Eclectic Ethical Decision Making For Autonomous Vehicles

Pranay Chowdary Jasti
xxp12@txstate.edu
Texas State University
San Marcos, USA

Henry Griffith
hgriffith5@alamo.edu
San Antonio College, San Antonio, Texas

Heena Rathore
heena.rathore@txstate.edu
Texas State University, San Marcos, Texas

ABSTRACT

Ensuring the safety of autonomous vehicles (AV) remains a critical challenge, particularly when navigating morally ambiguous scenarios. Enhancements in this area are crucial for increasing trust and boosting the adoption of AV. Existing state-of-the-art solutions, such as predictive controllers and the ethical valence theory, prioritize decisions based on preset priorities. In contrast, reinforcement learning-based models can be implemented to mimic actions based on widely established ethical theories. This paper proposes a novel approach to address the problem of random credence generation in moral uncertainty for AVs. We combine the simplex algorithm with the Dempster-Shafer (DS) combination technique to integrate the established ethical theories. We model the reward structure using a linear minimization approach. The simplex algorithm solves this model effectively, and its results are used as evidence for the Dempster-Shafer combination technique, which generates the plausibility of theories and serves as credence in our framework. Our approach demonstrates superior performance compared to existing models, offering a promising solution to the ethical challenges of autonomous driving.

CCS CONCEPTS

• **Computing methodologies** → **Multi-agent systems**; • **Theory of computation** → **Sequential decision making**; **Linear programming**; *Active learning*; • **Applied computing** → **Decision analysis**.

KEYWORDS

autonomous vehicles, moral uncertainty, ethical theories, credence, reinforcement learning

1 INTRODUCTION

Autonomous vehicles (AVs) have shown promise in exceeding human performance in certain areas [1], particularly in reducing traffic congestion and improving safety. Despite these advantages, AVs still face challenges in gaining public trust and widespread adoption [2]. While advancements have been made in navigation and security, a significant area of ongoing research is decision-making in ethically ambiguous situations [3]. There is ongoing debate about the most suitable ethical framework for AVs to navigate moral dilemmas [4]. There is also a need to explore eclectic strategies for integrating multiple ethical theories within the decision-making process.

1.1 Motivation

While numerous ethical theories exist, such as utilitarianism, deontology, and virtue ethics, philosophers continuously debate on which theory holds the most weight [5]. Developers tackling ethical decision-making in AVs must navigate this complex landscape of differing viewpoints. Current attempts include the ethical valence theory [6], which assigns value to outcomes based on the number of individuals affected. Additionally, the lexicographic optimization technique prioritizes certain moral principles over others in specific situations [7]. These methods, often coupled with a model predictive controller to optimize actions, aim to navigate complex ethical scenarios. However,they face limitations due to their static nature. Traffic environments are dynamic and require robust techniques to address unforeseen situations. Research suggests that implementing RL in AVs offers significant advantages [8]. Recent advancements in reinforcement learning (RL) offer promising alternatives [4]. RL's ability to adapt to dynamic scenarios and unforeseen circumstances makes it a potential solution for tackling the ethical challenges of autonomous driving.

1.2 Problem Statement

The authors in [4] took a significant step in merging RL with moral uncertainty by addressing most of the concepts outlined in the MacAskill's book [9]. Their approach hinges on selecting actions based on the number of theories, the associated action weights, and the credence (degree of belief in a theory). They introduced two primary theories: utilitarianism and deontology. While this is a commendable initial step, it is crucial to consider incorporating multiple theories, given the existence of numerous widely accepted ethical frameworks. The simulations were conducted in a grid world, featuring four scenarios where an uncontrollable trolley is faced with a moral dilemma: either hitting a large number of people or diverting to a side track and causing harm to fewer individuals. While these scenarios are simple to start with, they must be adapted to real-time traffic situations, which often present unforeseen challenges. One of the key challenges addressed in their work is the generation of credence. Currently, credence in the theories is randomly generated. However, when action selection depends on credence, random generation may lead to inconsistent results. Therefore, there is a need to refine the generation of credence based on the underlying theories.

1.3 Contribution

In this paper, we contribute to addressing the problem of random credence in moral uncertainty [4] by proposing a novel approach that combines the simplex algorithm with the DS combination technique to solve the minimization problem.

- We use simplex algorithm as it is widely used in various fields and applications due to its efficiency and effectiveness in finding optimal solutions for linear programming problems. In this paper, our objective is to minimize (Eq. 6) the total harm caused by autonomous systems in ethical dilemmas, where the optimal decision is paramount. The simplex algorithm's ability to quickly iterate through possible solutions and converge to an optimal or near-optimal solution is crucial in navigating the ethical landscapes that autonomous systems face. The scalability of the simplex algorithm ensures that it can handle increasingly complex scenarios as the machine's capabilities and the scope of its ethical considerations expand [10].
- Similarly, DS theory [11] offers a powerful tool for handling uncertainty and making decisions in situations. Since credence value is randomly generated it has uncertainty in the outcomes of the decision making. DS theory allows for the representation of partial ignorance or uncertainty by assigning belief masses to sets of outcomes. It allows the integration of diverse information to arrive at a more informed decision with the help of its systematic way of combining evidence from multiple sources.

