# An ethical decision-making framework with serious gaming: a smart water case study on flooding

Gregory Ewing and Ibrahim Demir **IWA**

## ABSTRACT

Sensors and control technologies are being deployed at unprecedented levels in both urban and rural water environments. Because sensor networks and control allow for higher-resolution monitoring and decision making in both time and space, greater discretization of control will allow for an unprecedented precision of impacts, both positive and negative. Likewise, humans will continue to cede direct decision-making powers to decision-support technologies, e.g. data algorithms. Systems will have ever-greater potential to effect human lives, and yet, humans will be distanced from decisions. Combined these trends challenge water resources management decision-support tools to incorporate the concepts of ethical and normative expectations. Toward this aim, we propose the Water Ethics Web Engine $(WE)^2$, an integrated and generalized web framework to incorporate voting-based ethical and normative preferences into water resources decision support. We demonstrate this framework with a 'proof-of-concept' use case where decision models are learned and deployed to respond to flooding scenarios. Findings indicate that the framework can capture group 'wisdom' within learned models to use in decision making. The methodology and 'proof-of-concept' system presented here are a step toward building a framework to engage people with algorithmic decision making in cases where ethical preferences are considered. We share our framework and its cyber components openly with the research community.

**Key words** | decision support, ethics, flooding, human-centered AI, smart water systems

**Gregory Ewing** (corresponding author)
**Ibrahim Demir** **IWA**
Department of Civil and Environmental Engineering,
University of Iowa,
300 S. Riverside Dr., Iowa City, IA 52245,
USA
E-mail: *gregory-ewing@uiowa.edu*

## HIGHLIGHTS

- Water Ethics Web Engine $(WE)^2$ is a web framework to incorporate voting-based ethical preferences into water resources decision support.
- We share $(WE)^2$ openly at our project repository: https://github.com/uihilab/waterethicswebengine.
- A proof-of-concept use case is presented where decision models are learned and deployed with flooding scenarios.
- Results indicate that the framework can capture group 'wisdom' in learned models.

## INTRODUCTION

Sensor networks, which is built on the back of the latest digital communication technologies, are increasingly being

deployed in urban sewer networks and at the regional scales to monitor flooding and water quality of rivers (Habibi *et al.* 2017; Mullapudi *et al.* 2017; Jones *et al.* 2018; Yildirim & Demir 2019). Concurrently, control technologies are being deployed alongside sensors which allow for operators to actively manipulate these systems in places and in a

way that was previously inconceivable (Kerkez *et al.* 2016; Mullapudi *et al.* 2018). Furthermore, the continued adoption of control technologies also introduces new stakeholders by which our water environment is *actively* manipulated (Rentsch 2019). These new technologies allow for higher-resolution monitoring via novel sensors (Sermet *et al.* 2019), social media (Sit *et al.* 2019), and decision making (Demir *et al.* 2018; Sermet *et al.* 2020) both in time and space. Where historically our water infrastructure, such as stormwater ponds, were designed as passive structures there is now the capability to control releases *actively* and *precisely* from a growing, distributed amount of water infrastructure, oftentimes as small as stormwater ponds.

A proliferation of new sensing and control locations means considerably more complex systems to address persistent water resources challenges such as flooding and water quality. Yet, to do so requires harnessing the unprecedented complexity of these new systems. Toward this end, there is a growing body of research on control schemes for water systems (Overloop 2006; Mollerup *et al.* 2017; Schütze *et al.* 2018). Recent studies mainly focus on the challenges of controller design to meet the objectives of physically based water *quantity* (flooding; Sadler *et al.* 2019; Mullapudi *et al.* 2020; Sun *et al.* 2020) and sometimes *quality* (total suspended solids or different 'indicator' pollutants; Muschalla *et al.* 2014; Sharior *et al.* 2019; Troutman *et al.* 2020). The collective work demonstrates the potential for improved system performance with a coordinated and increasingly automated control approach and the promise of the new 'smart' water paradigm.

However, little work exists to determine whether control schemes with physically based objectives are consistent with socially normative expectations of 'right' and 'wrong' actions. Because the primary objective of these control studies is to investigate the performance of control strategies, many experimental designs use toy networks or simplified abstractions of real networks. One consequence of this experimental design is that the social and economic variations within and across the network are not considered in the development of control schemes, nor in determining the impact of the controller performance. Yet, when considering the deployment of these distributed technologies across an entire city or region, these demographic factors become relevant. Consider the

case, for example, where some flooding is an inevitable outcome within the catchment area of a stormwater network with dynamic control capabilities; how *ought* a controller act to consider the societal implications of such flooding? Or, if a control scheme consistently recommends distributing damages in poor neighborhoods and benefits in richer neighborhoods, *ought* its directives to be followed? These normative questions of *ought* – across populations, landscapes, communities, etc. – pose serious ethical and moral dilemmas, especially where negative impacts are unavoidable and uneven.

While the ethical questions of civil infrastructure are as old as the infrastructure itself, the challenge to incorporate moral and ethical preferences into the automated, algorithmic decision-making tools that water resources management will rely upon are, indeed, new. The novelty of this challenge is supported by two trends. First, because sensor networks and control systems allow for higher-resolution monitoring and decision making in both time and space, it follows that greater discretization of control will allow for an unprecedented precision of impacts, both positive and negative. A second and complimentary trend is as the complexity of systems grows, humans will continue to cede direct decision-making powers to decision-support technologies such as data algorithms (Sermet & Demir 2018, 2019). Systems will have ever-greater potential to effect human lives, and yet, humans will be insulated from these direct decisions. An important topic to explore is whether decision-support tools for water resources management can integrate socially normative expectations with physically based objectives.

## Individual to institutional: ethics, morality, and machines

The challenge to consider social concepts into what are seemingly technical problems is non-trivial. The social and the technical are intertwined to make a sociotechnical problem (Jonoski 2002; Vojinović & Abbott 2012). This understanding motivates our review of both social and technical work as it relates to our efforts. Generally, we can consider ethics and morality along a spectrum that spans from the individual to the institutional. At the individual level, 'normative' ethics considers theories and schema to determine the appropriate actions of an individual or an agent. At the institutional

level, studies of ethics and morals consider the power structures that shape norms, their compliance, and their impact on individuals and groups. Work in biology, philosophy, psychology, and sociology investigates questions along with this spectrum. The field of machine ethics complements the efforts from the natural and social sciences by developing methods to incorporate decision making into artificial technologies that are consistent with human's (society's) normative expectations of their behavior (Moor 2006). Here, we provide a brief background on these fields and how they relate to incorporating normative ethics into smart water systems. Though some disciplines make distinctions between ethics and morality, we use them as interchangeable in the context of this effort. For consistency within various disciplines, we defer to their respective terminology when referencing their work.

## Normative ethical theories in philosophy

In the study of ethics, or moral philosophy, there are three fundamental traditions: *deontological*, *consequential*, and *virtue*. Respectively, these camps choose to inspect morality from three overlapping, but not identical, questions: 'What is the right thing to do?', 'How is the best possible state of affairs achieved?', and 'What qualities make for a good person?' (Grayling 1995). Importantly, each of these questions, and the responses of the ethical theories, are posed for *individuals*.

The *deontological approach*, or duty ethics, considers an action to be moral based upon *a set of rules* that deem an action permissible, impermissible, or obligatory (Alexander & Moore 2016). A strength of the deontological approach is the clarity of the rule to *direct* actions. However, multiple rules can require contradicting actions, which is a key weakness of a purely deontological approach.

In contrast, the *consequentialist* tradition judges the correctness of an act on its *outcome* (Sinnott-Armstrong 2019), in that the correct action is the action that leads to the best outcome by some specified *objective function*, such as maximizing happiness or minimizing costs. Egalitarianism and utilitarianism are well-known consequentialist moral theories. A concern for consequential approaches is which factors should be included in determining the normative value of an outcome. Though both consequential in the

approach, utilitarian and egalitarian models can produce incompatible solutions due to which factors are given importance. This tension is referred to as the equality-efficiency or the fairness-efficiency dilemma (Binmore 1998).

