

MoralAgreements: A game for crowdsourcing ethical opinions

113¹, 112², 131³, 120⁴, and 129⁵

Department of Information Science and Media Studies, University of Bergen

Abstract. As machines are becoming increasingly independent through rapid advances in artificial intelligence (AI) technology, questions arise about how ethical principles should guide their moral behavior. One major challenge involves the collection and assessment of social expectations to how these questions should be solved. In this paper, we propose a framework for collection and measurement of public opinion on ethical dilemmas. Using our framework as a tool, researchers can add their own collections of dilemmas, making it simple to investigate public opinion on dilemmas of their own choosing. In other words, the framework is a platform for crowdsourcing both public opinion on moral dilemmas as well as moral dilemmas. We describe how gamification can be implemented to improve the experience of participating and encourage participation in our tool. In addition, we demonstrate a method of assessing social expectations of AI ethics through comparison of collected data with existing moral theories. The capabilities of our framework is demonstrated through the deployment of a working prototype that implements gamification, compatibility with five moral theories, and a set of ten ethical dilemmas.

Code repository: <https://tinyurl.com/yaxujfyl>

Prototype: <https://moral-agreements.herokuapp.com/> (pin code "0001")

Keywords: Machine learning approaches · Codes of ethics · Crowdsourcing.

1 Introduction

Allowing machines to make decisions independent from human intervention requires them to make moral judgements that live up to certain ethical standards. Reaching a common moral code that can be implemented into artificial intelligence (AI) systems will require extraordinary joint efforts from ethicists, legal scholars, and moral philosophers, as well as computer scientists [43]. Artificial agents could be built by implementing predetermined ethical views, such as existing theories in moral philosophy [5]. However, it is not clear which theory would be best suited for the task. Quantitative contributions from the public can clarify what are the social expectations from machine-made decisions. A social agreement of what should be ethical behavior in machines can contribute to the foundation needed to build ethical artificial agents. In addition, it can help identify which moral theories would be seen as more desirable to the public.

We investigate a game-based approach to collecting public opinion on ethical dilemmas. In recent years serious games have been used to improve the experience of participating in crowdsourcing and encourage participation [33, 35, 59, 55, 41]. The Moral Machine Experiment [4] demonstrated that this type of tools can spur massive engagement and drive public discussion on machine ethics [3]. This was done through asking respondents to answer different variations of the Trolley Problem for Autonomous Vehicles. Our hypothesis is that serious games can be useful tools to collect public opinion on moral dilemmas.

The Moral Machine’s use of pictorial representations of dilemmas could resemble a game. We explore how other elements from games can be used in a tool that enables gathering and analysis of public opinion on moral dilemmas. In addition, we examine how a comparison between the social agreement and existing theories in moral philosophy can be integrated into such a tool.

Our aim is to build a framework that can be adopted to collect public opinion on moral dilemmas for AI ethics through a gamified experience. The collected data can be used to analyze the discrepancy between moral theory and public opinion on ethical behavior. In this context, public opinion can be seen as *social choice*. Social choice in *social choice ethics* refers to the aggregate ethical views of society, and include issues regarding the ethical views held by individuals [5].

Research question: How does public opinion on the right thing to do in an ethical dilemma compare with moral theories?

Our main contributions include:

- An adaptive framework for crowdsourcing public opinion on moral dilemmas and comparing them with theories of moral philosophy.
- A proof-of-concept of how game interactions can be used in such a framework.

We develop a prototype to explore the use of a game to compare public opinion on ethical dilemma solutions with theories of moral philosophies. In order for our prototype to be able to do this, it needs to:

- Present dilemmas in a clear and concise way.
- Have well-defined options to choose from.
- Collect and store data from users’ answers.
- Compare participant answers to different ethical theories.
- Increase engagement through gamification.

The game also needs to include the ability to easily implement different sets of dilemmas, through a function to create new sets within the application itself. By enabling researchers (hereby referred to as ”users”) to add new dilemmas, the framework also functions as a tool for crowdsourcing ethical dilemmas. In addition to this, anyone wishing to play the game should be able to do so easily in order to reach as high a number of participants (hereby referred to as ”players”) as possible. This can be solved by having the game openly accessible to play online.

The paper is organized as follows: Section 2 introduces the definitions of the basic concepts we rely on. Section 3 describes work related to our research

problem. Section 4 provides a description of the ideal game as a tool identified through exploration of game design for serious games. Section 5 contains a presentation of our prototype, the dilemmas used in the evaluation of our prototype, as well as a discussion of how constraints and criteria led to the functionality included into the design. Section 6 contains the evaluation of our framework, followed by a summary of the paper in Section 7.

2 Preliminaries

In this section we first present background on player motivation and concepts in game design, then describe moral theories that are used in our prototype.

2.1 Motivation and games

The incentive behind gamifying the activity of crowdsourcing opinions on moral dilemmas is to spur engagement, enjoyment, and participation among players. In this context, a moral dilemma means a Scenario that requires the decision-maker – in our case the player – to consider opposing moral values and then to choose between them. [30] In the following, we present background on motivation and define concepts we rely on in game design.

Motivation. Theoretical frameworks such as self-determination theory distinguish between different types of motivation to explain why individuals contribute to a task [48]. The two most fundamental distinctions is between *intrinsic* and *extrinsic* motivation. The former refers to doing something because it is inherently interesting or enjoyable. Extrinsic motives, by contrast, do not inherently relate to the task itself, but to external reasons such as approval from others [48].

Motives for participation in traditional surveys and crowdsourcing can be divided into three main categories: altruism (e.g., the survey advances a purpose important to the respondent, or the respondent answers to a social obligation), survey-related reasons (e.g., the respondent is interested in the topic), and egotistical reasons (e.g., money, or it brings the respondent joy) [51].

Gamification and serious games. In our approach we make use of *gamification* to create a *serious game*. Gamification is "the use of game elements in non-gaming systems to improve user experience and user engagement, loyalty and fun" [45]. This includes applying *extrinsic* motivators with the goal of creating and maintaining *intrinsic* motivation for the player. Since the main purpose of our game is collection of data rather than entertainment, it should be characterized as a serious game. Serious games are defined as "Game-based activities designed to promote a desired action" [45].

Serious games can, to a larger degree than traditional surveys and crowdsourcing tasks, satisfy egotistical motives for participation, as well as satisfy participants' curiosity for a topic [15, 31]. Previous studies have also demonstrated that gamification of crowdsourcing tasks is effective for increasing participation as well as the quality of contributions [36, 15, 35, 31].

2.2 Moral Theories

Our goal was a direct comparison of public opinion against suggestions of competing moral theories. For our prototype we focused on five theories. Among them, the three most known theories in moral philosophy, representing the three competing views of normative ethics:

- *Act Utilitarianism* as representative of Consequentialism
- *Kantian Ethics* as representative of Deontological Ethics
- *Aristotelianism* as representative of for Virtue Ethics

Further, we included two less popular theories that follow more extreme ethical principles. Even though these theories are less relevant in the domain of Machine ethics, benchmarking players against those extreme positions serves our idea of gamification and could help to engage our users:

- *Ethical Egoism*
- *Classical Natural Law Theory*

There exists a large variety of other moral theories and we will mention some of them in section 7.1 about potential enhancements of our framework. Now, we briefly introduce the selected theories:

Act Utilitarianism

As a form of consequentialism, act utilitarianism is based on the idea that the consequences of an act alone decides its moral permissibility [52]. The basis for the theory forms the principle of utility, which was first described by Jeremy Bentham in 1776 as "the greatest happiness for the greatest number" [7]. According to act utilitarianism, an act is considered morally right if it maximizes the net increase in happiness for all parties involved, where happiness is commonly understood as pleasure minus pain [52]. Given the possibility of travelling through time, act utilitarianism would therefore permit the use of a time machine in order to kill Adolf Hitler as a baby, as this would save millions of innocent lives [54].

Kantian Ethics

Kantian Ethics represents a deontological ethical theory that goes back to the German philosopher Immanuel Kant [28]. In contrast to utilitarianism, it does not take the consequences of an action into consideration. Instead, the theory assumes that a good will is the only virtue that can be unqualifiedly good [6]. According to Kant, morality requires one to be commanded by it [26] and therefore a good will must always be the result of reason and duty [58]. He claims that everyone should act according to the so called *Categorical Imperative* [28]:

- *Universalizability*: Act follows only moral principles that you will to be applied universally.
- *Humanity*: Act treats humanity always as an end and never as a means to an end.

- *Autonomy*: Agent follows the categorical imperative of one’s own free will and not because of external influences. [28]

Only actions that follow the the categorical imperative, are morally good. According to Kantian Ethics, it is a universalizable moral principle to not lie. Therefore one is obliged to always speak the truth, even in situations where lying would result in better consequences. Thus, lying to the Nazis about the location where Jews are hiding, would be considered morally wrong [57].