Our method involves modeling a minimization linear problem based on moral theories, with appropriate constraints. We then employ the simplex algorithm [12], known for its effectiveness in solving linear programming problems, to solve this model. The results obtained from the simplex algorithm serve as evidence for the DS combination technique. This technique generates the plausibility of the theories, which in turn is used as the credence in our framework. Our approach has demonstrated superior performance compared to the random credence generation method used in existing models.

2 RELATED WORK

Bogosian [13] argued that intelligent vehicles dealing with moral philosophy should accept uncertainty regarding morality while addressing the intractable nature of disagreement among theories. The author advocated for adopting moral uncertainty, similar to the voting problem proposed by William MacAskill. Along with discussing the reason behind employing a maximum expected choiceworthiness function, the author also described the necessity of an ordinal ranking and the impact of credence on a moral machine. He developed a computational framework to address the challenge of disagreement among moral philosophers regarding choosing the correct moral theory. The author also discussed the commercial viability of these moral machines. However, the practical challenges in the deployment of the model are not addressed as the author has not implemented the framework proposed, which limits the applicability and real-world impact of the research.

Authors in [14] conducted two experiments, involving scenarios where AVs had to decide between staying in the lane or swerving. Each scenario was characterized by potential collisions and varying probabilities. The study found that subjects consistently preferred the default action of staying in the lane, even when it did not minimize expected losses. Moral acceptability of the default option was higher, especially under uncertainty. The study highlights the importance of understanding moral judgments under risk and uncertainty to develop socially acceptable policies for AVs in critical conditions. However, it is important to note that the paper does not delve into the direct implementation of moral uncertainty in automated vehicles, leaving a gap in practical application.

Hong et al. [7] investigated the development of a predictive control framework for ethical decision-making in autonomous driving using rational ethics. The study aimed to utilize a lexicographic optimization technique along with a model predictive controller to solve ethical decision-making problems by establishing a priority order. The authors employed a simulation-based methodology, using Prescan for simulation and developing the Lexicographic Optimization Model Predictive Controller (LO-MPC) in MATLAB/Simulink. They generated an artificial potential field representing vehicles, road users, obstacles, and road boundaries, calculated potential crash severity using this field, and determined obstacle priority. The LO-MPC was then used to solve ethical decision-making based on these priorities.The results of the simulations were conducted in three different scenarios: Human vs. Vehicles (a moral dilemma), Road Regulation vs. Traffic Accidents, and Road Regulation vs. Unexpected Animal encounters. In all scenarios, the LO-MPC successfully minimized the total harm caused by the vehicle compared to the pathfinding-MPC approach. Despite their effectiveness, these methods are limited by their static nature, which may not adequately address the dynamic nature of traffic environments and the need for robust techniques to handle unforeseen situations.

The authors in [6] explored applying Ethical Valence Theory (EVT) to solve ethical decision-making in AVs. They framed AV decision-making as a form of claim mitigation, where different road users hold varying moral claims about the vehicle's behavior. The authors proposed using Markov Decision Processes (MDP) to evaluate all possible harms and select actions that minimize harm, prioritizing the safety of road users and passengers of the AV while considering traffic regulations. Claims are formulated as harms and valence, with valence indicating the degree of social acceptability attached to the claims. Two moral theories, Risk-Averse Altruism and Threshold Egoism, were applied. Risk-Averse Altruism prioritizes protecting road users with the highest valence unless the AV passenger's

risk is severe, while Threshold Egoism protects the AV passenger unless the high valence road user is severely at risk. The authors tested the framework in a dilemma where the AV must decide between saving a pedestrian and avoiding a collision with a peer vehicle. The results showed the choice of operative moral theory influenced the decision: Risk-Averse Altruism led to colliding with a wall, while Threshold Egoism resulted in colliding with the pedestrian. One limitation is that ethical decisions depend on the operative moral theory used to select the action, and the use of a possible valence hierarchy is not universally agreed upon.

The authors in [4] explored the intersection of moral philosophy and reinforcement learning (RL) to address moral uncertainty in artificial intelligence (AI) systems. Two voting systems, Nash voting and Variance voting, are introduced to guide AI agents in ethically complex decision-making using moral theories such as deontologist and utilitarian. The paper proposes a method for updating moral theories using credence values without retraining the policy. Experimental results in grid world environments mainly consists of classic, double, guard, and doomsday. However, this approach faces limitations, such as the difficulty of assigning and updating credence values for each theory and potential bias towards dominant theories in Nash voting. Additionally, the paper's use of simple grid world environments may not fully capture the complexities of real-world moral dilemmas.