In the *virtue ethics* tradition, virtues are the fundamental, irreducible unit by which to define normativity, meaning that they are derivative to neither the outcome of actions (consequential) nor duty to perform an action (deontological) (Hursthouse & Pettigrove 2018). In agent-based virtue ethics, agents' *motivations* ascribe the rightness and wrongness of an act and agents learn virtue from 'exemplars of goodness' (Zagzebski 2004). This understanding buoys the concept of learning models of normative behavior by observing the human performance of similar tasks.

## Morality in the sciences

Evolutionary biology tells us that morals are adaptations to social living; when prehistoric humans began to form larger groups, survival was dependent upon the group, and therefore, what was best for it could supersede the priorities of the individual (Krebs 2008). Psychology and the disciplines of biology are interested in the mechanisms and processes of the human brain as they relate to developing and making moral judgments (Haidt 2007). As a distinction from other fields, sociology interrogates morality at scales beyond the individual. When sociology does focus on the morality of individuals, it is almost always in relation to a larger group. Contemporary sociology of morality includes both moral theorizing and experimental science to uncover moral truths (Bykov 2019). Work is not a shared substantive focus, but the recognition that moral evaluations and categorizations are an essential part of struggles in 'social fields' (Hitlin & Vaisey 2013).

When considering how to incorporate moral and normative sentiments into intelligent infrastructure, the sociological literature provides intriguing insights. The theory of 'thick' and 'thin' moral concepts contends that there are thin moral concepts that are 'methodologically tractable', through hypothetical situational tests. Conversely, thick concepts – like dignity, integrity, humanness, etc. – do not lend themselves to parsimonious description or measurement, making thick concepts more difficult to explore by the experiment (Abend 2011). Furthermore,

there is a lack of evidence for how thin and thick concepts relate to each other. Thus, a holistic approach would incorporate exercises in which feedback on both thin and thick moral concepts can be collected. Furthermore, other studies show that higher social class predicts increased unethical behavior (Piff *et al.* 2012), that lower-class individuals are more likely to be compassionate to another's suffering (Stellar *et al.* 2012), that different social classes use different criteria in anticipated cost–benefit analyses (Trautmann *et al.* 2013), and that lower-class individuals are more likely to perform an unethical act if the act was to the benefit of others (Dubois *et al.* 2015). Together, these findings clarify the process and partners needed when operationalizing feedback on normative expectations of smart water systems.

## Integrating moral and normative sentiments into intelligent systems

Machine learning (ML), artificial intelligence (AI), and data algorithms generally have been – and stand to be – applied in a wide range of scenarios from data augmentation (Demiray *et al.* 2021) to forecasting (Sit & Demir 2019; Xiang *et al.* 2020). There is, however, growing acknowledgement that, unexamined, these techniques can institutionalize the biases and structural prejudices that persist in the world, especially in decision-making contexts that directly affect humans such as in job hiring, evaluating credit scores, and predicting repeat offenders during parole processes (O'Neil 2016). Consequently, popular culture, industry, government, and academia have focused attention upon the intentional and ethical application of AI and created a proliferation of AI ethics frameworks (Hagendorff 2019; Jobin *et al.* 2019). Two primary concerns of deploying these technologies in a complex world are as follows: (1) their instantiated purpose and behavior may not be well understood and (2) the systems may take irrevocable acts before humans have the data to discern their error (Samuel 1960).

These concerns connect to the questions of building frameworks and governance models to responsibly integrate algorithms into systems that affect humans. Toward this aim, the agenda of 'society-in-the-loop' (SITL) proposes to build an *algorithmic social contract* where human-in-the-loop principles and general stakeholder values are integrated into an iterative development process (Rahwan

2017). In its proposal, SITL forwards the tenets of *algorithmic regulation* (O'Reilly 2013). The algorithmic regulation proposed requires a deep understanding of the desired outcome that outcomes are monitored, that the algorithm adjusts based on new data, and periodic deeper analysis on algorithm performance.

Recent work in machine ethics can be understood in the context of building methodologies and frameworks toward SITL and algorithmic regulation (Tolmeijer *et al.* 2020). A crowdsourced, voting approach has been proposed as a flexible method to incorporate moral sentiments for AI applications (Conitzer *et al.* 2015, 2017). Crowdsourced voting has already been tested on a massive scale to query preferences on resolving moral dilemmas of autonomous vehicles using a pairwise comparison experimental setup (Awad *et al.* 2018). After data collection, concepts from computational social choice (Chevaleyre *et al.* 2007) and ML classification techniques can be used together to build preference models of individuals and groups (Noothigattu *et al.* 2017). These preference models, which in theory represent the real sentiments of the participants, can then be used in decision-support algorithms. Two examples of this process are in the development of an algorithm to decide tiebreaks in a theoretical kidney exchange market (Freedman *et al.* 2019) and an algorithm to support a fair and efficient dispatch of food donations (Lee *et al.* 2019). Critically, these examples apply voting-based preference aggregation within a larger, participatory framework. We observe that such frameworks to generally be able to (a) identify relevant belief features to base decisions upon; (b) assay preference along each relevant belief feature using pairwise comparison testing; (c) learn a preference model from the preference assays; (d) use the learned preference model in experimental decision-making scenarios; (e) analyze the outcomes of the preference model and identify if/where model-driven outcomes are incongruous with stated values or objectives; and (f) iterate on and/or deploy the learned preference model. In the 'Materials and methods' section, we provide further details for this process.

Currently, open frameworks which support all or part of the above workflow are limited (if any) in the literature. Previous studies have built custom web-voting applications (step b) for their specific applications but did not share generalized source code. Other studies used proprietary web

platforms to collect preference data. Furthermore, we are unaware of any framework that has shared analytical tools for post-play data processing and preference model development.

Toward this end, we propose an integrated and generalized framework to incorporate voting-based ethical and normative preferences into water resources decision-support schemes. We then demonstrate the framework with a proof-of-concept use case where decision models are learned and deployed to respond to flooding scenarios. The results indicate that the framework can capture group 'wisdom' in learned models and use this to make decisions. Furthermore, we share our generalized framework openly with the research community.

The remaining sections are organized as follows. Section 'Materials and methods' provides details on the methodology of each step of the framework listed above, and cyber components used in handling the step. Section 'Results and discussion' demonstrates the framework with a proof-of-concept use case for decision making for flood response with a discussion of the framework. Finally, the 'Conclusion' presents the larger context of incorporating normative expectations into smart water systems.

## MATERIALS AND METHODS

In this section, we describe the approach and technologies used to employ the methodological framework introduced previously. First, we describe how to identify relevant belief features. Next, we describe the process employed to collect people's preferences. This subsection includes a description of the generalized web framework developed to collect voting-based preferences, its web architecture, game play, and database architecture. Finally, this section details the post-play data analysis to derive preference models, how to use them to make decisions, and the analytics toolbox developed to support this exercise.

### Identifying and assaying relevant belief features

Example methodologies to identify relevant belief features for decision making (step a) can be found in Lee *et al.* (2019) and Freedman *et al.* (2019) and include survey and interview techniques. Relevant belief features are those that someone would use to make a deliberative action. For example, in the case of choosing between two flooding outcomes, relevant belief features to be considered could be public costs, private costs, injuries, deaths, and environmental impact. Once belief features have been established, different scenarios can be created that vary along with the belief features. The scenarios are presented to individuals in pairs. Participants must choose which outcome they prefer from each pair presented to them. Their choices are recorded and used later to learn a preference model.

To collect preference data from participants, we built an integrated web-based serious gaming platform. Serious gaming is used in a variety of fields for training, decision making, and education (Susi *et al.* 2007). In water resources, serious gaming is used to explore the multifaceted challenges such as multi-hazard mitigation (Carson *et al.* 2018). Web-based serious gaming offers easily accessible and user-friendly interfaces with flexible architectures for various skill levels (Xu *et al.* 2020). In many serious gaming applications, the game play offers the user an opportunity to explore real problems and strategies without the consequences of their actions impacting the real world. Instead, play informs a value system that guides behavior and action during a future, real-world event. Play is recognized as an important feature in the development of value systems and morals in humans. Furthermore, actions within the context of play (i.e. games) give rise to different value systems compared with a work context, even when considering the same topic (Bargheer 2018). These value systems result in different moral treatments of the same topic which can result from starting in the context of play or work. Gamification allows for parsimonious yet engaging descriptions of the ethical dilemmas at hand. More engaged users are more likely to play for longer and contribute more to the model development. The following subsections describe the web architecture, game play, and database architect of our serious game approach.