Aristotelianism

Virtue ethics describe a set of moral theories that define one’s inner identity and character as the root of morality. [17]: Because one’s actions are just a reflection of one’s inner identity, an action is considered moral, if one’s inner identity itself is moral. According to virtue ethicists, a moral inner identity is achieved through the possession of certain virtues (= character traits) [24]. Which virtues one must possess is what differentiates theories within virtue ethics. Aristotelianism represents one of those theories and defines the following moral virtues [1]:

- *Courage* in the face of fear
- *Temperance* in the face of pleasure and pain
- *Liberty* with wealth and possessions
- *Magnificence* with great wealth and possession
- *Magnanimity* with great honors
- *Proper ambition* with normal honors
- *Truthfulness* with self-expression
- *Wittiness* in conversations
- *Friendliness* in social conduct
- *Modesty* in the face of shame
- *Righteous indignation* in the face of injury

In order to make a judgement about right or wrong, one must analyze and conclude what an individual who possesses all those virtues, would do in that particular situation [17].

Ethical Egoism

Ethical Egoism goes back to the philosopher Henry Sidgwick [50] and represents a form of consequentialism. It differs from most other consequentialist theories (e.g. utilitarianism) in that it considers only one’s own interests rather than the happiness of all involved parties [50]. Accordingly an act is considered moral, if it maximizes the self-interests of the doer [49]. Self-interests are thereby defined in the long-term perspective and do not include short-term desires: Even if it was one’s personal wish to smoke a cigarette, ethical egoism would consider this immoral, because it contradicts with the self-interests of living a long and healthy life [42].

Classical Natural Law Theory

The idea of natural law theory reaches back to ancient Greek philosophy and builds on the idea that not only an act itself but also one's motives must be good in order to be truly moral. What is considered right depends solely on the rational nature of human beings [46]. In other words: Right is what is natural. The Christian philosopher Thomas Aquinas later defined what is natural through the so called primary precepts of natural law that represent characteristics of a good life [56]:

- Live peacefully
- Be reasonably prosperous
- Procreate
- Acquire wisdom through philosophy
- Pass on wisdom
- Express gratitude through praise

Furthermore, secondary precepts are rules and principles whose general observance support humans in living up either to those primary precepts or to other secondary precepts. Examples for secondary precepts are '*Do not murder*', '*Do not steal*' or '*Do not lie*', as humans are considered social animals and acts like killing, stealing or lying would negatively affect social relationships. Aquinas defined an action to be moral if it generally contributes to primary or secondary precepts [56]. Accordingly, getting drunk is considered immoral, because it contradicts with the primary precept of living peacefully and at the same time inhibits the naturally human ability to reason.

3 Related Work

Kim et al [29] present a concept called Deliberative Democracy in the context of bioethics. Motivated by the limitations of expert panels addressing ethical issues and the validity issues of traditional normative opinion surveys, they discuss how the general public opinion can best be incorporated in policy-making. The theory is based on the assumptions that informed and deliberative input from citizens is the ideal for democratic governance and that people might change their minds based on evaluation of other's opinions. Various methods that allow ordinary citizens to interact and discuss with each other in order to end up with a shared public opinion are therefore suggested. The suggested methods should always be based on an adequate, unbiased and balanced set of expert information and various perspectives. As the Deliberative Democracy methods do not consider public opinions to be fixed, they achieve a higher level of public conformity and should – according to Kim et al – be favored over traditional normative opinion surveys.

Awad et al [4] address the challenge of measuring societal expectations about how artificial agents act in the face of moral dilemmas and the ethical principles that should guide these decisions. By creating an online platform that present participants with moral dilemmas faced by autonomous vehicles, they gathered

40 million decisions from millions of people over the entire world. They point out that such a massive and cross-societal response would have been infeasible to achieve by using standard academic survey methods. A limitation of their approach is that the samples of participants is not guaranteed to match the socio-demographics of each country. The dilemmas in the Moral Machine are different variations of the trolley problem in the context of autonomous vehicles, communicated to participants through images. The reception of the Moral Machine gave the researchers a unique opportunity to summarize global ethical preferences, report variations in these based on demographics, and document variation across cultures. The results show that cultural differences exist, but also demonstrated a relative agreement over broad regions of the world.

Cushman and Young [13] argue that the conflicting outputs from different moral theories regarding moral dilemmas can be traced to human psychology. In the human mind, a moral dilemma may arise as a result of different answers received from different psychological systems, including cognitive and affective systems, when faced with a moral question. The set of principles that a moral theory applies in order to make a moral decision are explained to be formalized axioms rooted in different mechanisms of moral judgement. As a result, moral theories conflict in similar places as where psychological mechanisms conflict. A moral theory cannot simultaneously satisfy the different constraints of psychological systems and, accordingly, the conflict is unavoidable. In order to overcome these obstacles, the authors suggest that each individual theory should be perfected, non-moral practical reasoning should assist moral decisions, and finally reduce the frequency of which moral dilemmas occur through reshaping of our own psychology through argument and education, reevaluating moral conflicts, and remodeling social institutions.

In [17], Garnier argues that virtue ethics should be used in conjunction with consequentialism and the four principles of deontology as it would enhance the approach to moral dilemmas for medical professionals. The argument is that if we follow ethical principles based on reason alone, we forget the nature of the agent. Gardiner puts emphasis on emotion, as it is essential in the individual's judgment process and decision making. By including virtue ethics, a health worker will "flourish" which he describes as a person whom fulfils their purpose and function of human beings. The process will be natural as living by moral virtues over time will make them embedded in the individual. He does not propose a given set of virtues to inhabit, but rather encourage the individual to choose themselves based on the unique situation they are facing.

Young [60] discusses what he calls the "paradox of fiction", the concept of whether or not people can have genuine emotions in response to fictional events or people, and how it affects studies performed in virtual environments. Through analysing previous work in such studies, Young discusses different theories that have tried to solve this paradox. He presents two opposing views with differing theories on how humans experience emotion. Drawing comparisons to emotions people have for characters in books and movies, he argues whether or not these feelings are 'real' compared to emotions felt in real situations. Young concludes

that emotions subjects feel during studies in virtual environments are real and should be treated as such in psychological studies.

In their book, Christen et al.[11] set out to analyze moral behavior in common video games while demonstrating the potential that lies in these games. This demonstration intends to shed light on the lack of empirical studies connecting video games and morality in terms of the danger the former poses to the latter, which would be beneficial for both moral researchers and video game designers. Their analysis involves looking at a selection of video games that cover a large range of mechanisms representing ethical behavior. They argue that the relationship between video games and morality, in the current state of video game development, is much more complex than what is currently being discussed in public debates on the subject. Their end position is that, although methodologically demanding, conclusions about real world behavioral dispositions can be drawn from how players behave in game scenarios.

4 Gamifying the crowdsourcing of moral choices

We want to gamify the crowdsourcing of moral choice and enable comparison of the results to existing philosophies in moral theory. To this end we would need a tool that solicits a participant's opinion in moral dilemmas and provides game incentives to participate. In addition, the tool needs to take on new sets of dilemmas and provide respondent results to researchers. We here describe the necessary properties of this tool.

4.1 Tool properties

Flexibility. Our proposed tool needs to be flexible to serve its purpose of a framework for exploring public opinion on different types of ethical dilemmas and comparing it to established moral theories. Namely, the tool should, as far as possible, leave decisions on what questions are asked, how they are asked, and which theories should be considered, to the creator of a survey. Examples of how flexibility can be supported, includes providing the ability to customize answer options and the option of having illustrations supplement dilemmas.

Ease of use. The technical abilities of a person or team should not be a limitation to making use of our tool. With this in mind, design solutions should include the possibility for entering a set of dilemmas that is easy to use and minimize the room for error. One solution could be implementing a form that takes the user through the process of entering the dilemma set step by step. This would leave little room for error and would not require advanced technical abilities from the user. For those that are more technically capable, submitting dilemmas through a web form might be more tedious than it needs to be. These users might prefer uploading a file with their complete set of dilemmas as formally structured data instead.

Data collection. In order to serve its purpose as a tool for crowdsourcing ethical opinions, the game needs to collect and store the answers provided by players. Additional information, such as nationality and demography data about players could extend the possibilities for data analysis. For researchers using the game as a tool to collect ethical opinions, response data needs to be easily accessible in a structured format, ready for further analysis.

Sharing and access. After a set of dilemmas is successfully created, the game session needs to be easy to share and simple to access. The games can be shared through URIs provided for each individual dilemma set. In addition, players who want to explore different sets should be able to do so by browsing and entering openly available dilemma sets through the game menu. This could potentially increase the number of participants for the different dilemma sets, and even generate participation without sharing of particular URIs.

Platform and design. By developing the game as a web application, accessible through the use of a web browser with no additional downloading or installing, it creates little to no additional threshold for both users and players to enter. Considering that more than half of the traffic from the world’s 4.5 billion internet users comes from mobile devices, additional effort should be put into making the game’s layout work well for screen resolutions typical for these devices [12, 53].

4.2 Lessons from the Moral Machine Experiment

While the success of the Moral Machine Experiment [4] serves as an inspiration for creating our game, we have taken its critique and limitations into consideration. For instance, where the Moral Machine tackles the subject of Autonomous Vehicles specifically, our framework aims to be more flexible, letting researchers use it to design surveys for ethical dilemmas in any field. In addition, the use of images was found to be a source of misinterpretations from respondents, leading to unintended results [3]. We have therefore decided that our game will present dilemmas as text, with images being an optional feature when creating new surveys. The function that lets participants see analytics about their ethics, on the other hand, is something that we have chosen to build on as part of creating a more engaging experience for the players.