3 PROPOSED WORK

3.1 Overview of Random Credence

Figure 1 represents the framework developed by authors in [4]. The initial vector contains the credence assigned the theories. The degree of belief that the agent has in theory i is considered as the level of credence in theory C_i. The input vector, along with the initial vector, contains the state of the environment and the action space of the agent at each step. The authors replaced the standard reward function with a cardinal choice worthiness function $W_i(s, a, s')$, which is analogous to the reward function. Authors defined the function $Q_i(s, a)$ (Eq. 1)as the expected sum of future choice-worthiness discounted over time for theory i, starting from state s and taking action a, while all future actions follow the current policy.

$$Q_i(s,a) = E\left[\sum_{t=0}^{\infty} \gamma^t W_i(s_t, a_t, s_{t+1}) | s_0 = s, a_0 = a)\right] \quad (1)$$

The reward is calculated as the product of the credence in the theory and W_i at every step. RL agent is chosen depending on the voting method selected. The authors used a proximal policy optimization (PPO) agent [15] to train the agent with the Nash voting method and a SARSA agent [16] to train the agent with the variance voting method. The agent then selects an action using the optimal policy, which uses the input vector and the reward value. This action governs the trolley's direction, which is directly responsible for the harm caused to the people on the track.

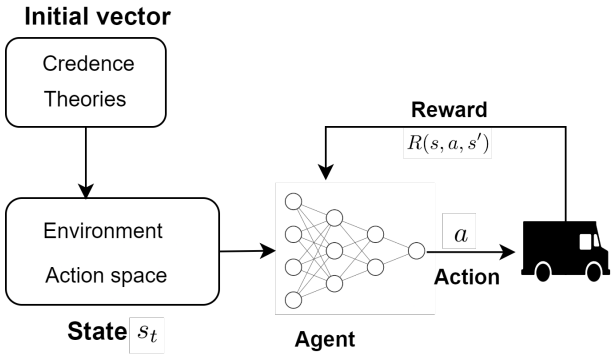


Figure 1: System Model

We have approached the problem of credence in two ways.

- Dempster Credence: Here, we combine the DS technique with random credence.
- Simplex Credence: Here, we have modeled the credence in theories using simplex algorithm. We have constructed a linear programming problem from the theories, starting by creating objective functions using the actions from theories, modeling the constraints with appropriate bounds, solving the objective function using the simplex algorithm to find optimal solution, generated mass functions using optimal solution and optimal values from simplex algorithm and used the plausibility score as the credence in theories.

3.2 Dempster Credence

As an initial step to strengthen the power of credence and stabilize it, we have combined the random credence with DS technique by using the credence generated randomly as the evidence to the DS combination to get the plausibility score of the theories.

3.2.1 Overview of Dempster-Shafer theory. The Dempster-Shafer (DS) theory of evidence is a mathematical framework that quantifies belief in statements by integrating independent evidence from various sources. It deals with uncertainty by assigning levels of belief to subsets of potential events, which differs from traditional probability theory. This theory operates under the assumption of inherent ignorance, which results in uncertainty, and employs the DS rule to merge belief functions. Below components together form the basis of the DS theory of evidence, providing a framework for reasoning under uncertainty and combining evidence from multiple sources.

- The frame of discernment (Θ) is a set of all possible outcomes, with each outcome representing a mutually exclusive, discretized value (utilitarian and deontology).

- The power set P(Θ) of Θ is a set of all subsets of (Θ), including individual elements, representing the DS frame of (Θ).
- Evidence consists of events or symptoms, where each evidence maps to a single hypothesis or a set of hypotheses. Different levels of evidence are considered, such as credence generated randomly.
- Mass function (m-value): The mass function relates to the weights of elements in P(Θ), indicating the belief assigned to each subset of Θ. It sums to 1 across all subsets, with the lower bound used as the belief function and the upper bound used to calculate the plausibility function.
- Plausibility function (Pl): The plausibility function determines the upper bound of the interval by summing the mass functions of subsets B that intersect with a given subset (A), indicating the degree of plausibility assigned to each subset of Θ. $(B \cap A \neq \phi)$, $Pl(A) : 2^{\Theta} \rightarrow [0, 1]$ [11].

$$Pl(A) = \sum_{B \cap A \neq \phi} m(B) \tag{2}$$

3.2.2 Dempster credence calculation. We use the randomly generated credence (C_r) as a base to generate the mass function $m(t_i)$ of theories. The system treats the credence of theory as the likelihood of the theory being superior while its complement (1- credence of the theory) is the likelihood of the theory being inferior. The system generates mass function tuple for both theories using the credence generated randomly. The mass function tuple contains $m(i)$ (the mass function of theory 1), $m(i')$ (the mass function of theory 2).

$$m_T(i) = C_{r_i} \tag{3}$$

$$m_T(i') = 1 - C_{r_i'} \tag{4}$$

In Eq. 3, i is the selected theory and in Eq. 4 i' is the other theory.

The plausibility score $Pl_d(T_i)$ is calculated using the mass functions generated using Eqs.3 and 4. This value is used as the credence for the theories in the simulation.