### Water Ethics Web Engine (WE)$^2$ architecture

The Water Ethics Web Engine (WE)$^2$ is an open and integrated framework that allows a rapid deployment of web applications to investigate moral preferences via pairwise

comparisons. The framework is comprised of a PHP-based application engine and use case web template and includes a database architecture on the back end (Figure 1). Full documentation of the framework and source code can be found in GitHub: https://github.com/uihilab/WaterEthics-WebEngine. Researchers provide their application and experimental information to the engine via two configuration files (site_meta.json and scenarios.json) using the JavaScript Object Notation (JSON) format. These files, along with supplied images, are stored on a Web server and accessed during the game play by the application engine. Data generated during the game play are logged in a web-accessible database. Once data are collected, researchers have access to analytics tools and a portal for data export.

Case study information described above is stored on a Web server in a unique directory. When a user navigates their browser to the specific game version, given to them by the researcher, the engine generates the Web page from the content stored in case study files (i.e. the site metadata, scenario content, and the images). Scenarios displayed to participants are chosen randomly from the total set supplied.

## Game play

Users are presented with a homepage that provides a brief mission statement on the purpose of the game and a button to start the game play. During the game play, users are presented with two scenario windows displayed side-by-side. In each scenario windows are the descriptions of



**Figure 1** | System architecture of the web-based decision-making framework. The application engine, database, and researcher-supplied data are located on a Web server. When clients, or users, navigate to a project directory, the application engine builds the project using the information supplied in the static JSON files. After the game play, the researcher may use the post-play analytical tools to investigate results and learn preference models.

the event and an action button with a user-defined decision, e.g. 'Flood This' or 'Save This'. Descriptions come in some combination of three forms: an image, info bar, and written description. All description types are supplied by the researcher and are customizable. Users are instructed to use the descriptions to determine what outcome they prefer for the scenario. To choose the preferred decision, users click on the action button on which the decision is recorded, and a new scenario is displayed. Once the user has provided their preference for all scenarios, they will be guided to a results page that provides the user with a description of their aggregate preferences in relation to all others who have played the game and to the absolute possible outcomes along the belief features provided in the descriptions.

## Post-play analysis

In this section, we describe the methods used in the post-play analysis. This includes learning preference models, making decisions with these models, and the iterative process of improving and deploying them. We also describe our analytic toolkit to facilitate these efforts.

## Learning preference models

By using the paired comparison experimental design, parsimonious random utility models can be leveraged to learn preference models for individuals or groups (step c) (Tsukida & Gupta 2011). When a participant votes that they prefer one outcome over another, one can hypothesize that the outcome has, on average, a greater utility. It is assumed that the utility of a decision relies on weighing the tradeoffs of each option across the belief features that describe them. When participants provide decisions across a set of scenarios, they provide a classified dataset from which a preference model can be learned via classification techniques used in machine learning.

There are numerous methods of using pairwise comparison data to build preference models (Conitzer et al. 2015; Noothigattu et al. 2017; Lee et al. 2019; Freedman et al. 2019). Here, we employ the Thurstone's Law of Comparative Judgement Case V model to convert paired comparisons into group quality scores (Tsukida & Gupta 2011). For a group of participants who have made judgments on the
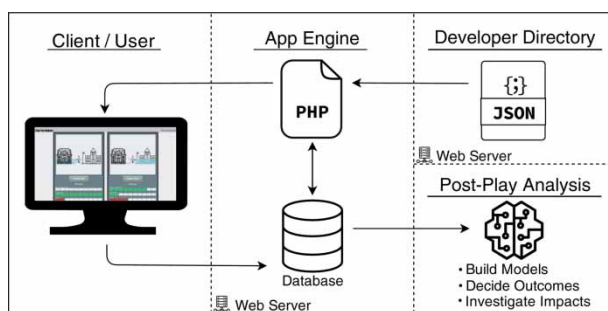
same scenarios, an estimate of the mean quality difference between option A and B, $\mu_{AB}$, is calculated for each scenario as follows:

$$\mu_{AB} = \Phi^{-1}\left(\frac{C_{A,B}}{C_{A,B} + C_{B,A}}\right) \tag{1}$$

where $\Phi^{-1}(x)$ is the inverse cumulative distribution function (CDF) of the standard normal distribution, and $C_{A,B}$, $C_{B,A}$ are the number of votes for each option received. For example, $C_{A,B} = 27$ would mean that 27 people prefer outcome A over outcome B.

Mean differences can be related to scenario belief features as:

$$\mu_{AB} = \beta^T (X_A - X_B) \tag{2}$$

where $X_A$, $X_B$ are the vectors of the belief features for option A and option B, respectively. The learned preference model of the group is the estimate of the belief feature weighting factors of $\beta^T$ and can be found using linear regression, or other classification techniques, given that:

$$\begin{bmatrix} \mu_{AB,1} \\ \mu_{AB,2} \\ \vdots \\ \mu_{AB,n} \end{bmatrix} = \beta^T \begin{bmatrix} (X_A - X_B)_1 \\ (X_A - X_B)_2 \\ \vdots \\ (X_A - X_B)_n \end{bmatrix} \tag{3}$$

where $\mu_{AB,i}$ is the estimated mean utility difference of scenario $i$ of the $n$ shared scenarios and $(X_A - X_B)_i$ is the difference in belief features of $i$ of the $n$ shared scenarios.

After a preference model $\beta^T$ is learned, we can investigate its behavior. To do so, we need to calculate $\mu_{AB}$ in Equation (2) using the belief features of new scenarios. If $\mu_{AB}$ is positive, then choice A is selected. If $\mu_{AB}$ is negative, then choice B is selected. Decisions generated using the preference model can then be compared against various benchmarks, such as the historical performance of a system or against some specified definition of fairness and efficiency. Committing to this activity will require a reflection on whether the algorithm is meeting its purpose. After interrogating the behavior of the learned preference model, the researcher has a clear choice to continue to refine the preference model iteratively or incorporate the preference model into their operation. As a practical matter, one can choose to pursue both continuously and simultaneously.

To facilitate the workflow of data retrieval, learning preference models, and using them to make decisions on new scenarios, we developed a data analytics toolkit using the Python programming language. The toolkit includes a data service interface with a PostgreSQL database, data structures and methods to streamline the workflow. The toolkit and documentation of workflow are provided in the project repository.

## RESULTS AND DISCUSSION

In this section, the web framework and methodology are applied in a proof-of-concept use case with a group of undergraduate engineering students at the University of Iowa ($n = 409$). The use case explores the preferences of individuals to flood scenarios and the result of using their preference models to make decisions.

Between 1980 and 2018, Iowa experienced 26 flood disasters where damages exceeded $1 billion (Immerman & Immerman 2018). Most recently, damages due to the spring 2019 flooding events in Iowa are estimated at $1.6 billion (Hardy & Cannon 2019). Furthermore, between 1988 and 2016, there were a total of 951 flood-related presidential disaster declarations in the state (Eller 2018). In total, the sum of damages over the last 40 years is estimated at $41 billion, or a little more than $1 billion per year. In response, the State of Iowa has supported flood mitigation and flood preparedness through the Iowa Flood Center (IFC) and the Iowa Watershed Approach (Weber *et al.* 2018). Yet still, when floods occur, the response requires lay people and decision makers alike to make fraught decisions that take on ethical and moral dimensions (Bosman *et al.* 2019; Kelley 2019; Marso 2019). Furthermore, actions taken in the lead up to and during a flood event can result in litigation from damaged parties which contribute to a fractured response. As such, decision-support recommendations from a voting-based framework could increase confidence and coordination in a community's flood response.