Communicating the dilemmas mainly through text instead of with images to avoid ambiguity requires additional considerations. First, the description of a dilemma needs to be precise in presenting the problem. Participants should not have to make assumptions about any of the factors involved in a dilemma. Second, the description of a dilemma needs to be concise. A long description would require excessive effort and attention to read, potentially leading to players not reading the whole text, or diminishing motivation among players. Imprecise or lengthy dilemma descriptions, leading to misinterpretations or unmotivated answers, can potentially damage the reliability of data collected with our framework.

4.3 Player incentives

In order to gain useful data for research there need to be people actually playing the game. While some people would be interested in the game because of their interest in ethics, there could be additional incentives included in order to increase the player-base. Incentives might distort data by affecting player decisions, and should therefore be implemented with care. Here we present some incentives for players that could be included.

Learning and curiosity. Depending on the set of dilemmas, the game can be a learning tool for ethical theories. One option is to offer a player to go through a set of dilemmas again and try to answer as if they were strictly utilitarian, testing how familiar they are with the utilitarian theory. The game could give detailed feedback on a wrong answer, explaining why a theory fits with the correct answer. Letting players compare their decisions with moral theories that often collide with common understanding of ethics, such as Divine Command Theory and Ethical Egoism could contribute to generate a fun learning experience for players. Experiments conducted by Law et al [31] demonstrated curiosity as an intrinsic motivational driver to incentivize workers on Amazon’s Mechanical Turk. With the use of information as a currency, worker retention improved without degrading the performance.

Scenarios. A set of creative and thought-provoking dilemmas could be an incentive in itself. A lot of entertainment is based on tackling ethical problems (e.g. Star Trek, The Good Place, etc.), and an interesting enough set of dilemmas would be intellectually stimulating. Themes like sci-fi and fantasy could add additional layers of interest, as well as expanding the scope of possible dilemmas [9].

Gamification. Giving the player rewards for playing can be an easy incentive to encourage them. Richter et al [45] discuss the social aspects of gaming and how it motivates players to stay engaged. Points for correct answers or completed sets is a way for players to keep track of their progress, and could easily be used for scoreboards where players can compare their score to friends or other players. However, rewards and social comparison should be implemented with care not to push player decisions one way or the other.

“Personality” test. Similar to the learning incentive, dilemma sets could be set up to tell players how they align with ethical theories. Based on their answer the game would give feedback for how much they fit with different ethical theories, and why [27].

Sensation. Features that directly targets the senses of a player can contribute to an enhanced game experience [14]. This includes visual features, such as art style and animation, as well as audio. Implementing animations and audio as features in the game could increase a player’s sense of interactivity and fun, while a pleasing design would make the game more compelling to a player [14].

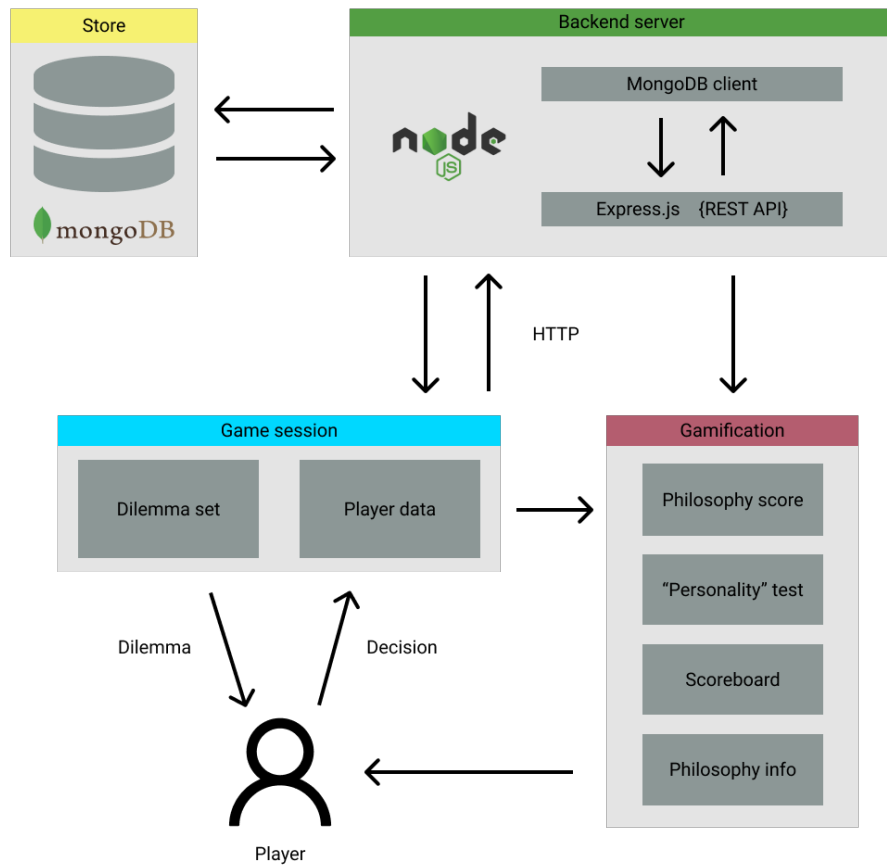


Fig. 1: The framework architecture and game components

5 Prototype

The prototype of our framework implements several of the elements described in the previous section. Figure 1 illustrates the architecture and game components of the prototype. In this section, we present the prototype and its functions, and considerations made when developing it. The prototype can be accessed here: <https://moral-agreements.herokuapp.com/>. Use pin code "0001" to access the sample game.

5.1 Components

Create game

Our prototype lets users input their own sets of dilemmas through a simple web form. Here, the creator of the dilemma enters a description of each dilemma and their respective possible responses, as well as a short introduction to the theme of the game. The descriptions may include images, but this is optional. It is the responsibility of the creator of the set that dilemma descriptions are precise and concise in their form.

When adding a set of dilemmas, users are presented with the option to choose which moral theories should be considered when playing through. We assume that users interested in entering their own sets of dilemmas are experts, and it is left to the creator of a dilemma set to provide correct answers for each moral theory. The moral theories available in the prototype are limited to the ones considered in the preliminaries. The technical abilities of a person or team should not be a limitation to making use of our tool. With this in mind, design solutions should include the possibility of entering a set of dilemmas that is easy to use and minimize the room for error. One solution could be implementing a form that takes the user through the process of entering the dilemma set step by step. Future versions can include additional moral theories.

Dilemma set summary and data retrieval

For analyzing the response to a set of dilemmas, our prototype implements a summary page, as well as the possibility of downloading the collected response data of a set of dilemmas. The summary page includes built-in visualizations of results and usage statistics for a selected dilemma set. Built-in analytics enable stakeholders to get a quick preview of the data collected without requiring additional resources, time, and effort. The complete data linked to a dilemma set is available for download in JSON format for further analysis.

Game

Access. Having deployed our prototype and made it openly available online, a game session can be easily initiated. Each dilemma set has a unique 4-digit ID that can be entered in the home screen of our prototype application in order to access a game. A game can also be accessed through a direct URI. The home

page, shown in Figure 2, is also where the "create game" component can be entered by users.

Game introduction. After entering a game, players are provided with a brief introduction to the theme explored in the dilemmas, as well as the moral theories that are applied. The purpose of this part is to prime the players for the game and spur curiosity in them. Figure 3 shows the game introduction page for our sample game.

Gameplay. In the main part of the game, shown in Figure 4, a player is presented with dilemmas one by one. There are three available options, two choices given by the game in question, and a third option being to "flip a coin". Simultaneously, the player's score for each moral theory is visible at the top of the page. For each answer a player gives, the score of the moral theories is updated corresponding to which answer each of the theories would give. Icons are used as representations for the moral theories.

Scoreboard and social comparison. When answers for all dilemmas in a game session are given, the player is presented with a summary of how these answers compare with standings of moral theories. In addition, the aggregate scores of other players who have previously played are shown. Figure 5 shows a screenshot from the summary page.

"Personality" test and learning. Together with the summary of a game session, players receive additional info about the moral theory with which their answers align the most. This element combines the incentive of curiosity with a "personality" test - the players gain knowledge about moral philosophy while receiving feedback on which moral theory is most in line with the values reflected by their answers in the game.

5.2 Sample Game

To test and demonstrate the functionality of our prototype, we developed and implemented a set of ten ethical dilemmas. Due to the relevancy of the topic, we chose dilemmas related to the current COVID-19 outbreak. The first five dilemmas address the problem of triage and raise the question whether lives can be weighed against each other, and if so in which way. Because some moral theories assume every life to be worth the same, we included an additional option "Flip a coin" for some dilemmas. This option represents a random decision between the given alternatives for the scenario. The other five scenarios are motivated by public discussions about political and individual decisions related to the pandemic. They require evaluating two opposing moral considerations in order to make a decision.