3.3 Simplex Credence

3.3.1 Simplex Model. The system architecture (See Figure 2) involves modeling a linear programming problem based on moral theories. The problem is then solved to find the optimal solution using the simplex algorithm. Next, the plausibility values of the theories are determined using the optimal solution as evidences with the DS theory. These plausibility scores are used as the credence of the theories. We call the credence generated in our framework as simplex credence to differentiate from existing credence. The credence in the initial vector of Figure 1 is replaced by simplex credence. In the framework, the theories (t_i) are formulated in the form of a dictionary with keys as the action (a_t) and the values as the rewards (r_a) associated with the actions. Equation 5 showcases the structural representation of theory.

$$t_i = \{a_t : r_a\} \tag{5}$$

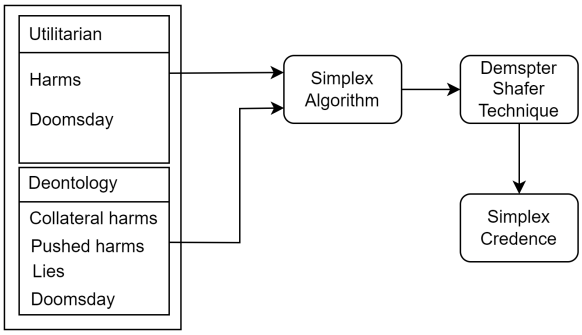


Figure 2: Simplex Model

Objective function generation. The objective function $O(f)$ is defined as the sum of the unique actions a_i present in both theories.

$$O(f) = \sum_{i \in t} a_i \tag{6}$$

In the system, there are two theories: utilitarianism and deontology. These theories are defined such that every action has a corresponding weight, which is used to calculate the reward at the end. Deontology considers all actions and their associated weights, focusing on the actions and the harm caused by the actions of an agent. On the other hand, utilitarianism only has two values: harms and doomsday. For utilitarianism, the primary goal is to maximize happiness, without consideration for how it is achieved. Actual representation of theories in the simulation is given below:

- Deontologist: { pushed harms : -4, collateral harms : -1, lies : -0.5, doomsday : -10 }.
- Utilitarian: { harms : -1, doomsday : -300 }

From the above representation of the deontologist theory, "pushed harms" is associated with the action "push" in a double and guard environment where the agent has a choice to push the fat man onto the track. When the agent chooses to switch, the "collateral harms" are used to calculate the penalty for the agent. Similarly, "lies" is associated with the action "lie" in the guard environment. The "harms" in the utilitarian theory is a generic penalty used to calculate the total reward for the agent after its choice. The value of "doomsday" in both theories is used to calculate the penalty for the agent when the agent's preference is doomsday in the doomsday environment.

From the above theories, we get pushed harms, collateral harms, lies, doomsday and harms. The sum of the above unique keys forms the objective function which is later solved using simplex algorithm using the constraints.

Constraints generation. The constraints (C_t) are generated as sum of product of actions a_i and the reward value r_a associated with the actions in a given theory.

$$C_{t_i} = \sum a_i r_a < r_{doomsday_i} \tag{7}$$

The constraints C_{t_i} reflect the theories in the form of an equation. Regardless of the theory, the overall aim here is the safety of the humans involved. To illustrate this, the constraints are designed to prevent the selection of "doomsday" as an action. This is achieved by limiting the value of a constraint to be less than the value of the reward for "doomsday," as prescribed by the theory. We solve the objective function $O(f)$ with constraints C_{t_i} as a minimization problem using the simplex algorithm. The Simplex algorithm halts when an optimal solution $Os(O(f))$ is found. For this, we have used linprog function from the scipy package [17] which uses tableau-based simplex method.

Simplex algorithm finds the optimal values of the variables Ov_a in objective function $O(f)$. We use the optimal values to generate the mass functions of the theories.

3.3.2 Dempster - Shafer Technique.

Mass function generation using optimal value. The optimal values generated using the simplex algorithm serve as a base to generate mass functions $m(f)$ supporting the theories which are used to calculate the plausibility values of the theories. For each action value in the theory, the system generates mass functions consisting of evidence supporting theory 1 ($m_a^i(1)$), and evidence supporting theory 2 ($m_a^i(2)$).

$$m_a^i(1) = \begin{cases} \frac{Ov_a}{Os(O(f))} & If \quad \frac{Ov_a}{Os(O(f))} > 0.2 \\ 0.1 & Else \end{cases} \tag{8}$$

$$m_a^i(2) = 1 - m_a^i(1) \tag{9}$$

Plausibility calculation. The generated mass functions using Eq. 8 and Eq. 9 are combined using the DS rule of combination [11] which is later used to calculate the plausibility score of the theories.

Plausibility as credence. The plausibility scores of theories generated from the previous step are used as the credence of the theories, denoted as SC_i. We refer to this credence as simplex credence. This step ensures that the credence of the theories is generated from the preferences of theories instead of random generation. We believe that the incorporation of theories in credence calculation will play a crucial role in the agent's preference of actions and will enhance its performance.