The primary goal of this use case was to demonstrate how voting-based preferences of specific flooding scenarios can be used to decide responses for a collection of flood scenarios. A secondary goal is to analyze the cumulative effect of these responses. To design flooding scenarios to use with the web-based ethics framework, we first determined relevant belief features. Belief features were chosen in an exploratory fashion based upon stated priorities of stakeholders as detailed in news publications and personal experience. Relevant belief features used to describe the impact of various flooding outcomes, such as *public costs*, *private costs*, *injuries*, *deaths*, and *environmental damage*. Given the proof-of-concept nature of our application, the process of identifying and clarifying relevant belief features

is beyond the scope of this paper. Next, 17 different scenarios were randomly generated. Scenarios present two options of flooding outcomes and consist of two assets with varying descriptors (e.g. flood national chain grocer in low-income neighborhood or flood middle-income multi-family home). For each asset, unitless values were assigned for the impact of flooding along with each belief feature. Furthermore, each asset description includes an illustration and a text (Figure 2).

To collect preference data, we had freshmen vote on a five-scenario subset of the 17 scenarios during an ethics module in the University of Iowa, College of Engineering's Introduction to Engineering Problem Solving course. To build class section-specific preference models, all students



**Figure 2** │ Example scenario and game play provided by the (WE)² framework. Participants are served multiple scenarios where they are provided two outcomes and their descriptions as photos, info bars, and text (displayed when information button, button is click). Upon clicking the action button, e.g. 'Flood This', the response is logged in the database and the participant is served another scenario.

within a section were served the same five scenarios (Table 1). All responses were recorded in a web-accessible database and related to unique user ids and the class section keyword name.

To learn models from the collected data, votes were aggregated by class section. Participants who did not provide a vote for all five scenarios were not included in model generation. Preference models were learned using multiple linear regression. Five models were learned including one for each class section and one model that is the average of the four class sections. Five model parameters ($\beta$ values) describe each model, one parameter for each belief feature provided in scenarios (Figure 3 and Table 2). A negative model parameter indicates the model's preference to minimize the impact along the given dimension. A positive model parameter indicates the preference to maximize the impact along the given dimension. This is due to

**Table 1** | Each class section is referenced by a letter, section name

| Section name | N | Scenarios |
|---|---|---|
| a | 103 | 2, 3, 10, 15, 16 |
| b | 111 | 2, 3, 4, 5, 6 |
| c | 106 | 8, 9, 10, 11, 12 |
| d | 89 | 13, 14, 15, 16, 17 |

N details that how many students provided preferences through the game play. Scenarios column lists which five scenarios were served to each class section.
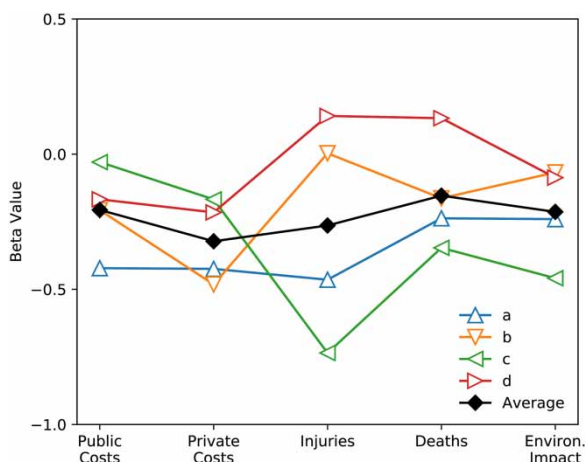


**Figure 3** | Learned preference models for each of the four class sections and their average. Negative beta values indicate a preference to minimize the impact for a category.

the experimental design and our definition of marginal utility. For example, if it is preferred to minimize public costs, then the difference in public cost between the preferred option and the undesired option would be negative; or, that the preferred option results in fewer public costs. This negative value is multiplied by a negative $\beta$ value to result in a positive contribution to the utility of that choice. All learned models minimize public costs to some degree; model a gives the most priority to public cost minimization and model c gives the least. However, no model gives minimizing public cost the highest utility against the other belief features. Models b, d, and Average give the highest utility to private cost minimization. Models a and c give the highest utility to the minimization of injuries. However, model a maximizes the minimization of public costs, private costs, and injuries almost equally.

To analyze the impact of the learned models, we simulated decisions on all 17 scenarios using the learned models. To add reference outcomes, we also simulated decisions of five models that, respectively, only prioritized one of the belief features. In effect, these reference models made decisions based on a single criterion instead of some combination of the five criteria of the learned models. Decisions, realizing the flood of the left or right scenario, for all 17 scenarios were made (Figure 4). The procedure is as follows: the dot product was taken between the vector $(X_A - X_B)$ and each of the ten $\beta^T$ models. The result, the mean utility difference between option A (left) and option B (right) determines how each model would decide between the two outcomes; for a negative difference, chose option A (left), and for a positive difference, chose option B (right). Ties, which only occurred with the reference models, were noted and interpreted as that outcome could go either way. In general, learned models voted similarly, showing the agreement of vote on 11 out of 17 scenarios.

To compare how the outcomes differed between the models, we calculated the total impact each model *avoided* along with each belief feature and normalized them against the minimum and maximum possible avoidable damage (Figure 5(a)). A score of one indicates that the maximum damage was avoided, while a score of zero indicates that the minimum damage was avoided. All learned models avoided high levels of death and environmental impacts,

**Table 2** | Beta values for each preference model used to make decisions

| | Beta values | | | | |
|---|---|---|---|---|---|
| **Experimental model** | **Public Costs** | **Private Costs** | **Injuries** | **Deaths** | **Environmental Impact** |
| a | − 0.42 | − 0.42 | − 0.46 | − 0.24 | − 0.24 |
| b | − 0.21 | − 0.48 | 0.00 | − 0.16 | − 0.07 |
| c | − 0.03 | − 0.17 | − 0.74 | − 0.35 | − 0.46 |
| d | − 0.17 | − 0.22 | 0.14 | 0.13 | − 0.09 |
| Average | − 0.21 | − 0.32 | − 0.26 | − 0.15 | − 0.21 |
| Minimize Public Costs | − 1 | – | – | – | – |
| Minimize Private Costs | – | − 1 | – | – | – |
| Minimize Injuries | – | – | − 1 | – | – |
| Minimize Deaths | – | – | – | − 1 | – |
| Minimize Environmental Damages | – | – | – | – | − 1 |

Models a, b, c, and d are the learned models from each class section, and Average is the average of each of these learned models. These models weigh multi-dimensional effects to determine a decision. These models contrast with the reference models, which only consider the impact along with a single category.
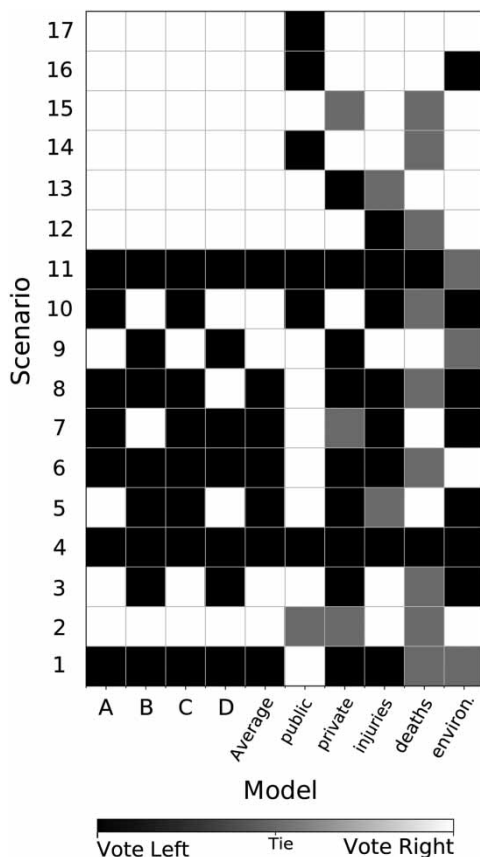


**Figure 4** | Voting results from each of the five learned models and five reference models. Each model was used to vote, left or right, on the outcome of all scenarios generated for the flooding use case.

while showing less agreement and less ability to avoid public costs. The same exercise was performed for the reference models (Figure 5(b)). Although each model scored maximally in their respective categories, it does not suggest with these example scenarios that it is a productive strategy to prioritize only one category over all others.

Because each of the 17 scenarios has only two options, flood left or flood right, there are $2^{17}$ – or 131,072 – unique decision combinations. We normalized and ranked all outcomes along with each category, allowing us to build percentile curves (Figure 6). Only model a score above the 50th percentile in all categories. The remaining four learned models performed above the 50th percentile in all but one category, public costs. Across the five categories and five models, 17 out of the 25 outcome scores ranked above the 90th percentile, which indicates strong performance for all learned models.