We will now present the dilemmas in detail and describe how different moral theories would resolve them. Each dilemma addresses certain ethical challenges that we highlight below the scenario's description. These underlying challenges

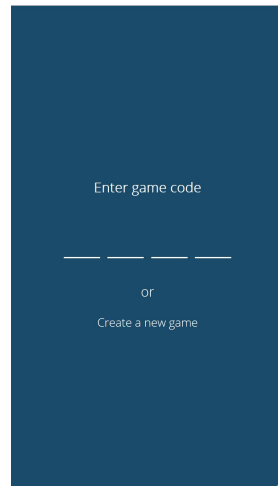


Fig. 2: Home page.

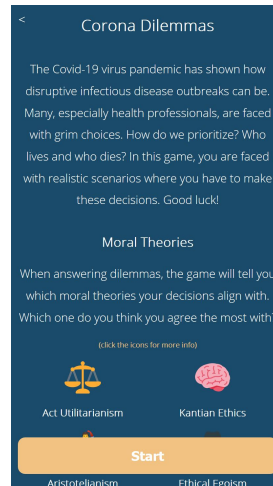


Fig. 3: Game introduction.

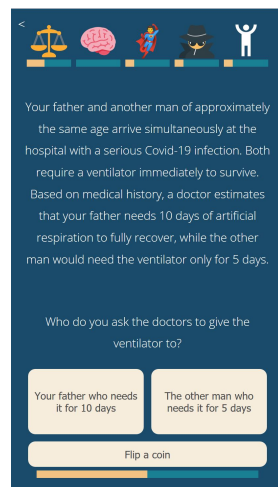


Fig. 4: Gameplay.

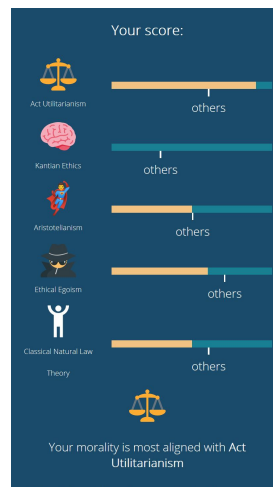


Fig. 5: Summary page.

only explain our motivation behind the chosen dilemmas. They do not appear to the users as part of the game. We are aware that many of the scenarios are possibly addressable by the ethical code of conduct for medical professionals, however we are here interested in what members of the general public would do.[8]

5.3 Dilemmas

Scenario 1: You and your 27-year-old daughter have been tested positive with the COVID-19 virus. Suddenly both of you experience difficulties breathing. When you arrive at the hospital, a doctor states that both of you require an immediate ventilator treatment in order to survive. Unfortunately there's only one ventilator left at the hospital. Who do you ask the doctors to give the ventilator to?

- A: Yourself
- B: Your 27-year-old daughter
- Flip a coin

This scenario addresses the following ethical challenges:

- Would you sacrifice your own life to save someone else?
- Does an interpersonal relationship (parent and child) affect your decision?
- What role does the age of the characters play?

Scenario 2: A 75-year old man and a 16-year old boy both require a ventilator urgently. If they receive one immediately, doctors estimate a survival chance of 90% for the boy; 60% for the elderly man. The chances of survival without a ventilator are 30% for the boy and 0% for the elderly man who would die immediately. You are the head of the hospital and you know that there's only one ventilator left. Who do you decide receive the last ventilator?

- A: The 75-year old man
- B: The 16-year old boy
- Flip a coin

Addressed ethical challenges:

- Should the patients' survival chance affect who receives the treatment?
- In both alternatives, the expected value of saved lives is the same (0.9 lives). But option B accepts a higher risk of both characters dying in order to have a chance that both survive. Is this risk morally permissible?
- What role does the age of the characters play?

Scenario 3: You and your friend David both experience a very serious COVID-19 infection and both of you need a ventilator immediately in order to survive. Because the hospital has only one ventilator left, they can only save one of you. Daniel has a wife and three children, whereas you live alone without close family connections. The doctors ask you to make the decision: Who do you give the ventilator?

- A: Yourself
- B: Your friend David
- Flip a coin

Addressed ethical challenges:

- Would you sacrifice your own life to save someone else?
- Does an interpersonal relationship (friends) affect your decision?
- Should the family situation of the patients play a role in the decision?

Scenario 4: A 95-year old woman at a hospital receives intensive care with a ventilator, but as her attending physician you know that her chance to survive is very small. Suddenly, a 12-year old girl with asthma arrives at the hospital. Without a ventilator she would die, but unfortunately there's no other ventilator available. Do you take the ventilator from the 95-year old woman and give it to the 12-year old girl?

- A: Yes
- B: No

Addressed ethical challenges:

- Should the patients' survival chance affect who receives the treatment?
- Would you actively kill as means to an end?
- What role does the age of the patients play?

Scenario 5: Your father and another man of approximately the same age arrive simultaneously at the hospital with a serious COVID-19 infection. Both require a ventilator immediately to survive. Based on medical history, a doctor estimates that your father needs 10 days of artificial respiration to fully recover, while the other man would need the ventilator only for 5 days. Who do you ask the doctors to give the ventilator to?

- A: To your father who needs it for 10 days
- B: To the other man who needs it for 5 days
- Flip a coin

Addressed ethical challenges:

- Given a shortage in medical resources, should the expected length of the treatment affect the distribution of the limited resources.
- Does an interpersonal relationship (child-father) affect your decision?

Scenario 6 [25, 47, 38, 37]: The society is in a lock down and as a member of the national parliament you need to vote whether schools should open again. On the one hand, lifting restrictions may increase the spread of the disease that is still not completely under control and can result in a higher death toll. On the other hand, children experiencing domestic violence are more vulnerable than before and the risk to their safety is increased. Furthermore, parents can be negatively affected since they need to handle both their jobs and childcare. What do you vote for?

- A: Keep schools closed
- B: Open schools

Opposing moral considerations:

- Reduce the number of deaths by slowing down the spread of the disease
- Reduce the risk for children to experience domestic violence or abuse and the negative effects on the economy

Scenario 7 [10]: Showing that a vaccine works is typically done through a large study, in which thousands of people receive either a vaccine or a placebo, while researchers keep track of how many patients from each group get infected with the disease over a period of several months. A faster alternative is to infect healthy people and check whether those vaccinated don't develop any symptoms. As a scientist working on a vaccine, would you infect healthy volunteers, if you were legally allowed to?

- A: Yes
- B: No

Opposing moral considerations:

- Reduce number of deaths caused by the virus through faster access to vaccine
- Increase the risk for healthy people.

Scenario 8 [22]: Mobile apps using tracking information for infection control could help to slow down the spread of an epidemic and thereby save many lives. They are meant only for the time of emergency, but they share similarities with surveillance systems of authoritarian states and always pose a risk of being misused for other purposes. As a normal citizen, would you personally install such an app?

- A: Yes
- B: No

Opposing moral considerations:

- Avoid unnecessary deaths by better controlling the spread of the disease
- Protect your own privacy.

Scenario 9: A vaccine for COVID-19 has been developed, but the supplies are still low. As Minister of Health, you are in charge of the distribution of the vaccines. In order to save as many lives as possible, a consulting doctor suggests to exclude people with other life-threatening conditions who are expected to die very soon regardless of the vaccine. Would you follow the suggestion and exclude those patients?

- A: Yes
- B: No

Opposing moral considerations:

- Save more lives by vaccinating those with a higher survival chance first
- Respect the right to fair treatment

Scenario 10 [21]: Tonight will be the prom of your girl-/boyfriend and you promised to be there. Just before you leave the house, you receive the information that your brother - who lives in the same household as you - tested positive for COVID-19. You really want to go, but even though you don't show any symptoms, you are advised to self-quarantine for 14 days. The prom itself is a public event with 200 people and legally permitted by the authorities. What do you do?

- A: Still go to the prom
- B: Stay home

Opposing moral considerations:

- Respect the individual right to liberty and the interest to keep ones promises
- Respect other people's right to health by not risking to infect them

5.4 Applying the ethical theories to the scenarios

Moral Reasoning for Act Utilitarianism

Recalling the definition from section 2.2, act utilitarianism defines an act to be moral, if it maximizes the net increase in happiness for all affected parties. Based on this definition, we first need determine the effect of a person's death on the overall level happiness – defined by pleasure minus pain – in order to make judgements on the above dilemmas. For that purpose we will assume that the expected net balance of happiness for a human being is positive, meaning that on average and in the long run, everyone is expected to experience more pleasure than pain in their life. Consequently, the longer a patient lives, the higher their net amount in happiness. Further, we assume that well-being diminishes with age [32, 2], meaning that years of life of young patients have a higher utility than years of life of elderly patients. Based on these assumptions we can conclude: (1) Younger patients benefit more than older ones (because they live longer and

experience a higher net amount of happiness) and (2) women benefit more than men (because women have a higher life-expectancy on average).

Accordingly, the last available ventilator should be given to the 27-year old daughter in Scenario 1 (instead of to her parent), in Scenario 2 to the 16-year old boy (instead of the 75-year old man) and in Scenario 4 to the 12-year old girl with asthma (instead of the 95-year old woman with a small survival chance). As act utilitarianism focuses on the consequences and not on the act itself, it does not matter for the decision whether the ventilator is freely available (Scenario 1 and 2) or already assigned to one of the patients (Scenario 4). In Scenario 3, additional emotional pain (= decrease in net happiness) for the family members of David can be avoided by giving the last ventilator to him. Scenario 5 describes a situation where in either way, one of the middle-aged man dies. In this case, the man who is expected to require the ventilator for less days, should receive it. This way the limited resource becomes available earlier again, what could potentially help saving more lives in the future.