Incorporating credence into the existing framework. To understand the power of credence, it is helpful to provide an overview of the various contexts in which credence is used. The RL agent at the switch follows two different voting methods: Nash voting and variance voting. The authors incorporated the PPO RL framework with Nash voting, and the SARSA framework is used with variance voting. At each step, the input to the RL agent contains the credence in theory C_i, the state of the environment s_t, and the reward from the previous step $R(s_{t-1})$, which is calculated using the values in the theories. The action space comprises particular actions associated with the selected environment. The agent is trained to select an action a using the optimal policy π. The selected action is multiplied by the credence, and the value is considered as votes $V_a i$ as referred in Eq. 10.

$$V_a i = a_i \times C_i \tag{10}$$

The action associated with the highest votes is selected as the final action in both theories. In addition to the selection of action, the credence also plays a crucial role in the calculation of the reward for the agent. The authors used a simple reward structure in Nash voting Eq. 11, which is the product of the credence in theory and the choiceworthiness function $W_i(s, a, s')$.

$$R(s, a, s') = \sum_i C_i W_i(s, a, s') \tag{11}$$

In variance voting, the credence is replaced by the variance-normalized credence Eq. 12.

$$R(s, a, s') = \sum_i w_i W_i(s, a, s') \tag{12}$$

where:

$$w_i = C_i / (\sqrt{\sigma_i^2} + \epsilon)(\mu_i(s)) \tag{13}$$

where ϵ is a small constant (10^{-6} in our experiments). In order to select the values of parameters of the affine transformation of Q_i instead of direct vote theory i the variance voting suggests that the Q_i function should be normalized by the expected value $E_{s\sim S}$ of its variance across timesteps.

$$\sigma_i^2 = E_{s\sim S}\left[\frac{1}{k}\sum_a (Q_i(s, a) - \mu_i(s))^2\right] \tag{14}$$

where k is number of actions in a discrete action space :

$$\mu_i(s) = \frac{1}{k}\sum_a Q_i(s, a) \tag{15}$$

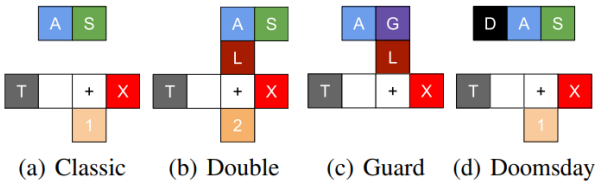
The random credence C_i in the equations Eq. 10, 11 and 13 is replaced with simplex credence SC_i which is generated in the previous steps in the simulation.

4 SIMULATION SETUP AND RESULTS

4.1 Datasets

We utilize an open-source framework, [18], that implements moral uncertainty using reinforcement learning as the basis for our work, which extends this existing framework. The simulation involves a trolley that, without intervention, will hit people standing on the main track. An agent, positioned at a switch, faces a moral dilemma: whether to divert the trolley to a side track with fewer people. There are four environments: classic, guard, double, and doomsday. In the classic scenario, (See Figure 3(a)), the agent must decide between pulling the switch to save

fewer people or doing nothing and letting more people be harmed. The double environment in Figure 3(b) introduces a fat man on the side track, adding complexity to the decision-making process. The agent has three options: do nothing, push the fat man onto the track, or pull the switch to divert the trolley. In the guard environment (See Figure 3(c)), the agent faces a similar moral dilemma, but with the added option of lying to the guard to push a heavy person onto the track, potentially saving more lives but involving dishonesty. In the doomsday environment (Figure 3(d)), the agent is given the additional option of intentionally choosing to kill a large number of people, contrasting sharply with the other options aiming to minimize harm.



(a) Classic (b) Double (c) Guard (d) Doomsday

Figure 3: Trolley environment

4.2 Simulation Environment

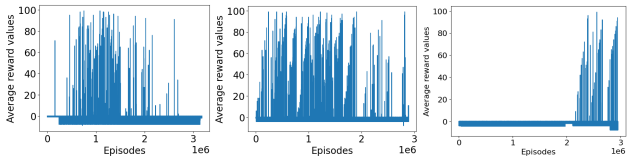
The simulations are carried out on a NVIDIA dual GPU workstation AMD(R) Ryzen threadripper pro 3955wx 16 cores x32 with 128 GB RAM on a Ubuntu 20.04.5 LTS. We ran the simulations for 10^7 episodes with random, Dempster and simplex credence. For visualization purposes we had represented the plots for 3×10^6 episodes of random, Dempster and simplex credence.

4.3 Results

We have compared the performance of random credence (the work done in [4]) with Dempster Credence and Simplex credence. We have presented two types of results to compare the performance of the agent in different environments and learning rates using the reward gained by theories with different credence methods.

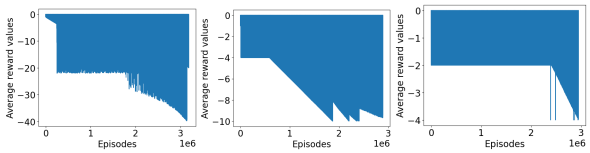
4.3.1 Reward Analysis. The results show the average reward gained by the RL agents using random credence, Dempster credence, and simplex credence. The reward plots contain the rewards gained by the agents in different environments such as classic, double, guard, and doomsday using deontology and utilitarian theories. The results from $0 - 0.75 \times 10^6$ episodes represent the reward gained by the agent using a specified theory in the classic environment. Results from $0.75 \times 10^6 - 1.5 \times 10^6$ episodes illustrate the reward gained by the agent in the double environment. The reward gained by the agent in the guard environment is represented in the graph during $1.5 \times 10^6 - 2.25 \times 10^6$ episodes, and the reward gained by the agent during $2.25 \times 10^6 - 3.0 \times 10^6$ episodes represents the agent's performance in the doomsday environment.

