the region has spatial biases related to age, ethnicity and gender, as shown in Figure 1. Like other cities [7], Oxford has neighborhoods of higher concentration of minority ethnicities and of older populations than

Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for third-party components of this work must be honored. For all other uses, contact the owner/author(s).

KDD 2020 Workshop on Humanitarian Mapping, August 24, 2020, San Diego, CA

© 2020 Copyright held by the owner/author(s).

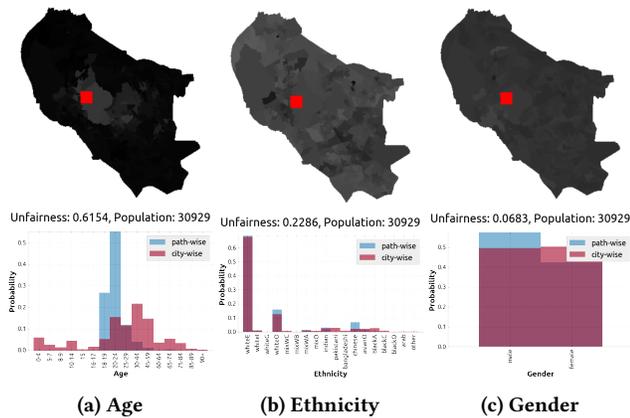


Figure 1: Distribution of age, ethnicity and gender over the whole city and a robot path around the city center.

the city center. The consequence for our rescue robot example is the following: planned paths will have skewed distributions of these personal characteristics. If for example the drone thoroughly explores the area immediately around the base station it will find many people because of population density, however most of whom are young and healthy.

Let us suppose that the navigation planner is such that it thoroughly explores the region around the base-station because it is highly populated. Figure 1 shows this path on top of the city map, as well as the path-wise and city-wise personal characteristic distributions of age, ethnicity and gender. Figure 1 shows that, as qualitatively seen on the maps, the distribution of age in the center (along the robot’s path) is highly biased towards that of undergraduate students, while the city-wide distribution is considerably more uniform. The figure also shows that both the center and the city as a whole are highly biased towards a “white English” ethnicity. The center has an overrepresented “white other” and “Chinese” population compared to the rest of the city.

2.2 Issues raised by this example

2.2.1 Indirect discrimination within robot navigation. The example shows that robot navigation paths can be biased in favor or disfavor of different people, as paths inherit spatial distribution biases. In particular, the probability of being found by the robot was strongly correlated with age and ethnicity.

In disaster response, such a robot could continue or even reinforce common criticisms in disaster response missions: that policies for selecting disaster response locations usually benefit particular groups of people [11]. Avoiding such discrimination explicitly through algorithms could be a way not only to promote distributive justice, but also to enforce a certain degree of political or commercial neutrality in disaster response (i.e. to make sure that disaster response agencies using robots do not favor a particular group).

2.2.2 Inequality can be unfair. While one of the goals of a humanitarian rescue robot is to find as many people as possible, notions of priority also exist [11]. One accepted principle is to attend to the people most-at-risk first [10, 11]. These could be people with

the lowest health-state, those living in low-quality accommodation susceptible of collapse, those that have lower chances of survival if not attended to, such as children and older adults, etc. In our particular example, the fact that there is indirect discrimination of age, with a bias towards the younger population, is unfair according to such view of disaster response ethics. So while our robot example is doing part of a disaster response team’s job—finding as many people as possible—it is not respecting the context’s notion of distributive fairness.

This example also raises the issue of identifying in which personal characteristics indirect discrimination is unfair. What about the case where a rescue robot finds many people, primarily those at highest risk, but at the same time is biased towards the white population of the city—since it is slightly closer to the base station than the neighborhoods of high minority-concentration? Should this give rise to concern? According to defendants of affirmative action it should, since it reinforces social inequalities that have been ingrained in society (through urban policy, housing markets, etc.) for generations.

It then becomes important to provide stakeholders with tools to help identify these possible inequality outcomes of robot deployment so that better design decisions can be made before robot deployment. One such tool is the methodology that we use in this paper—of simulating the robot’s deployment and predicting possible inequalities that can give rise to concerns of distributive fairness. The example further raises the issue, however, of how one will reach the “final” choice of fairness principles and protected characteristics to use in an algorithm. This requires involvement of all stakeholders in the decision process.

2.2.3 Robots must face dilemmas that humans already face. Thinking about the fairness issues related to disaster response locations is not something new about rescue robots, but an inherent concern of disaster response itself. For example, discussions and claims of unfairness are raised regarding disaster response hospital locations when they favor people of specific backgrounds, sometimes politically or commercially favorable to the country backing the response [11]. Principles such as most-at-risk first [10, 11] are already used in the field by human doctors. Thus, the rescue robot example also shows that when robots are used to solve problems currently solved by humans, they must face similar dilemmas currently faced by human decision-makers.