For Scenario 6 two perspectives must be considered: In addition to the economical burden of school closures, children experiencing domestic violence will suffer from both immediate physical pain and potentially long-term mental health problem. On the other hand, people who die due to the faster spread of the virus take away the amount of net happiness they would have otherwise experienced throughout their life. Because no further information is available on the number of affected children, the number of potential deaths or the economical impact, it is not possible to take a utilitarian decision for this scenario without having to make many assumptions. Therefore, we accept both decisions for this dilemma.

Scenario 7 weighs the number of deaths due to the unavailability of a vaccine against the number of deaths caused by infecting healthy volunteers during the vaccine development process. We assume that: (1) clinical studies usually have only a few hundred patients [40], (2) there is no selection bias in the group of volunteers, (3) there is a realistic chance for the vaccine to be effective and (4) the number of daily natural infections is at least ten times higher than the number of active infections as part of such studies. In this case, infecting healthy volunteers for the sake of a faster vaccine development is the preferred option according to act utilitarianism, because (1) more lives are expected to be saved through the vaccine and (2) the age of the deaths does not matter as no selection bias is present in the participants group.

In Scenario 8, mobile tracking data should be used for infection control, because saving lives prevents a large amount of net happiness of getting lost. In comparison, the loss in data privacy can be assumed to have a smaller net happiness.

According to act utilitarianism, a vaccine should be given to those, who are expected to increase their life expectancy the most through the vaccine. Thus, people with other life-threatening conditions should be deprioritized in Scenario 9, as they show a low life expectancy despite a vaccine.

For the final Scenario 10, one should stay home and not attend the prom. There is a high chance of being infectious and infecting other people would increase their level of pain significantly. The short-term out on happiness due to not attending the prom can't balance out this risk.

Moral Reasoning for Kantian Ethics

This moral philosophy assumes an act to be moral, if it follows rational, universalizable duties and ignores personal desires. Kant himself defined "Do not kill" as one of these universal duties and consequently the act of killing as morally impermissible [28]. Thus, taking away the ventilator from the elderly woman in Scenario 4 is morally wrong, because the act would directly lead to her death (= active killing). The death of the 12-year old girl on the other hand is not a direct consequence of the act, but of the disease itself. Although a ventilator could prevent her death, withholding it from her is not an act of killing, because it does not represent the cause of her death.

For the scenarios that raise the question of who should receive the last available ventilator, the universal duty not to kill does not provide a direct response: Even though, one of the characters dies in the end, the act of giving the ventilator to one character is not an act of killing the other, but rather an act saving at least one life. Without treatment, both would have died. We therefore suggest the perfect duty "Treat every life to be worth the same" based on Article 1 of the Universal Declaration of Human Rights [39]. Because agreed by the United Nations and considered a fundamental human right, we conclude that this duty is willed to be applied universally and therefore fulfils Kant's requirements for a perfect duty.

Consequently, flipping a coin is the preferred option for Scenarios 1 and 5, as it gives both patients the same chance of 50% to survive. Also in Scenario 3, the life of the one patient should not be valued less only because he has no family; thus, randomness should decide. In Scenario 2, flipping a coin would leave the boy with a 60% chance to survive, the elderly man with only a 30% chance. On the other hand, giving the ventilator to the elderly man gives him a 60% survival chance and the boy only a 30% chance. So, one must throw the coin potentially twice: First to decide between option B and 'Flip a coin', then to decide between option A and B.

In Scenario 6, the act of opening schools again is morally impermissible, as it will lead to a higher number of deaths and therefore would be an act of killing. Consequently, one should vote in favor of keeping schools closed. Similarly, infecting healthy volunteers with a deadly virus (Scenario 7) is an act of killing and therefore morally wrong. As for Kantian ethics only the act itself and not the consequences matter, the act would be still wrong, even if the vaccine is effective and nobody dies.

The imperfect duty of beneficence defined by Immanuel Kant himself [28] provides the answer to Scenario 8: The principle requires everyone to take actions that can prevent or remove harm from others, as long as those actions don't

contradict with perfect duties. By installing the app, one can help to prevent others from getting infected and therefore prevent harm.

For Scenario 9, the perfect duty to treat every life equally applies again: Everyone should have equal access to the vaccine and no patients should be excluded.

In Scenario 10, we must also consider the perfect duty to keep one's promise [28]. There is no case in which one is allowed to violate a perfect duty and therefore staying home is morally impermissible. Although you willingly accept infecting others by going to the prom, it must be considered that you don't have any symptoms and that the chances of infecting someone who will ultimately die from the disease, are low. Therefore we assume going to the prom not to be a direct act of killing and still permissible according to Kantian Ethics.

Moral Reasoning for Aristotelianism

For virtue ethics as described by Aristotle, an act is moral if it is consistent with the actions of an excellent character. What determines an excellent character, was briefly explained in section 2.2. Thus, in order to make judgements about our dilemmas, we need to conclude what a virtuous person according to Aristotle would do.

In the first Scenario, sacrificing one's child to save one's own life, would contradict with the characteristic of being righteous. At the same time, just flipping a coin would not be courageous (to take a decision). Instead, courage and magnanimity would lead the parent to sacrifice herself for her child. Following the same argumentation, one would also sacrifice oneself in Scenario 3 in order to save the lives of your friend.

The hospital director in Scenario 2 would seek an open and honest conversation with the two patients about the situation, but then not hesitate to take a decision. Without any further information and having to act righteous, he/she would order to give the ventilator to the elderly man, who requires the treatment more urgently than the boy. Also, the virtue of proper ambition would encourage one in trying to save both lives. And this goal could only be achieved by treating the elderly man instead of the boy.

The relationship between a patient and a virtuous physician is characterized by high level of trust, honesty and wittiness. The attending physician in Scenario 4 would therefore not take away the ventilator from the elderly woman without her explicit content, because this would violate their trusted relationship. Also the active act of killing would not be a sign of temperance in the face of pain.

In Scenario 5, the virtues of magnanimity, liberty and courage would require one to set aside personal desires and to make a righteous decision. Therefore, one should not ask to prioritize one's father. At the same time, even though one patient needs the ventilator for less days, giving it to him would not be a righteous thing to do either, as this would discriminate the other patient because of his medical history. Instead one should ask to flip a coin, as this represents the fairest option.

Aristotelianism requires respecting liberty and righteous indignation in the face of injury. Therefore, one should vote in favor of relieving the lock down that fundamentally restricts people in their liberty, even if this results in a higher death toll (Scenario 6). Also, opening schools, decreases the risk of children to experience domestic violence, what is in line with the virtue of temperance in the face of pain. Instead of using force and closing schools completely, a virtuous person would appeal to the general public in an open, honest and friendly way and ask everyone to do their part to prevent infections. This includes leaving the decision of sending children to school up to their parents.

A virtuous scientist (in Scenario 7) would ensure to inform the participants about the risks of the study in an honest and sincere manner and would never try to convince or force anyone to participate. At the same time, he/she is supposed to act out of courage and proper ambition and would therefore risk infecting healthy volunteers for the sake of developing a vaccine that can potentially save hundreds of deaths - always assuming the explicit consent of the participants.

Aristotelianism would never support an obligation to use such an app as described in Scenario 8, as it restricts liberty and righteous indignation requires one to step up against those kind of restrictions. At the same time, as an individual one would still be morally obliged to install the app, because one should set personal interests aside (magnanimity) in order to prevent others from getting infected (temperance in the face of pain). As this step can serve others, this can also be considered a sign of friendliness in social conduct. Based the same argumentation, one should not follow personal desires and rather decide not go to the prom in order to prevent infections (Scenario 10).

Finally, excluding certain groups from the vaccine (Scenario 9) would discriminate those groups and one should position oneself against discrimination (righteous indignation). Also, proper ambition would encourage one to do all in one's power to save those people as well, even if their survival chances are very little.

Moral Reasoning for Ethical Egoism

As summarized in Section 2.2, ethical egoism tries to maximize one's self-interests. Accordingly we need to determine the doer's self-interests for each dilemma.

In Scenario 1, the self-interest could either be to save ones own life or to save the life of one's child. Similarly in Scenario 3, you may either save yourself or sacrifice your life. As both decisions are subjective to the player, our prototype accepts any of the two options to be fitting with this theory. And also in Scenario 5 there could be a self-interest in letting one's father die, so both options are accepted there as well. Given one's personal involvement in all three dilemmas, flipping a coin on the other hand is not considered a fitting answer for ethical egoism, as this would indicate that one doesn't have a clear self-interest.

For Scenario 2 it is up to the head of the hospital to make the decision and we assume public perception to be their dominating self-interest. If revealed that

the hospital director uses a coin to determine how to allocate medical resources to patients, it could potentially cause a public furore. Arguing instead that patients who require treatment most urgently get prioritized, would probably be a publicly accepted decision. Therefore, giving the ventilator to the elderly man, who needs it more urgently, would fulfil the self-interests of the hospital director best.