To compute an overall outcome score for each of the $2^{17}$ voting possibilities, we performed a simple summation of the normalized outcome scores of each category for each model. These overall scores, too, were ranked and percentile scores calculated (Table 3 and Figure 5(c)). Shockingly, model a achieved the highest possible overall score of 3.922 (out of a maximum of 5) for rank 1 of $2^{17}$. Model c achieved a cumulative outcome score of 3.849 for a rank of 6 of $2^{17}$. All learned models achieved scores that put them in the 98th percentile of all outcomes in avoiding
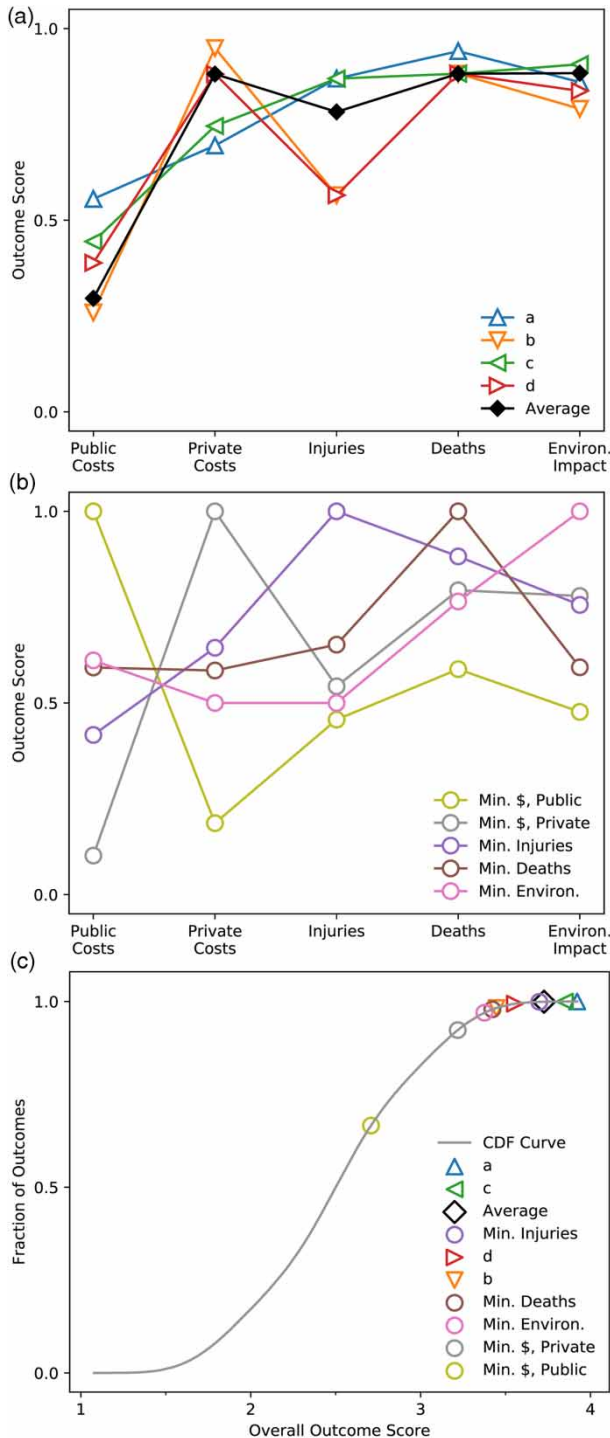
**Figure 5** │ Outcome scores are calculated as the avoided damage along a belief feature normalized by the maximum possible avoidance. Outcome scores were calculated for the learned models (a) and reference models (b). A model's cumulative outcome score is calculated as the sum of all outcome scores. Each model's cumulative score is ranked against all possible cumulative outcomes (c). Model a performed achieved the highest possible overall outcome score of 3.92 and a rank of 1 in $2^{17}$.

total damages. Interestingly, only one of the single objectives, reference models performed better than any of the learned models.

## DISCUSSION

The framework presented here demonstrates the ability to use a voting-based system to aggregate human preferences to ethical decisions in smart water systems. Collected data can be used to learn models of preferred behavior which can then be used to make decisions on new scenarios. This data-driven approach is novel in helping researchers *learn* the utility function from a large, potentially very large cohort of people and not assume an understanding of the utility function to be used to judge outcomes *a priori*. To this end, water professionals and researchers can investigate how algorithmic components of smart water systems or disaster response perform in relation to people's normative expectations of right and wrong.

The framework follows a consequentialist ethical theory, as the definitions of utility and performance rankings are based upon the *outcomes* of each scenario. However, learning utility functions without any *a priori* knowledge mitigates the limitations of traditional consequentialist approaches. As discussed earlier, consequentialist approaches come with limitations such as what factors (belief features) to use to describe outcomes and the fairness-efficiency tradeoffs between the following different moral theories (utilitarian vs. egalitarian strategies). One strength of this approach is that belief features may also be crowdsourced, which allows many people to define an inclusive list of important belief features. Another strength is that by learning the models of action from people's decision, it is unnecessary to proclaim beforehand what fairness-efficiency tradeoffs should be made. Instead, the fairness-efficiency priorities are captured within the learned models. However, these learned models may not agree with the institutional understanding of rights or justice. Further effort may be required to integrate learned preference models with our ideals and aspirational sense of justice and fairness. Finally, this approach currently does not address the difficulty of decision making under uncertainty. This is an area of future work. Overall, we identify the
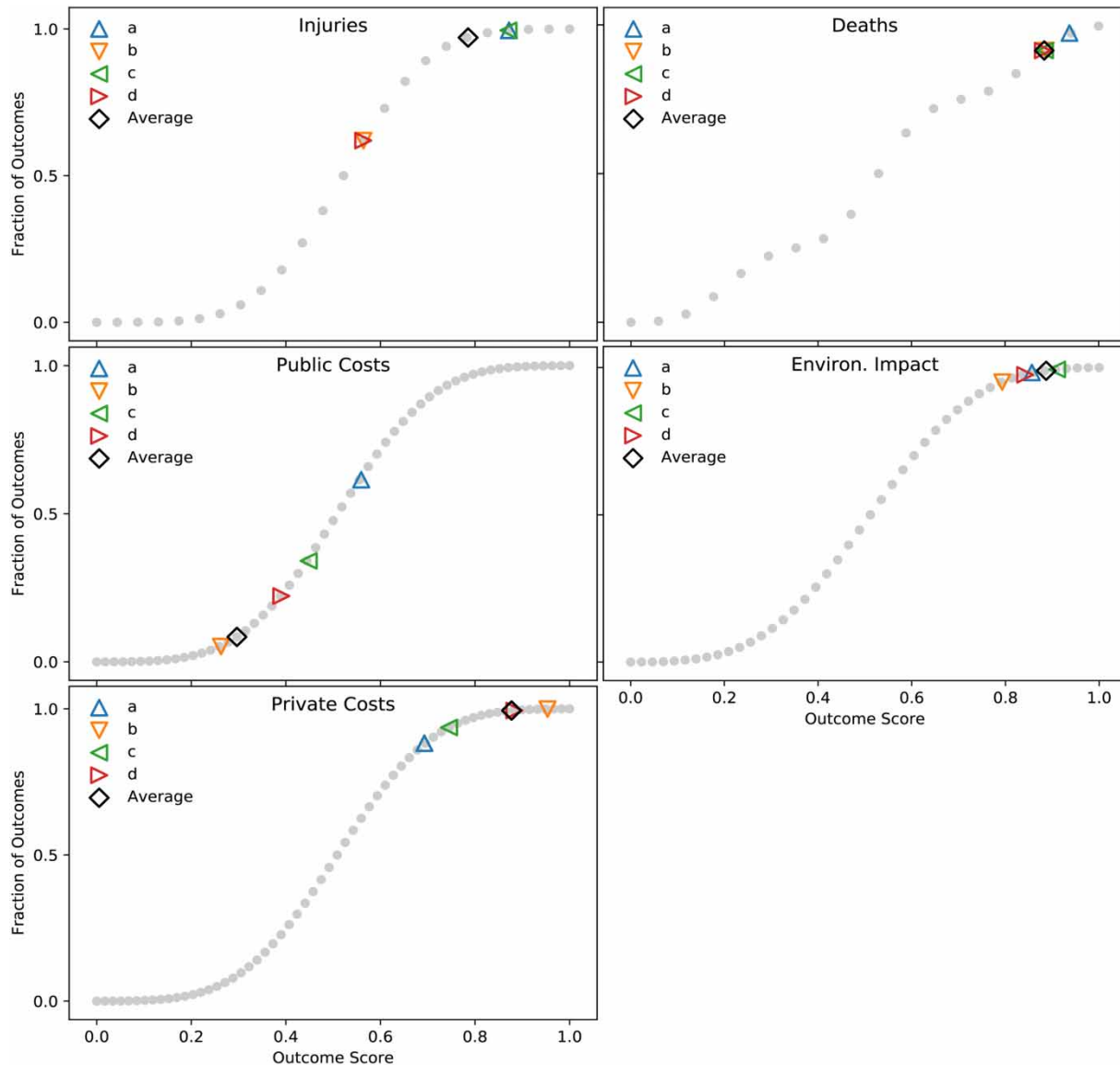
**Figure 6** | Ranking curves for each of the five learned models along each of the five relevant belief features. Overall, the learned models performed well. Model a, which had the highest cumulative score when considering minimization of damages, scored highest among the models in three categories: deaths, injuries, and public costs. Yet, it also performed the worst in minimizing private costs and was third in environmental impact minimization.