3 DESIGNING FAIR ROBOT NAVIGATION

3.1 Objects of fairness in navigation

We now briefly formalize fairness in the context of robot navigation. We distinguish between different objects of fairness in navigation: **Locations:** Being fair to locations (e.g. neighborhoods) means that we care about which locations will be visited by a robot: in particular, how often they will be visited, either in absolute value or in comparison to other locations.

Protected characteristics: Being fair to protected characteristics means that we care about the distribution of protected characteristics of the people found along the robot’s path.

3.2 Specifications of fairness in navigation

In [3] we provide formal specifications for a set of distributive justice principles in the context of navigation. These are, for example:

Demographic parity: Enforcing (or minimizing the distance to) demographic parity. In the context of navigation this implies that the event of a person being found along a robot’s path is independent from group membership (i.e. a protected characteristic).

Rawlsian egalitarian fairness: maximizing the utility of the worst-off, i.e. maximizing the visit counts of the least-visited region, or the probability of being found for the least-likely group.

Affirmative action: Enforcing a desired distribution of location-visits (e.g. region of priority 1 visited M times more than region of priority 2) or protected characteristics (e.g. ratio of younger and older people found along the planned path). This definition is general enough to encompass demographic parity—which is equivalent to a preference towards the distribution of the whole population.

3.3 Developing a fair navigation planner

In [3] we reach several conclusions regarding the issues involved in the design of a fair system. Here is a brief summary:

3.3.1 Fairness may be infeasible, requires trade-offs. We found that to decrease unfairness it is necessary to reduce the total population found. We arrived at this conclusion by computing the Pareto-front of two objectives: minimization of unfairness (Jensen-Shannon distance to perfect demographic parity for the sake of example) and maximization of efficiency (number of people found). Figure 2 shows this Pareto-front. Each point along the curve is a different path of different unfairness and efficiency values. Figure 3 shows the paths corresponding to the two extremes of the Pareto-front (i.e. lowest and highest unfairness). In our example, lowering unfairness requires lowering efficiency because older people also live more scattered and further away from the center than the younger population. The figures also show that it was impossible for the method to find a path of strict fairness, i.e. where demographic parity is satisfied exactly, since the curve does not reach zero. This is understandable because, firstly, extreme luck should exist for a subsample of the city to exhibit exactly the same statistics as the city as a whole. Secondly, Pareto-front estimation methods such as the one we used are not guaranteed to find global minima since that would require exhaustively searching all possible paths within the map, which is unfeasible for the size of our problem.

In this situation, a stakeholder such as a decision-maker, emergency responder or policy maker could use the Pareto-fronts themselves to make a more informed decision about efficiency and fairness of the response, in a way that reflects the priorities and values at play in the specific situation. The decision-maker could select one of the solutions within the Pareto and thus explicitly weigh the conflicting objectives according to context.

3.3.2 Current planning methods provide few guarantees. The method we used to generate these paths is not a traditional one in robotics: it was based on evolutionary algorithms to compute Pareto-fronts [3], which are algorithms that do not offer optimality guarantees. Traditional methods (e.g. A^* with an admissible heuristic) could also have been used but they would require cost functions to be Markovian, i.e. the cost over a path to be equal to the sum of per-state

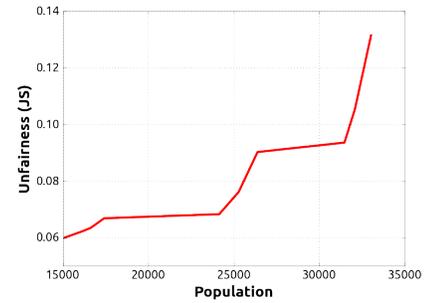


Figure 2: Pareto-curve showing the trade-off between the total population found along the robot’s path and the distance to the desired distribution of age.