Deontologist reward. Figures 4(a), 4(b) and 4(c) represent average reward values of agent with random credence, Dempster credence and simplex credence using deontology theory. From Figure 4(a), we observe that the plot is left-skewed and denser from 0.5×10^6 to 1.5×10^6 episodes. The agent was able to achieve an average reward value of more than 50 for over 50% of the time. In Figure 4(b), representing Dempster credence, we see that the reward is distributed evenly for approximately 2.0×10^6 episodes and moderately distributed from $2.0 \times 10^6 - 3.0 \times 10^6$ episodes. The agent was able to achieve an average reward of more than 40. From the simplex credence plot in Figure 4(c), we can see that it is right-skewed, and the average reward value is below 80 most of the time.



(a) (b) (c)

Figure 4: (a) Reward with random credence, (b) Reward with Dempster credence, (c) Reward with simplex credence



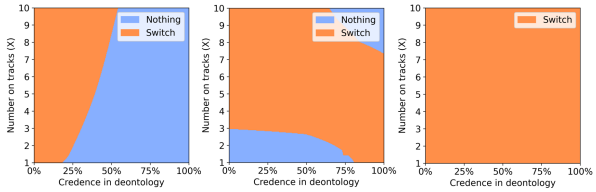
(a) (b) (c)

Figure 5: (a) Reward with random credence, (b) Reward with Dempster credence, (c) Reward with simplex credence

Utilitarian reward. Figures 5(a), 5(b) and 5(c) represent the average reward values of agent with random credence, Dempster credence and simplex credence using utilitarian theory. From the average reward value plots of agent using utilitarian theory, we can observe that the plots are right skewed in nature. From Figure 5(a), we observe that the minimum average reward value is -40. The agent maintained a constant -20 for approximately 1.75×10^6 episodes from 0.2×10^6 to 2.0×10^6. However, from 2.0×10^6 episodes, we can see that the agent's average reward value decreased from -20 to -40 by 3.0×10^6 episodes. Figure 5(b) represents the average reward value the utilitarian agent received using Dempster credence. We can see that the minimum reward value is -10. It is evident from the graph that the agent was able to maintain a constant reward value of -4 until 0.7×10^6 episodes when dealing with the classic environment. It then decreased to -10 by 1.7×10^6 episodes

and remained moderately low with few elevations in the reward until the 3.0×10^6 episode. From Figure 5(c), we observe that the agent's minimum average reward value is -4, and the agent was able to achieve a constant -2 reward value for most of the time during the simulation. The agent was constantly able to maintain a -2 average reward for about the first 2.5×10^6 episodes, and the reward value started to decrease during the last 0.5×10^6 episodes from 2.5×10^6 episode to 3.0×10^6 episode.

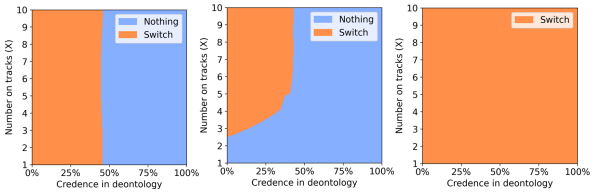
4.3.2 Performance Analysis. In this section we discuss about the performance of the agent with random credence, Dempster credence and simplex credence in different environments using Nash voting and variance voting methods as described in the paper [4]. In the plots, the x-axis represents the credence in deontology theory ranging from 0% - 100%. When the credence in deontology is 10%, the credence in utilitarian will be 100-10 that is 90%. The y-axis represents the no of people on track. The graph illustrates the preferences of an agent across various levels of credence while considering different numbers of people on the track.



(a) (b) (c)

Figure 6: (a) Random credence, (b) Dempster credence, (c) Simplex credence

4.3.3 Classic Environment : Nash Voting. Figures 6(a), 6(b) and 6(c) represent the preference of the choice by the agent with the Nash voting method using random credence, Dempster credence and simplex credence respectively. The Figure 6(a) shows that an agent with random credence tends to prefer switching when there is one person on the track, with a 20% credence in deontology. This preference continues up to a 50% credence in deontology with ten persons on the track. However, when the credence in deontology exceeds 50%, the agent tends to prefer doing nothing, regardless of the number of people on the track. Figure 6(b) depicts the preference of the agent with Dempster credence. The graph indicates that the agent tends to prefer doing nothing when there are fewer than three people on the track, with a 75% credence. However, as the number of people increases beyond three, the agent's preference shifts to switching, regardless of the credence. This preference holds until there are eight people on the track. Additionally, when the credence in deontology reaches around 90%, the agent chooses to do nothing even with eight people on the track. Notably, the agent prefers doing nothing with a 75% credence in deontology when there are ten people on the track. From Figure 6(c) we can see that regardless of no of people on track and credence in deontology the agent with simplex credence always preferred to switch.