inability of a single person, namely the researcher, to insert bias into the calculations as a positive outcome of using this framework.

Results from the experiment show the remarkable possibility for the models to choose outcomes that rank highly when considering the cumulative outcome. Models a and c appear to achieve their astonishingly high ranking because they chose outcomes that did not favor private cost minimization at the expense of public costs. This observation is

supported by the $\beta$ values for public and private costs that are equal for model a. These findings are relevant only within the context of our theoretical proof-of-concept. More robust studies are required to make that claim generally. Furthermore, the cumulative scores and ranking method used here communicate only a single concept of 'success'. It is reasonable to justify other performance metrics beyond summing the normalized results of each category. Rather, the cumulative outcome score analysis can be

**Table 3** | Cumulative impact scores for each learned and reference model in the rank order

| Experimental model | Cumulative outcome | Percentile | Rank |
|---|---|---|---|
| a | 3.922 | 0.99999 | 1 |
| c | 3.849 | 0.99995 | 6 |
| Average | 3.726 | 0.99944 | 73 |
| Minimize Injuries | 3.699 | 0.99911 | 117 |
| d | 3.555 | 0.99406 | 779 |
| b | 3.447 | 0.98280 | 2,254 |
| Minimize Deaths | 3.423 | 0.97894 | 2,760 |
| Minimize Environmental Damages | 3.376 | 0.96988 | 3,948 |
| Minimize Private Costs | 3.219 | 0.92303 | 10,089 |
| Minimize Public Costs | 2.708 | 0.66605 | 43,772 |

A total of $2^{17}$ possible outcomes.

instructive as a first step toward an application-specific evaluation technique.

Critiques of paired comparison for modeling moral preferences follow two forms: misplaced moral subject and lack of meaning of dilemmas descriptors. The first form states that the subject, or the moral actor, in the dilemma should move from the individual to the institutional. Consider the Trolley Problem, a classic philosophical dilemma in which the subject, the trolley driver, must choose between two bad decisions (Thomson 1985). One could ask: How did society fail to enforce safety standards such that an individual must intervene in a life and death scenario? Or, why has society allowed public infrastructure to be so underfunded that it poses the risk of catastrophic failure? Broadly, this critique emphasizes that moral and ethical dilemmas should be interrogated from an institutional or societal level and not exclusively at the individual. As such, a goal of smart water research and institutional design may ask: how can organizations be structured such that moral decisions made by an autonomous agent are minimized?

The second criticism is that the characteristic classifiers used in the moral dilemmas under describe scenarios and, in doing so, elicit not a users' moral ideas but their biases along the dimension of the classifiers (Everett Jaques 2019). Put another way, if an experimental setup provides only information on the age or race of a candidate to receive medical treatment what is the experiment doing but forcing participants to display their ageist and racist biases to make healthcare decisions? And, by extension, if we use these data to learn a model of 'ethical' decision making, are we not simply training a model to be biased just like us?

These criticisms are addressed in the methodological design. Contextually, pairwise comparison preference testing should not be understood to possess deep ethical or normative meaning independently of all the exercises in the stated framework. Next, because it is not well understood how thin moral concepts relate to thick moral concepts (Abend 2011), we cannot *a priori* assume that these thin normative preferences collected in pairwise comparison tests translate to thicker concepts. As such, thicker moral and ethical concepts can be incorporated via activities beyond pairwise comparison testing, such as the numerous rounds of interviews with stakeholders (Lee *et al.* 2019). Finally, strong reaction to an action or a method derived from the framework is itself a measure of normative values and can be integrated via the iterative process; collect data and act upon both thin and thick concepts of morality to improve system performance.

In practice, it is unlikely that a strictly consequentialist framework would be operational in real-world scenarios. Instead, a hybrid decision-making process would be employed. For example, a system could trigger automatic human oversight if a decision is anticipated to reach a specified damage threshold. Likewise, because the learned models predict whether one outcome is better than another, a mean utility difference between two outcomes close to zero suggests that there is a very weak preference. Thus, in these cases, a human review could also be triggered. These rule-based heuristics can set guardrails on the strictly consequentialist models, while also providing further opportunity for society-in-the-loop principles. Furthermore, a critical step for future work will be to explore how decisions derived from $(WE)^2$ preference models impact system outcomes via their integration with hydraulic models.

## CONCLUSION

Societal values are embedded into our built world – water systems are no exception – but these values are rarely inspected as part of the scoping of technical solutions.

Instead, values are treated as priors – immutable, unstated, and implicit as they relate to the objectives of infrastructure. Yet, our infrastructure itself evolves. Increasing resolution in sensing and control of our water environment will allow for unprecedented precision of impacts, both positive and negative. At the same time, humans will continue to cede direct decision-making powers to decision-support technologies such as data algorithms. This new paradigm of smart and autonomous water systems will create new operational capabilities and new opportunities to [re]evaluate these values and explicitly incorporate them into operations.

The methodology and 'proof-of-concept' presented here are a first step toward building a framework for engaging people in algorithmic decision making in cases where normative and ethical preferences are considered. We developed the web-based (WE)$^2$, which is a generalized framework with serious gaming to collect normative preferences through paired comparison testing. Although our framework was designed for water applications, the framework is generalizable and can be used for any paired comparison exercise in any field. Preferences collected using (WE)$^2$ can then be used with our data analytics toolbox to build decision-support preference models and investigate their behavior. These resources, including documentation and tutorials, are shared openly and can be found in the project repository.

We observe that the strength of this framework is that it can prime conversations on values and system expectations at every step of the process, forwarding an iterative process. By doing so, practitioners can work to unobscure AI, ML, and data-driven techniques from behind jargon and demystify the 'black box' processes of decision-support algorithms. We anticipate benefits in deploying our integrated framework in education, operational, and outreach contexts.

Efforts toward incorporating ethics and norms into smart systems must be considered in a sociotechnical context. Importantly, this means that the solution to the development of a technology that is faithful to a society's values may not necessarily be technical in nature at all. Instead, findings from studies could support a structural, social solution as opposed to a solution reliant upon a technological artifice. Though aspects of the work can be technological, it should not preclude results finding that a structural or institutional solution is preferred. The application of AI, ML, and data-driven techniques to water sector problems does not alone make a system 'smart'. Instead, 'smart' water should be conceived as the use of these tools to forward an explicitly recognized objective of the society.

## DATA AVAILABILITY STATEMENT

All relevant data are available from an online repository or repositories (https://github.com/uihilab/waterethicswebengine).

## REFERENCES

Abend, G. 2011 Thick concepts and the moral brain. *Archives Europeennes de Sociologie* **52** (1), 143–172. https://doi.org/10.1017/S0003975611000051.

Alexander, L. & Moore, M. 2016 Deontological ethics. In: *The Stanford Encyclopedia of Philosophy* (E. N. Zalta, ed.). Winter 201. Metaphysics Research Lab, Stanford University.