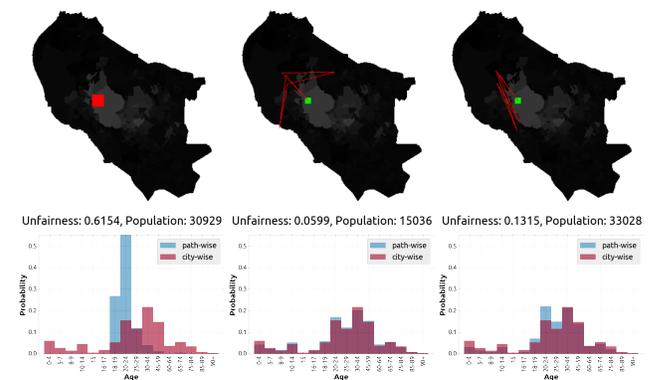


Figure 3: A thorough search path around the base (left) and the two extreme Pareto-optimal paths (middle and right).

costs. This is not the case for our protected characteristic fairness specification, and so a cumulative-cost approximation would have to be used. In [3] we show that optimizing a cumulative-cost approximation would lead to twice higher unfairness in this case. So current methods provide few guarantees: they will either try to optimize the desired fairness metric without optimality guarantees, or they will optimize (with guarantees) a proxy metric that is different from what we care about.

3.3.3 Fairness specifications can be counter-productive. We found that some fairness specifications can lead to lower utility for all groups, compared to demographic parity. For example, we compared the result of promoting demographic parity on gender, with the result of promoting strict-equality affirmative action (i.e. a 50-50% ratio). We found out that strict-equality can lead to lower utility for both male and female classes. In other words, all classes were made worse-off just in order to find higher equality solutions.

Again, the use of fairness-related visualizations such as distributions and trade-offs is important to inform decision-makers about the impact and effectiveness of the technical choice. Furthermore, this example shows the advantage of simulating system deployments implementing different fairness definitions, in order to better evaluate and predict their results. Finally, the example suggests that

the process of specification—of selecting a definition of fairness—could be the product of an iterative design and validation process.

3.3.4 Intuitive understanding. The particular metric we have used for “unfairness” in Pareto-curves—distance to the fair distribution—also raises questions about which metrics are most intuitive for stakeholders to evaluate the degree to which a distributive principle is satisfied. In this example we used Jensen-Shannon distance between distributions, but such a metric could arguably be considered unintuitive. The intuitiveness and interpretability of the choice of metric is an important topic of further research.

3.3.5 Design is an iterative process. A further issue that the example raises is that there are possibly multiple personal characteristics that a user or decision-maker cares about and would like to pay respect to. These may not be obvious from the onset of robot deployment. For example, a disaster response team might program a fair navigation planner to respect a certain health and age-related feature, only to later find out they have a bias towards high-income neighborhoods that they would like to avoid. Alternatively, optimizing for fairness in one characteristic may introduce new biases in the paths that are again morally relevant.

Part of the specification process will hence be in the discovery of which values matter for the application at hand, and how. This process can only succeed if there exist value-sensitive tools [6] that identify or support the identification of these issues. Our methodology in [3] was to develop a resource of applicable formal definitions that can be simulated to guide discussions and anticipate issues *before* deployment, at the early stages of development.

4 DISCUSSION

4.1 Developing fair navigation planners

As we have shown, building “fair navigation planning” algorithms certainly requires more than optimizing a fairness metric. We believe the ethical development and deployment of fair navigation algorithms requires work on multiple fronts.

4.1.1 Transparency. Planners should provide an intuitive understanding, through appropriate visualizations, statistics, or metrics, of the fairness characteristics of navigation plans. Planners should be equipped with data and tools for the analysis of fairness across multiple variables and specifications, in order to allow impact and fairness-related issues to be spotted by stakeholders.

4.1.2 Human autonomy. In order to allow humans to responsibly use such tools, they need to be able to understand and control the trade-offs between fairness and other task objectives. This will require the use of visualization and human-in-the-loop design features. For example, Pareto-curves of the different objectives can serve as interesting visualizations and decision-aids. Additionally, users should be able to interface with the planning methods in order to include considerations of relevant expert knowledge, such as adding intermediate goals, biasing paths to certain solutions, or adjusting estimates of risk and utility (e.g. building damage or population density in the rescue case).

4.1.3 Privacy. Promoting distributive fairness over protected characteristics in navigation requires data on the distribution of these

features themselves. This comes at the cost of having to gather such data, but also of potential privacy issues within the collection, analysis or security breaches of the data.

4.1.4 Data availability and input. Fairness with respect to locations cannot be fulfilled if there are missing locations in a map, and similarly for fairness on protected characteristics. For fair navigation planning systems to be reliable, data and models of the relevant fairness features will be required. For example, population statistics over a map should be available to the planning algorithms. Some of these already exist—census data can be detailed in some countries—though they could arguably only be available in privileged circumstances. This calls for implementing ways for operators and stakeholders to populate maps with demographics from unstructured data and human knowledge. Finally, even if demographic statistics are difficult to obtain in many humanitarian situations, implementing fairness at the level of locations may still be desirable.