As the attending physician of the elderly woman in Scenario 4, one's self-interest is to act professionally, because one's future career depends on it. Taking the ventilator from the elderly woman makes one directly responsible for her death, while the death of the 12-year old girl due to a lack of medical resources does not lie in the physician's area of responsibility. Assuming no opposing guidelines have been provided, one should therefore not take away the ventilator.

For the member of the national parliament, the main self-interest lies in representing the opinion of the target voter group, as this might affect their reelection and career. Because the scenario description indicates that the disease is not under control yet and because school closures are usually publicly accepted as long as the disease is not under control, the right thing to do is to vote in favor of the school closures.

In Scenario 7, the scientist working on the vaccine has the primary self-interest in developing a vaccine as fast as possible. This interest justifies infecting healthy volunteers.

The use of mobile apps for infection control (Scenario 8) is again a very subjective decision. Thus, both options are correct answers for ethical egoism.

The Minister of Health in charge of distributing the vaccines is primarily interested in the public perception and in maximizing the chances of getting reelected. Therefore, he/she would hesitate to take the risk of a potential furore caused by excluding certain groups from the vaccine. He/she would rather decide in favor of an equal distribution.

Although the final scenario expresses the strong wish to go to the prom, this wish represents a short-term desire and not a long-term self-interest. As one would have to expect serious consequences for violating quarantine rules, the clear self-interest according to ethical egoism still lies in staying home.

Moral Reasoning for Classical Natural Law Theory

Following the Classical Natural Law Theory, an act is moral if it is part of human nature. Typical characteristics of human nature were defined in section 2.2. Those include a peaceful life, procreation, not to murder, not to lie, and to pass on wisdom.

Accordingly, the parent should sacrifice himself/herself in order to save the daughter, because this supports procreation. The same argument applies to Scenario 4, where the 12-year old girl should be saved for the sake of procreation, even if this means killing the 95-year old woman.

Because biologically a man can impregnate multiple women, but a woman can be impregnated only once per year, women are considered more important

for procreation than men. Consequently the girls in Scenario 1 and 4 should be protected over the male patients. At the same time, a 75-year old man is expected to possess a high level of wisdom that he can pass on. Therefore, in Scenario 2, it should be the 75-year old man who receives the ventilator.

In Scenario 3, the personal interest to live a peaceful life dominates and therefore taking the ventilator for oneself is the right thing to do – even if that means a friend must die. Similarly one would want to protect one’s own tribe/family in Scenario 5 and rather give the ventilator to one’s father.

Sending one’s own children to school while a deadly disease is spreading (Scenario 6), contradicts with the natural interest of having a peaceful life for oneself and one’s family. Being home with one’s family is a very natural thing, even if that includes domestic violence or child abuse.

For Scenario 7 the perspective of the doer matters: One would never volunteer to be infected with a potentially deadly disease, as this violates one’s interest in living a peaceful life. At the same time, if there would be volunteers for the study, a scientist working on the vaccine would not hesitate conduct the study, as this could help to find a vaccine (acquisition of wisdom) and because a vaccine serves society (humans are social).

In exceptional situations, it is natural to protect one’s own tribe/family. The use of the tracking app described in Scenario 8 would serve this interest and is thus required by Classical Natural Law Theory.

Scenario 9 does not describe any personal relationships of the decision maker. So, naturally, the goal of the minister of health would be to save as many lives as possible. Accordingly, those people with other life-threatening conditions should be excluded from the vaccine as long as supplies are low.

For the final scenario with the prom, one would be advised to stay home, because this serves the interest of not risking to infect oneself (live a peaceful life), and of not jeopardizing others (humans are social). At the same time, a romantic night with one’s partner will most likely result in sex afterwards, what naturally leads to procreation. Although this Scenario is highly unlikely nowadays due to contraceptives, we still decided that our prototype will accept both answers for this dilemma, as the need to be with one’s partner represents a deeply human desire.

Summary

Figure 6 summarizes the answers of the five theories on the dilemmas that we presented in this section.

6 Evaluation

In section 1 we outlined requirements for our prototype framework to follow. These requirements were transformed into specific properties that the tool should inhibit in section 4. By developing a working prototype as described in section 5, and deploying it online, the tool was tested in practice. In this section, we

	Utilitarianism	Kantian Ethics	Aristotelianism	Ethical Egoism	Cl. Natural Law Theory
Scenario 1	B	Flip a coin	A	A or B	B
Scenario 2	B	Flip a coin	A	A	A
Scenario 3	B	Flip a coin	A	A or B	B
Scenario 4	A	B	B	B	A
Scenario 5	B	Flip a coin	Flip a coin	A or B	A
Scenario 6	A or B	A	B	A	A
Scenario 7	A	B	A	A	A
Scenario 8	A	A	A	A or B	A
Scenario 9	A	B	B	B	A
Scenario 10	B	A	B	B	A or B

Fig. 6: Summary of the moral theories' answers to the dilemmas

evaluate the extent to which the prototype meets our initial success criteria and property requirements, and describe experiences from the online experiment.

6.1 Property evaluation

Scenarios. The descriptions of the ten scenarios in our sample game range from three to six sentences in length. Keeping descriptions concise and precise was kept in mind when forming these. In addition, shorter descriptions will fit better on smaller screens, such as mobile phones. Each of the scenarios have two options to chose from, with some having the optional third 'flip a coin' option. These characteristics address the goals of presenting dilemmas in a clear and concise way, as well as having well-defined options to choose from. However, the prototype does not implement any form of quality control when new dilemma sets are added. The responsibility of forming precise and concise dilemmas relies on the user when creating them. This can potentially lead to games that do not meet these requirements being created in our framework. Results of such cases can include bad data and unmotivated participation in these games. The lenient approach taken in our prototype gives the users freedom to decide for themselves exactly how their scenarios are best presented. In the future, stricter approaches can be explored, for instance by implementing mechanics for quality control.

Flexibility. The ability for users to create their own scenarios represents one of the core mechanics of the prototype. This enables exploration of dilemmas of almost any kind, from any field, fulfilling the goal of flexibility for our framework. Adding to the flexibility, is providing users with the option of adding images to their scenarios. This gives them power to describe dilemmas in the manner they find the most meaningful. As we discuss in section 4.2, images can be sources of ambiguity, and should therefore only be added after careful consideration.

Ease of use. To ensure ease of use, users can add their dilemma sets through the "create game" component. The hope is that such component can lower the

threshold for usage of our tool. However, it is too early to conclude the actual effect this has in practice. In the future this component should be assessed for its usability. The prototype is also missing an option for more technically able users to upload a complete set of dilemmas as formally structured data.

Accessibility. Implemented as an openly available web application, our prototype is easy to access and share online, the prototype fulfills the 'sharing and access' property. In regards to 'platform and design', the prototype is designed to work with mobile devices, and does not require any additional installation on the user side to enter.

Player incentives Section 4.3 suggests five potential incentives for getting players invested in the game:

- Learning and curiosity
- Scenarios
- Gamification
- "Personality" test
- Sensation

Four out of five incentives were implemented to different degrees in the final prototype.

"Learning and curiosity" is an inherent incentive in the prototype due to the academic nature of the subject. It was however emphasised in the prototype as the creation of dilemmas needs an associated ethical theory attached.

The "Scenarios" part talks about having interesting scenarios for the dilemmas. Our own set of scenarios takes on COVID-19 related dilemmas, which is an obviously relevant topic and allows users to contemplate ethics in potentially real situations. The interestingness of other scenarios will be up to the individuals who invent them.

In the gamification incentive section we highlighted the importance of not letting gamification elements influence player decisions. One of the gamified parts of the prototype is in the scoreboard and social comparison. It motivates users to keep playing and also lets end-users get a sense of the crowdsourced data. Social comparison could be extended, for instance through a per-dilemma comparison in addition to the existing moral theory score. The comparison with the ethical theories also addresses the "'Personality' test" incentive. The "Sensation" incentive discusses using audio and visuals to enhance the gameplay experience. No audio or animation elements are in the prototype.

6.2 Online experiment

Over the course of 6 days, our prototype collected 105 responses from players of the included sample game. Although the tool was only made available online for

a limited amount of time before evaluation, we were able to gather some useful insight about its abilities and potential.

In terms of functionality as a data gathering tool, the results are promising. More than one thousand answers from 105 players in 10 different scenarios related to COVID-19 were successfully collected and compared with moral theories. Although the number of respondents is limited, we clearly see some patterns emerging about which moral theories players agree with at the highest rate in different dilemmas and the sample game as a whole. For instance, in Scenario 1, a significant majority of players would save the life of their daughter over themselves. Less than 3% of players align with Kantian ethics to flip a coin in Scenario 1. Other dilemmas, such as scenarios addressing whether patients' survival chance should affect who receives treatment, are more controversial. In the total set of dilemmas, player answers align with Ethical Egoism and Act Utilitarianism most frequently, while Kantian Ethics is most frequently disagreed with. The high agreement with Ethical Egoism is nearly inevitable for the scenarios in our sample game, as the theory agrees with two of the possible answers in four of the scenarios. The results demonstrate that our framework is successful in gathering opinions about moral dilemmas and compare them with moral theories.