Awad, E., Dsouza, S., Kim, R., Schulz, J., Henrich, J., Shariff, A., Bonnefon, J. F. & Rahwan, I. 2018 The moral machine experiment. *Nature* **563** (7729), 59–64. https://doi.org/10.1038/s41586-018-0637-6.

Bargheer, S. 2018 A pragmatist theory of morality. In: *Moral Entanglements: Conserving Birds in Britain and Germany*. University of Chicago Press, Chicago, p. 336.

Binmore, K. 1998 Egalitarianism versus utilitarianism. *Utilitas* **10** (3), 353–367. https://doi.org/10.1017/S0953820800006269.

Bosman, J., Turkewitz, J. & Williams, T. 2019 In the Midwest, relentless floods dredge up 'shadow' of 1993. *New York Times*, June 3, 2019. Available from: https://www.nytimes.com/2019/06/03/us/midwest-floods.html.

Bykov, A. 2019 Rediscovering the moral: the 'old' and 'new' sociology of morality in the context of the behavioural sciences. *Sociology* **53** (1), 192–207. https://doi.org/10.1177/0038038518783967.

Carson, A., Windsor, M., Hill, H., Haigh, T., Wall, N., Smith, J., Olsen, R., Bathke, D., Demir, I. & Muste, M. 2018 Serious gaming for participatory planning of multi-hazard mitigation. *International Journal of River Basin Management*. https://doi.org/10.1080/15715124.2018.1481079.

Chevaleyre, Y., Endriss, U., Lang, J. & Maudet, N. 2007 A short introduction to computational social choice, pp. 51–69. https://doi.org/10.1007/978-3-540-69507-3_4.

Conitzer, V., Brill, M. & Freeman, R. 2015 Crowdsourcing societal tradeoffs. In *Proceedings of the 2015 International Conference on Autonomous Agents and Multiagent Systems*, pp. 1213–1217. Available from: http://dl.acm.org/citation.cfm?id=2773305.

Conitzer, V., Sinnott-Armstrong, W., Borg, J. S., Deng, Y. & Kramer, M. 2017 Moral decision making frameworks for artificial intelligence. In AAAI Conference on Artificial Intelligence, pp. 4831–4835. https://doi.org/10.1111/j.1467-8276.2009.01357.x.

Demir, I., Yildirim, E., Sermet, Y. & Sit, M. A. 2018 FLOODSS: Iowa flood information system as a generalized flood cyberinfrastructure. International Journal of River Basin Management. https://doi.org/10.1080/15715124.2017.1411927.

Demiray, B. Z., Sit, M. & Demir, I. 2021 D-SRGAN: DEM super-resolution with generative adversarial networks. SN Computer Science 2, 48. https://doi.org/10.1007/s42979-020-00442-2.

Dubois, D., Rucker, D. D. & Galinsky, A. D. 2015 Social class, power, and selfishness: when and why upper and lower class individuals behave unethically. Journal of Personality and Social Psychology. https://doi.org/10.1037/pspi0000008.

Eller, D. 2018 Flooding has slammed every Iowa county since 1988, some as many as 17 times. Des Moines Register, April 29, 2018. Available from: https://www.desmoinesregister.com/story/money/agriculture/2018/04/29/iowa-flood-center-ranks-disaster-damages-billions-wapsipinicon-river/422336002/.

Everett Jaques, A. 2019 Why the moral machine is a monster. In University of Miami Law School: We Robot Conference, Miami, p. 10. Available from: https://robots.law.miami.edu/2019/wp-content/uploads/2019/03/MoralMachineMonster.pdf.

Freedman, R., Borg, J. S., Sinnott-Armstrong, W., Dickerson, J. P. & Conitzer, V. 2019 Adapting a kidney exchange algorithm to align with human values, p. 115. https://doi.org/10.1145/3278721.3278727.

Grayling, A. C. 1995 Philosophy: A Guide Through the Subject. Oxford University Press. Available from: https://books.google.com/books?id=YLR8ngEACAAJ.

Habibi, H., Nazari, B., Norouzi, A., Noh, S. J., Seo, D. J., Sinha, S., Yu, X., Bartos, M., Kerkez, B., Lakshman, L., Zink, M., Lyons, E., Philips, B. J., Jangyodsuk, P. & Gao, J. 2017 Integrated sensing and prediction of flash floods for the Dallas-Fort Worth Metroplex (DFW). In 97th American Meteorological Society Annual Meeting.

Hagendorff, T. 2019 The Ethics of AI Ethics – An Evaluation of Guidelines, No. 2018 (February). Available from: http://arxiv.org/abs/1903.03425.

Haidt, J. 2007 The new synthesis in moral psychology. Science. https://doi.org/10.1126/science.1137651.

Hardy, K. & Cannon, A. 2019 Iowa flooding: damage from floodwaters reaches $1.6B, Gov. Kim Reynolds estimates. Des Moines Register, March 22, 2019. Available from: https://www.desmoinesregister.com/story/news/2019/03/22/iowa-flooding-damage-estimate-governor-kim-reynolds-federal-flood-relief-trump-nebraska-missouri/3232934002/.

Hitlin, S. & Vaisey, S. 2013 The new sociology of morality. Annual Review of Sociology 39 (1), 51–68. https://doi.org/10.1146/annurev-soc-071312-145628.

Hursthouse, R. & Pettigrove, G. 2018 Virtue ethics. In: The Stanford Encyclopedia of Philosophy (E. N. Zalta, ed.), Winter 201. Metaphysics Research Lab, Stanford University.

Immerman, M. & Immerman, E. 2018 Getting from Push to Pool – An Alternative Starting Point in the Water Quality Debate. Des Moines, IA. Available from: http://regionalstrategic.com/assets/rsq_2018_summer_essay3.pdf.

Jobin, A., Ienca, M. & Vayena, E. 2019 The global landscape of AI ethics guidelines. Nature Machine Intelligence 1 (9), 389–399. https://doi.org/10.1038/s42256-019-0088-2.

Jones, C. S., Davis, C. A., Drake, C. W., Schilling, K. E., Debionne, S. H. P., Gilles, D. W., Demir, I. & Weber, L. J. 2018 Iowa statewide stream nitrate load calculated using in situ sensor network. Journal of the American Water Resources Association. https://doi.org/10.1111/1752-1688.12618.

Jonoski, A. 2002 Hydroinformatics as Sociotechnology. Ihe Dissertation, Taylor & Francis, p. 27. Available from: https://books.google.com/books?id=IVD-FVDxbdAC.

Kelley, T. J. 2019 The fight to tame a swelling river with dams that may be outmatched by climate change. New York Times2, March 21, 2019. Available from: https://nyti.ms/2TLVuKa.

Kerkez, B., Gruden, C., Lewis, M., Montestruque, L., Quigley, M., Wong, B., Bedig, A., Kertesz, R., Braun, T., Cadwalader, O., Poresky, A. & Pak, C. 2016 Smarter stormwater systems. Environmental Science and Technology. https://doi.org/10.1021/acs.est.5b05870.

Krebs, D. L. 2008 Morality: an evolutionary account. Perspectives on Psychological Science 3 (3), 149–172. https://doi.org/10.1111/j.1745-6924.2008.00072.x.

Lee, M., Kusbit, D., Kahng, A., Kim, J. T., Yuan, X., Chan, A., See, D., Noothigattu, R., Lee, S., Psomas, A. & Procaccia, A. D. 2019 WeBuildAI: participatory Framework for Algorithmic Governance. Proc. ACM Hum.-Comput. Interact. 3 (CSCW). https://doi.org/10.1145/3359283.

Marso, A. 2019 Act of god? Or government? In fight over flood, farmers blame army corps of engineers. The Kansas City Star, March 24, 2019. Available from: https://www.kansascity.com/news/state/missouri/article228237519.html.

Mollerup, A. L., Mikkelsen, P. S., Thornberg, D. & Sin, G. 2017 Controlling sewer systems – a critical review based on systems in three EU cities. Urban Water Journal 14 (4), 435–442. https://doi.org/10.1080/1573062X.2016.1148183.