4.2 Methodological contribution

Rather than only reflecting on general principles of fairness or requesting stakeholders to anticipate the consequences of an innovation, we believe that formal models and simulation-based investigation provide a more solid foundation on which to initiate discussions that can anticipate the consequences of an innovation. So, in the case of robot navigation it is possible to provide examples of particular decisions made in the design of an algorithm and their consequences. This can be a resource in stakeholder workshops where potential users, developers, policy makers and members of the general public seek to anticipate the implications of a technological innovation. Modelling, formal specification and simulation can help provide a more systematic and informed foundation to such discussions, prior to any development taking place.

5 CONCLUSION

In this paper we explored the concept of fairness in the seemingly mundane, value-neutral, technical problem of robot navigation. We showed that there is a fairness dimension to robot navigation, using a walkthrough example of a humanitarian rescue robot. We discussed how mobile robots will have to face similar dilemmas that humans already face. We then applied theories of distributive justice to our navigation problem and used them to simulate robot deployment outcomes, ground discussions of fairness and design across multiple stakeholders, and anticipate issues. We showed that fairness-aware navigation planners will involve efficiency-fairness trade-offs, that their design should be an iterative process of understanding the fairness issues of the context at hand, and that current planning methods have downsides that need be addressed.

This paper also sets the ground for a new research field of fair planning. Several challenges still lie ahead. Part of those are technical challenges of designing efficient, interpretable, formally verifiable and optimal methods. Another part is related to responsible innovation and value-sensitive design through appropriate analysis, visualization and participation tools.

REFERENCES

- [1] Julia Angwin, Jeff Larson, Surya Mattu, and Lauren Kirchner. 2016. Machine Bias: there’s software used across the country to predict future criminals. And

- it's biased against blacks. ProPublica 2016.
- [2] Martim Brandao. 2019. Age and gender bias in pedestrian detection algorithms. In *Workshop on Fairness Accountability Transparency and Ethics in Computer Vision, CVPR*. <https://arxiv.org/abs/1906.10490>
- [3] Martim Brandao, Marina Jirotko, Helena Webb, and Paul Luff. 2020. Fair navigation planning: a resource for characterizing and designing fairness in mobile robots. *Artificial Intelligence* 282 (2020). <https://doi.org/10.1016/j.artint.2020.103259>
- [4] Joy Buolamwini and Timnit Gebru. 2018. Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification. In *Proceedings of the 1st Conference on Fairness, Accountability and Transparency (Proceedings of Machine Learning Research, Vol. 81)*, Sorelle A. Friedler and Christo Wilson (Eds.). PMLR, New York, NY, USA, 77–91.
- [5] Alexandra Chouldechova. 2017. Fair Prediction with Disparate Impact: A Study of Bias in Recidivism Prediction Instruments. *Big Data* 5, 2 (2017), 153–163. <https://doi.org/10.1089/big.2016.0047>
- [6] Batya Friedman, Peter H Kahn, Alan Borning, and Alina Hultgren. 2013. Value sensitive design and information systems. In *Early engagement and new technologies: Opening up the laboratory*. Springer, 55–95.
- [7] Aaron Golub, Richard A Marcantonio, and Thomas W Sanchez. 2013. Race, space, and struggles for mobility: Transportation impacts on African Americans in Oakland and the East Bay. *Urban Geography* 34, 5 (2013), 699–728.
- [8] Andre Holzapfel, Bob Sturm, and Mark Coeckelbergh. 2018. Ethical dimensions of music information retrieval technology. *Transactions of the International Society for Music Information Retrieval* (2018).
- [9] Patrick Lin. 2015. *Why Ethics Matters for Autonomous Cars*. Springer Berlin Heidelberg, Berlin, Heidelberg, 69–85. https://doi.org/10.1007/978-3-662-45854-9_4
- [10] Ofer Merin, Nachman Ash, Gad Levy, Mitchell J Schwaber, and Yitshak Kreiss. 2010. The Israeli field hospital in Haiti-ethical dilemmas in early disaster response. *New England Journal of Medicine* 362, 11 (2010), e38.
- [11] Dónal P O'Mathúna, Bert Gordijn, and Mike Clarke. 2013. *Disaster bioethics: Normative issues when nothing is normal*. Vol. 2. Springer Science & Business Media.
- [12] Gordon Walker. 2012. *Environmental justice: concepts, evidence and politics*. Routledge.
- [13] Alan F. T. Winfield and Marina Jirotko. 2018. Ethical governance is essential to building trust in robotics and artificial intelligence systems. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences* 376, 2133 (2018), 20180085. <https://doi.org/10.1098/rsta.2018.0085>