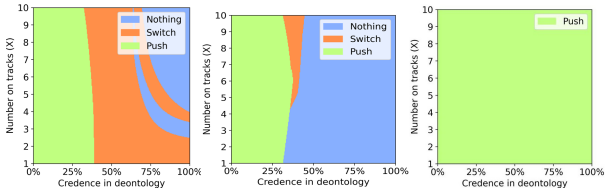
(a) (b) (c)

Figure 7: (a) Random credence, (b) Dempster credence, (c) Simplex credence

4.3.4 Classic Environment : Variance Voting. Figures 7(a), 7(b) and 7(c) represent the agent's choice preference using the variance voting method with random credence, Dempster credence, and simplex credence in classic environment. From Figure 7(a) representing the choice preference of the agent using random credence, we can observe that the agent, with approximately 50% credence in deontology, prefers to switch regardless of the number of people on the track. However, this trend changes when the credence in deontology increases above 50%, leading the agent to prefer doing nothing. Figure 7(b) illustrates the preference of the agent using Dempster credence. From the graph, we observe that the agent prefers doing nothing over switching when there are fewer than 3 people on the track. However, when the number of people on the track increases beyond 3 and the credence in deontology is less than 50%, the agent chooses to switch over doing nothing. Regardless of the count of people on the track, when the credence in deontology is greater than 50%, the agent prefers doing nothing. It is evident from Figure 7(c) that the agent preferred switch over nothing regardless of credence and no of people on track using simplex credence.

4.3.5 Double Environment : Nash Voting. Figures 8(a), 8(b) and 8(c) represent the agent's choice preference in the double environment using the Nash voting method with random credence, Dempster credence, and simplex credence, respectively. Figure 8(a) represents the preference of the agent with random credence. We can observe that when the credence in deontology is less than 30%, the agent prefers to push over doing nothing and switch, regardless of the count of people on the track. As the credence increases beyond 30%, the agent changes its preference to switch. The agent continues to choose switch when there are fewer than 3 people on the track, with 100% credence in deontology. However, when the number of people on the track increases beyond 3, with 75% credence, the agent prefers doing nothing. This preference remains consistent until there are 10 people on the track with slight variations between switch and nothing when the no of people on track ranges from 3-5. Figure 8(b) represents the preference of the agent using Dempster credence. From the

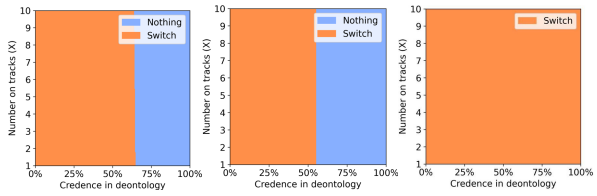
graph, we can observe that the agent chooses to push when the credence in deontology is around 30%, regardless of the number of people on the track. The agent changes its preference to doing nothing when the credence in deontology reaches 50%, a preference that constant until the end. The agent only chooses to switch when the number of people on the track is more than 5 and the credence in deontology is between 30% and 50%. From Figure 8(c) we can observe that the agent with simplex credence always preferred to push irrespective of credence and number of people on track.



(a) (b) (c)

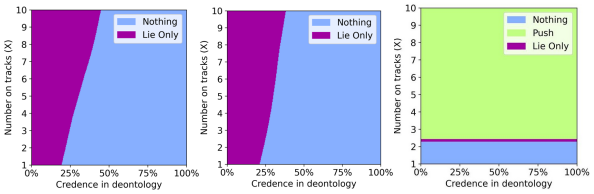
Figure 8: (a) Random credence, (b) Dempster credence, (c) Simplex credence

4.3.6 Double Environment : Variance Voting. Figures 9(a), 9(b) and 9(c) represent the agent's choice preference in double environment using the variance voting method with random credence, Dempster credence, and simplex credence, respectively. Figure 9(a) represents the preference of the agent with random credence. From the graph we can see that the agent choose to switch till 60% of credence in deontology and from 60% to 100% of credence in deontology the agent choose to nothing over switch irrespective of number of people on tracks. The Figure 9(b) represents the preference of the agent using Dempster credence. From the graph, we can observe that the agent chooses to switch when the credence in deontology is around 50%, regardless of the number of people on the track. The agent changes its preference to doing nothing when the credence in deontology ranges from 50% to 100%. From Figure 9(c) we can observe that the agent with simplex credence always preferred to switch.



(a) (b) (c)

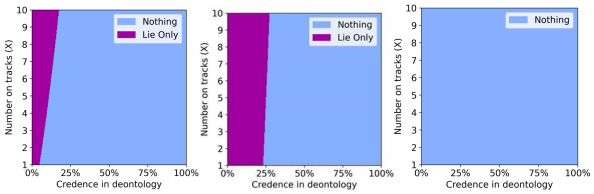
Figure 9: (a) Random credence, (b) Dempster credence, (c) Simplex credence



(a) (b) (c)

Figure 10: (a) Random credence, (b) Dempster credence, (c) Simplex credence

4.3.7 Guard Environment : Nash Voting. Figures 10(a), 10(b) and 10(c) represent the preference of agent with Nash voting using random, Dempster and simplex credence. From Figure 10(a) we can observe that there is a correlation between the number of people on track and choice of the agent until the credence in deontology is around 50%. As the number of people increase from 1-10 the the agent choose to lie only. Once the credence in deontology increases more than 40% the agent changes its preference to nothing and this action continues till the credence in deontology is 100%. From Figure 10(b), which represents the preference of the agent with Dempster credence, we observe a correlation between the number of people on the track and the agent's preference. However, this correlation is only noticeable when the credence is less than 40%. As the credence increases beyond 40%, the agent changes its preference to doing nothing, maintaining this preference until 100% credence. In Figure 10(c), we see that the agent chooses to do nothing when there are fewer than 3 people on the track, regardless of the credence in deontology. When the number of people on the track is greater than 4, the agent chooses to push instead of doing nothing. This preference is consistent across different levels of credence. However, when the number of people on the track is around 3-4, the agent chooses to lie, irrespective of the credence in deontology.