It is still too early to determine the effects gamification has had for participation in our prototype. In order to assess how successful our approach is in increasing enjoyment and engagement among players, the development of popularity over time needs to be documented. In the best-case event, participants will recruit new players after having tried the game, and it will spread online in the likes of the Moral Machine. However, in the case that the game reaches low recruitment of new players, reasons might include the form dilemmas are presented, the dilemmas themselves, lack of player incentives, or a combination of these. It is likely that some player incentives require further development to engage more.

Over the course of the testing period, no new dilemma sets were added to the game. This was an expected result, considering the limited amount of time between deployment and evaluation.

7 Summary

In this paper, we investigated how to compare public opinion on ethical dilemmas against the answers suggested by moral theories. We created an openly available framework for a gamified collection of user responses on custom dilemmas: Researchers can deploy custom sets of dilemmas, collect responses and benchmark those responses against the positions of moral theories. Through the use of game interactions we demonstrate how design the process of data collection in a playful, engaging, and at the same time, unbiased way.

Our framework demonstrates one approach to reach an answer for our research question. Making it publicly available enables the the research community

to use our infrastructure for benchmarking users' opinion on ethical dilemmas against moral theories.

By deploying the tool online, we demonstrated how it can be used in work towards an answer for our research question. The results of our evaluation show comparison of public opinion in moral dilemmas to moral theories in practice. In addition to this, analysis of crowdsourced moral dilemma sets and moral theory decisions in these, our tool can contribute to identify the number of dilemma decisions it takes to identify a moral theory.

7.1 Future work

Extension of moral theories. For the purpose of demonstrating our framework, we have limited the scope of our prototype to only five theories. Of course, there exists a larger spectrum of other moral philosophy theories. Integrating more theories would allow to compare the user's opinion against a wider and more extreme set of moral principles and at the same time provide an enhanced and more gamified experience to the user. Below we present some of those other moral philosophy perspectives:

- *Divine Command Theory*: An act is moral if it follows the will of God [19]
- *Rawls Theory of Justice*: An act is moral if it supports the ideas of liberty and equality [44]
- *Situation Ethics*: An act is moral if it is the most loving thing to do in the particular situation [16]
- *Rule Utilitarianism*: An act is moral if it follows a moral principle whose general observance creates the most happiness [18]
- *Preference Utilitarianism*: An act is moral if it maximizes the preferences and interests of all involved parties [23]
- *Doctrine of double effect*: An act is moral, if it is intended to lead to a good end that will outweigh potential harm [34]
- *Ethics of Care*: An act is moral, if it is consistent with the actions of a caring character [20]

Additional testing. Our framework, together with a sample game was only available online for a short period of time before our evaluation. While this was sufficient as a proof-of-concept, additional time is needed in order to explore the true potential of our tool. For instance, the effects of gamification to increase player engagement and popularity could not be assessed. Additional time and a potential increase in popularity could prompt researchers to make use of the tool to make their own games. An answer on how public opinion on the right thing to do in a moral dilemma compares with moral theory can only be reached with our tool through popularity on both the researcher and the player side.

Game design and usability. Some of the player incentives identified in this paper, such as animation and audio features, were not implemented in our prototype. Gamification, such as social comparison can also be further developed in order

to engage more. Future versions should look into adding such features, as well as looking for other features with potential of enhancing the game experience. Furthermore, evaluations of the style and usability of the interface could help identify potential improvements in the tool.

References

1. Aristotle: *Nicomachean Ethics*. Cambridge Texts in the History of Philosophy, Cambridge University Press (2000). <https://doi.org/10.1017/CBO9780511802058>
2. Arrospide, A., Machón, M., Ramos-Goñi, J., Ibarrondo, O., Mar, J.: Inequalities in health-related quality of life according to age, gender, educational level, social class, body mass index and chronic diseases using the spanish value set for euroqol 5d-5l questionnaire. *Health and Quality of Life Outcomes* **17** (12 2019). <https://doi.org/10.1186/s12955-019-1134-9>
3. Awad, E., Dsouza, S., Bonnefon, J.F., Shariff, A., Rahwan, I.: Crowdsourcing moral machines. *Commun. ACM* **63**(3), 48–55 (Feb 2020). <https://doi.org/10.1145/3339904>, <https://doi.org/10.1145/3339904>
4. Awad, E., Dsouza, S., Kim, R., Schulz, J., Henrich, J., Shariff, A., Bonnefon, J.F., Rahwan, I.: The Moral Machine experiment. *Nature* **563**(7729), 59–64 (2018). <https://doi.org/10.1038/s41586-018-0637-6>, <http://dx.doi.org/10.1038/s41586-018-0637-6>
5. Baum, S.: Social choice ethics in artificial intelligence. *AI & Soc* **35**, 165–176 (2020). <https://doi.org/https://doi.org/10.1007/s00146-017-0760-1>
6. Benn, P.: *Ethics*. UCL Press (1998)
7. Bentham, J.: A comment on the commentaries and a fragment on government. In: Burns, J. H.; Hart, H.L.A. (ed.) *The Collected Works of Jeremy Bentham*. Oxford University Press (1997). <https://doi.org/10.1093/actrade/9780199553471.book.1>
8. Beauchamp, T.L., Childress, J.F.: *Principles of biomedical ethics*. Oxford University press pp. 10–25 (2013), http://www.allisonkrilethornton.com/wp-content/uploads/Medical_Ethics_Readings/BandC-Moral-Dilemmas.pdf
9. Burton, E., Goldsmith, J., Mattei, N.: Teaching AI ethics using science fiction. In: Walsh, T. (ed.) *Artificial Intelligence and Ethics, Papers from the 2015 AAAI Workshop*, Austin, Texas, USA, January 25, 2015. AAAI Workshops, vol. WS-15-02. AAAI Press (2015), <http://aaai.org/ocs/index.php/WS/AAAIW15/paper/view/10139>
10. Callaway, E.: Should scientists infect healthy people with the coronavirus to test vaccines? (Mar 2020), <https://www.nature.com/articles/d41586-020-00927-3>
11. Christen, M., Faller, F., Götz, U., Müller, C.: *Serious Moral Games. Analyzing and Engaging Moral Values Through Video Games*. Zürcher Hochschule der Künste (01 2012)
12. Clement, J.: Global digital population as of april 2020. <https://www.statista.com/statistics/617136/digital-population-worldwide>, accessed: 2020-04-29
13. Cushman, F., Young, L.: The psychology of dilemmas and the philosophy of morality. *Ethical Theory and Moral Practice* **12**, 9–24 (02 2009). <https://doi.org/10.1007/s10677-008-9145-3>
14. Danelli, F.: *Implementing Game Design in Gamification*, pp. 67–79. Springer International Publishing, Cham (2015), https://doi.org/10.1007/978-3-319-10208-5_4

15. Feyisetan, O., Simperl, E., Van Kleek, M., Shadbolt, N.: Improving paid micro-tasks through gamification and adaptive furtherance incentives. In: Proceedings of the 24th International Conference on World Wide Web. p. 333–343. WWW '15, International World Wide Web Conferences Steering Committee, Republic and Canton of Geneva, CHE (2015). <https://doi.org/10.1145/2736277.2741639>, <https://doi.org/10.1145/2736277.2741639>
16. Fletcher, J.: Situation Ethics: The New Morality. Library of theological ethics, Westminster John Knox Press (1997), <https://books.google.no/books?id=Y4759nkMFq0C>
17. Gardiner, P.: A virtue ethics approach to moral dilemmas in medicine. *Journal of medical ethics* **29**(5), 297–302 (October 2003). <https://doi.org/10.1136/jme.29.5.297>, <https://europepmc.org/articles/PMC1733793>
18. Garner, R., Rosen, B.: Moral Philosophy: A Systematic Introduction to Normative Ethics and Meta-ethics. Macmillan (1967), <https://books.google.no/books?id=7vbWAAAAAAAJ>
19. Geirsson, H., Holmgren, M.: Ethical Theory: A Concise Anthology - Third Edition. Broadview Press (2018), <https://books.google.no/books?id=NP0qswEACAAJ>
20. Gilligan, C.: Moral orientation and moral development, pp. 19–33. Rowman and Littlefield Publishers (1987)
21. Godin, M.: Covid-19 patient's family break quarantine for school dance (Mar 2020), <https://time.com/5799412/missouri-coronavirus-patient-family-break-quarantine/>
22. Hamilton, I.A.: Edward snowden says covid-19 could give governments invasive new data-collection powers that could last long after the pandemic (Mar 2020), <https://www.businessinsider.com/edward-snowden-coronavirus-surveillance-new-powers-2020-3?r=US&IR=T>
23. Hare, R.: Moral Thinking: Its Levels, Method, and Point. Clarendon Press (1981), <https://books.google.no/books?id=SverDwAAQBAJ>
24. Hursthouse, R., Pettigrove, G.: Virtue ethics. In: Zalta, E.N. (ed.) *The Stanford Encyclopedia of Philosophy*. Metaphysics Research Lab, Stanford University, winter 2018 edn. (2018)
25. Høydal, H.F.: Slår alarm: – har sett overgrepststeder krasje på grunn av for stor pågang. VG (Apr 2020), <https://www.vg.no/nyheter/innenriks/i/vQ76o1/slaar-alarm-har-sett-overgrepststeder-krasje-paa-grunn-av-for-stor-paagang>
26. Johnson, R.N.: Was kant a virtue ethicist? Kant's Ethics of Virtue pp. 61–76 (2008)
27. Joinson, A.N., McKenna, K.Y.A., Postmes, T., Reips, U.D., Buchanan, T.: Personality testing on the internet what we know, and what we do not (09 2012), <https://www.oxfordhandbooks.com/view/10.1093/oxfordhb/9780199561803.001.0001/oxfordhb-9780199561803-e-028>
28. Kant, I.: *Groundwork for the Metaphysics of Morals*. Oxford University Press (2002 [1785])
29. Kim, S., Wall, I., Stanczyk, A., De Vries, R.: Assessing the public's views in research ethics controversies: Deliberative democracy and bioethics as natural allies. *Journal of empirical research on human research ethics : JERHRE* **4**, 3–16 (12 2009). <https://doi.org/10.1525/jer.2009.4.4.3>
30. "Kvalnes, Ø.: "Moral Dilemmas", pp. "11–19". "Springer International Publishing", "Cham" ("2019"), https://doi.org/10.1007/978-3-030-15191-1_2
31. Law, E., Yin, M., Goh, J., Chen, K., Terry, M.A., Gajos, K.Z.: Curiosity killed the cat, but makes crowdwork better. In: Proceedings of