Moor, J. H. 2006 The nature, importance, and difficulty of machine ethics. IEEE Intelligent Systems 21 (4), 18–21. https://doi.org/10.1109/MIS.2006.80.

Mullapudi, A., Wong, B. P. & Kerkez, B. 2017 Emerging investigators series: building a theory for smart stormwater systems. Environmental Science: Water Research and Technology 3 (1), 66–77. https://doi.org/10.1039/c6ew00211 k.

Mullapudi, A., Bartos, M., Wong, B. & Kerkez, B. 2018 Shaping streamflow using a real-time stormwater control network. Sensors (Switzerland) 18 (7), 1–11. https://doi.org/10.3390/s18072259.

Mullapudi, A., Lewis, M. J., Gruden, C. L. & Kerkez, B. 2020 Deep reinforcement learning for the real time control of

stormwater systems. *Advances in Water Resources* **140**, 103600. https://doi.org/10.1016/j.advwatres.2020.103600.

Muschalla, D., Vallet, B., Anctil, F., Lessard, P., Pelletier, G. & Vanrolleghem, P. A. 2014 Ecohydraulic-driven real-time control of stormwater basins. *Journal of Hydrology* **511**. https://doi.org/10.1016/j.jhydrol.2014.01.002.

Noothigattu, R., Gaikwad, S. ʻ. S., Awad, E., Dsouza, S., Rahwan, I., Ravikumar, P. & Procaccia, A. D. 2017 A voting-based system for ethical decision making, pp. 1587–1594. Available from: http://arxiv.org/abs/1709.06692.

O'Neil, C. 2016 *Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy*. Crown, New York, NY. https://www.tandfonline.com/doi/full/10.1080/01972243.2017.1354593.

O'Reilly, T. 2013 Open data and algorithmic regulation. *Beyond Transparency: Open Data and the Future of Civic Innovation* **21**, 289–300.

Overloop, P.-J. 2006 *Model Predictive Control on Open Water Systems*. IOS Press, Amsterdam.

Piff, P. K., Stancato, D. M., Côté, S., Mendoza-Denton, R. & Keltner, D. 2012 Higher social class predicts increased unethical behavior. *Proceedings of the National Academy of Sciences of the United States of America*. https://doi.org/10.1073/pnas.1118373109.

Rahwan, I. 2018 Society-in-the-loop: programming the algorithmic social contract. *Ethics Inf Technol* **20**, 5–14. https://doi.org/10.1007/s10676-017-9430-8.

Rentsch, J. 2019 Can a retention pond be high-tech? Walmart and Maryland officials hope so. *Delmarva Now*, November 20, 2019. Available from: https://www.delmarvanow.com/story/news/2019/11/20/walmart-state-maryland-retrofit-water-retention-ponds-bay-health/4239417002/.

Sadler, J. M., Goodall, J. L., Behl, M., Morsy, M. M., Culver, T. B. & Bowes, B. D. 2019 Leveraging open source software and parallel computing for model predictive control of urban drainage systems using EPA-SWMM5. *Environmental Modelling and Software*. https://doi.org/10.1016/j.envsoft.2019.07.009.

Samuel, A. L. 1960 Some moral and technical consequences of automation – a refutation. *Science* **132** (3429), 741–742. https://doi.org/10.1126/science.132.3429.741.

Schütze, M., Lange, M., Pabst, M. & Haas, U. 2018 Astlingen – a benchmark for real time control (RTC). *Water Science and Technology* **2017** (2), 552–560. https://doi.org/10.2166/wst.2018.172.

Sermet, Y. & Demir, I. 2018 An intelligent system on knowledge generation and communication about flooding. *Environmental Modelling and Software*. https://doi.org/10.1016/j.envsoft.2018.06.003.

Sermet, Y. & Demir, I. 2019 Towards an information centric flood ontology for information management and communication. *Earth Science Informatics*. https://doi.org/10.1007/s12145-019-00398-9.

Sermet, Y., Villanueva, P., Sit, M. A. & Demir, I. 2019 Crowdsourced approaches for stage measurements at ungauged locations using smartphones. *Hydrological Sciences Journal*. https://doi.org/10.1080/02626667.2019.1659508.

Sermet, Y., Demir, I. & Muste, M. 2020 A serious gaming framework for decision support on hydrological hazards. *Science of the Total Environment*. https://doi.org/10.1016/j.scitotenv.2020.138895.

Sharior, S., McDonald, W. & Parolari, A. J. 2019 Improved reliability of stormwater detention basin performance through water quality data-informed real-time control. *Journal of Hydrology*. https://doi.org/10.1016/j.jhydrol.2019.03.012.

Sinnott-Armstrong, W. 2019 Consequentialism. In: *The Stanford Encyclopedia of Philosophy* (E. N. Zalta, ed.), Summer 201. Metaphysics Research Lab, Stanford University.

Sit, M. & Demir, I. 2019 Decentralized flood forecasting using deep neural networks. *arXiv Preprint* arXiv:1902.02308.

Sit, M. A., Koylu, C. & Demir, I. 2019 Identifying disaster-related tweets and their semantic, spatial and temporal context using deep learning, natural language processing and spatial analysis: a case study of Hurricane Irma. *International Journal of Digital Earth*. https://doi.org/10.1080/17538947.2018.1563219.

Stellar, J. E., Manzo, V. M., Kraus, M. W. & Keltner, D. 2012 Class and compassion: socioeconomic factors predict responses to suffering. *Emotion*. https://doi.org/10.1037/a0026508.

Sun, C., Svensen, J. L., Borup, M., Puig, V., Cembrano, G. & Vezzaro, L. 2020 An MPC-enabled SWMM implementation of the Astlingen RTC benchmarking network. *Water* **12** (4), 1034. https://doi.org/10.3390/w12041034.

Susi, T., Johannesson, M. & Backlund, P. 2007 *Serious Games – An Overview* (Technical Report). *Elearning*.

Thomson, J. J. 1985 The trolley problem. *The Yale Law Journal* **94** (6), 1395. https://doi.org/10.2307/796133.

Tolmeijer, S., Kneer, M., Sarasua, C., Christen, M. & Bernstein, A. 2020 Implementations in machine ethics: a survey. Available from: http://arxiv.org/abs/2001.07573.

Trautmann, S. T., van de Kuilen, G. & Zeckhauser, R. J. 2013 Social class and (un)ethical behavior: a framework, with evidence from a large population sample. *Perspectives on Psychological Science*. https://doi.org/10.1177/1745691613491272.

Troutman, S. C., Love, N. G. & Kerkez, B. 2020 Balancing water quality and flows in combined sewer systems using real-time control. *Environmental Science: Water Research & Technology* 1357–1369. https://doi.org/10.1039/c9ew00882a.

Tsukida, K. & Gupta, M. R. 2011 *How to Analyze Paired Comparison Data*. UWEE Technical Report.

Vojinović, Z. & Abbott, M. B. 2012 *Flood Risk and Social Justice: From Quantitative to Qualitative Flood Risk Assessment and Mitigation*. Water Intelligence Online. https://doi.org/10.2166/9781780400822.

Weber, L. J., Muste, M., Bradley, A. A., Amado, A. A., Demir, I, Drake, C. W., Krajewski, W. F., Loeser, T. J., Politano, M. S., Shea, B. R. & Thomas, N. W. 2018 The Iowa watersheds

project: Iowa's prototype for engaging communities and professionals in watershed hazard mitigation. *International Journal of River Basin Management*. https://doi.org/10.1080/15715124.2017.1387127.

Xiang, Z., Yan, J. & Demir, I. 2020 A rainfall-runoff model with LSTM-based sequence-to-sequence learning. *Water Resources Research* **56** (1), e2019WR025326.

Xu, H., Windsor, M., Muste, M. & Demir, I. 2020 A web-based decision support system for collaborative mitigation of multiple water-related hazards using serious gaming. *Journal of Environmental Management* **255**, 109887. https://doi.org/https://doi.org/10.1016/j.jenvman.2019.109887.

Yildirim, E. & Demir, I. 2019 An integrated web framework for HAZUS-MH flood loss estimation analysis. *Natural Hazards*. https://doi.org/10.1007/s11069-019-03738-6.

Zagzebski, L. T. 2004 *Divine Motivation Theory*. https://doi.org/10.1017/CBO9780511606823.