- the 2016 CHI Conference on Human Factors in Computing Systems. p. 4098–4110. CHI '16, Association for Computing Machinery, New York, NY, USA (2016). <https://doi.org/10.1145/2858036.2858144>, <https://doi.org/10.1145/2858036.2858144>
32. Lord, J., Willis, S., Eatock, J., Tappenden, P., Trapero Bertran, M., Miners, A., Crossan, C., Westby, M., Anagnostou, A., Taylor, S., Mavranzeouli, I., Wonderling, D., Alderson, P., Ruiz, F.: Economic modelling of diagnostic and treatment pathways in national institute for health and care excellence clinical guidelines: The modelling algorithm pathways in guidelines (mapguide) project. *Health technology assessment (Winchester, England)* **17** (12 2013). <https://doi.org/10.3310/hta17580>
 33. Mavandadi, S., D.S.F.S.Y.F.S.U.Y.O.P.S.N.K..O.A.: Distributed medical image analysis and diagnosis through crowd-sourced games: a malaria case study. *PloS one* **7** (2012). <https://doi.org/https://doi.org/10.1371/journal.pone.0037245>
 34. McIntyre, A.: Doctrine of double effect. In: Zalta, E.N. (ed.) *The Stanford Encyclopedia of Philosophy*. Metaphysics Research Lab, Stanford University, spring 2019 edn. (2019)
 35. MK., F.: The use of crowdsourcing and the role of game mechanics in identifying erroneous disease burden estimates (2013)
 36. Morschheuser, B., Hamari, J., Koivisto, J., Maedche, A.: Gamified crowdsourcing: Conceptualization, literature review, and future agenda. *International Journal of Human-Computer Studies* **106**, 26–43 (2017). <https://doi.org/https://doi.org/10.1016/j.ijhcs.2017.04.005>, <http://www.sciencedirect.com/science/article/pii/S1071581917300642>
 37. Muladal, A., Ertesvåg, F.: Barnas minister om coronastengte skoler og barnehager: Frykter omsorgssvikt. VG (Apr 2020), <https://www.vg.no/nyheter/innenriks/i/LAE1Q1/barnas-minister-om-corona-stengte-skoler-og-barnehager-frykter-omsorgssvikt>
 38. Myrvang, S.E.: Bekymret for vold i norske hjem under coronapandemien. VG (Apr 2020), <https://www.vg.no/nyheter/innenriks/i/6jAkq3/bekymret-for-vold-i-norske-hjem-under-coronapandemien>
 39. Nations, T.U.: Universal Declaration of Human Rights. The United Nations (1948)
 40. Paul-Ehrlich-Institut: Pressemitteilungen - erste klinische prüfung eines covid-19-impfstoffs in deutschland genehmigt (08 2020), <https://www.pei.de/DE/newsroom/pm/jahr/2020/08-erste-klinische-pruefung-sars-cov-2-impfstoff-in-deutschland.html>
 41. Prandi, C., Salomoni, P., Roccetti, M., Nisi, V., Nunes, N.J.: Walking with geozombie: A pervasive game to engage people in urban crowdsourcing. In: 2016 International Conference on Computing, Networking and Communications (ICNC). pp. 1–5 (2016)
 42. Rachels, J.: *The Elements of Moral Philosophy*. The heritage series in philosophy, McGraw-Hill College (1999), <https://books.google.no/books?id=fgfXAAAAAAAJ>
 43. Rahwan, I.: Society-in-the-loop: programming the algorithmic social contract. *Ethics Inf Technol* **20**, 5–14 (2018). <https://doi.org/https://doi.org/10.1007/s10676-017-9430-8>
 44. Rawls, J.: *A Theory of Justice*. Harvard University Press (2009), <https://books.google.no/books?id=kvpby7HtAe0C>
 45. Richter, G., Raban, D.R., Rafaei, S.: Studying gamification: The effect of rewards and incentives on motivation. In: T. Reiners, L.W. (ed.) *Studying Gamification: The Effect of Rewards and Incentives on Motivation*, chap. 2, pp. 21–46. Springer International Publishing Switzerland (2015)

46. Rommen, H.: *The Natural Law. European political thought*, Arno Press (1947), <https://books.google.no/books?id=aod4BR6xTE8C>
47. Rosenlund-Hauglid, S.: Norsk politi bekymret: Nedgang i antall anmeldelser av vold i nære relasjoner. VG (Apr 2020), <https://www.vg.no/nyheter/innenriks/i/4qeWzV/norsk-politi-bekymret-nedgang-i-antall-anmeldelser-av-vold-i-naere-relasjoner>
48. Ryan, R.M., Deci, E.L.: Intrinsic and extrinsic motivations: Classic definitions and new directions. *Contemporary Educational Psychology* **25**(1), 54–67 (2000). <https://doi.org/https://doi.org/10.1006/ceps.1999.1020>
49. Sanders, S.M.: Is egoism morally defensible? *Philosophia* **18**, 191–209 (1988)
50. Sidgwick, H.: *The Methods of Ethics*. Thoemmes Press (1874)
51. Singer E, C.M.: Do incentives exert undue influence on survey participation? experimental evidence. *J Empir Res Hum Res Ethics* **3**(3), 49–56 (2008). <https://doi.org/https://doi.org/10.1525/jer.2008.3.3.49>
52. Sinnott-Armstrong, W.: Consequentialism. In: Zalta, E.N. (ed.) *The Stanford Encyclopedia of Philosophy*. Metaphysics Research Lab, Stanford University, summer 2019 edn. (2019)
53. Statcounter: Desktop vs mobile vs tablet market share worldwide. <https://gs.statcounter.com/platform-market-share/desktop-mobile-tablet/worldwide>, accessed: 2020-04-29
54. Strauss, M.: A philosopher weighs in on whether or not you should kill baby hitler (2015), <https://www.inverse.com/article/7570-a-philosopher-weighs-in-on-whether-or-not-you-should-kill-baby-hitler>, [Online; posted 29-October-215]
55. Susumpow, P., Pansuwan, P., Sajda, N., Crawley, A.: Participatory disease detection through digital volunteerism: how the doctorme application aims to capture data for faster disease detection in thailand. In: WWW '14 Companion: Proceedings of the 23rd International Conference on World Wide Web. pp. 663–666 (04 2014). <https://doi.org/10.1145/2567948.2579273>
56. Thomas, A.: *The "Summa theologiae" of St. Thomas Aquinas*. Washburne, ltd. (1920), <https://search.library.wisc.edu/catalog/999609076002121>, the Supplement was compiled probably by Fra Rainaldo da Piperno; First published 1912-1925; Also available on the Internet.
57. Varden, H.: Kant and lying to the murderer at the door . . . one more time: Kant's legal philosophy and lies to murderers and nazis. *Journal of Social Philosophy* **41**(4), 403–421 (2010). <https://doi.org/10.1111/j.1467-9833.2010.01507.x>, <https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1467-9833.2010.01507.x>
58. Wood, A.W.: *Kant's Ethical Thought. Modern European Philosophy*, Cambridge University Press (1999). <https://doi.org/10.1017/CBO9781139173254>
59. Yiannakoulis, N., Gordon, J.N., Darlington, J.C.: The decision game: a serious game approach to understanding environmental risk management decisions. *Journal of Risk Research* **23**(1), 81–94 (2020). <https://doi.org/10.1080/13669877.2018.1517379>
60. Young, G.: Virtually real emotions and the paradox of fiction: Implications for the use of virtual environments in psychological research. *Philosophical Psychology* **23**, 1–21 (2 2010). <https://doi.org/10.1080/09515080903532274>