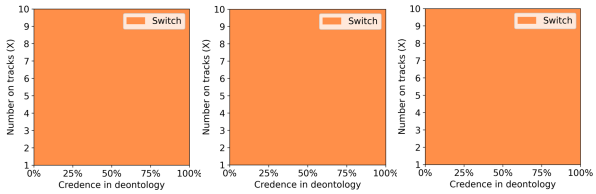


(a) (b) (c)

Figure 11: (a) Random credence, (b) Dempster credence, (c) Simplex credence

4.3.8 Guard Environment : Variance Voting. The Figures 11(a), 11(b) and 11(c) represents the preference of agent with variance voting using random, Dempster and simplex credence. From Figure 11(a), we can see a correlation between the number of people and the choice made by the agent. As the number of people increases, the agent's preference to choose to lie continues, even when the credence

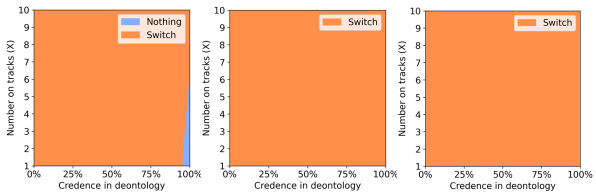
increases up to 20%. However, when the credence in deontology increases beyond 20%, we can see a switch in the agent's preference to do nothing, which continues until the end. In Figure 11(b), representing the preference of the agent with Dempster credence, we observe that the agent prefers to lie only when the credence in deontology is less than 25%. Once the credence increases beyond 25%, the agent chooses to do nothing, continuing this preference until 100%. From Figure 11(c), representing the agent's preference using simplex credence, we can observe that the agent prefers doing nothing over lying or pushing throughout the entire simulation.



(a) (b) (c)

Figure 12: (a) Random credence, (b) Dempster credence, (c) Simplex credence

4.3.9 Doomsday Environment : Nash Voting. Figures 12(a), 12(b) and 12(c) represent the preference of the agent with Nash voting using random credence, Dempster credence and simplex credence. From the figures below, it is clear that the agent always preferred to switch throughout, regardless of the credence and number of people on the track. The same preference is observed in all three cases with random, Dempster, and simplex credence.



(a) (b) (c)

Figure 13: (a) Random credence, (b) Dempster credence, (c) Simplex credence

4.3.10 Doomsday Environment : Variance Voting. Figures 13(a), 13(b) and 13(c) represent the preference of the agent with variance voting using random, Dempster and simplex credence. From Figure 13(a), which represents the agent's preference with random credence, we can see that the agent consistently chose to switch over doing nothing or selecting the doomsday option. This preference remains constant until the credence in deontology reaches 95%. However, the agent changes its preference when the number of people on the track is less than 5, but with a credence greater than 95%. The agent's preference was similar in a doomsday scenario when using Dempster and Simplex credence, as shown in Figures 13(b) and 13(c). The agent preferred switch over nothing and the doomsday option throughout the simulation.

5 DISCUSSION

The preference for a moral theory depends on the credence generated. With random credence, the agent's performance is unstable, whereas with Dempster credence, the agent's performance is uniform and better than with random credence. The agent using the simplex credence performs better than the agent using random and Dempster credence. This can be observed from Figure 5(c), where the maximum average reward value gained by the utilitarian agent using simplex credence is -4, while the agents using random and Dempster credence gained -40 (see Figure 5(a)) and -10 (see Figure 5(b)), respectively.

5.1 Key Scientific Insights

From the analysis of agent preferences, it is evident that the agent using simplex credence consistently outperformed the agents using random credence and Dempster credence by consistently choosing the option that saves the greatest number of people.

- This preference is clearly demonstrated in the classic environment with both voting methods, where the agent consistently preferred to switch over doing nothing (see Figures 6 , 7).
- In the double environment, when the agent used Nash voting, the agent using simplex credence only preferred to push over doing nothing and switch. In contrast, when the agent used variance voting, the agent preferred to switch over push and nothing. In both cases, the agent's preference save more number of people when compared to agents preference using random and Dempster credence.
- The superiority of simplex credence is further highlighted in the guard environment with Nash voting. The agent preferred to do nothing when the number of people on the track was less than 2, and to push when there were 3 or more people on the track. Pushing the large man onto the track requires the agent to first lie to the guard. In order to save more people, the agent chose to push. When the agent used variance voting with simplex credence, the agent preferred to do nothing over lying only and pushing. In contrast, the agents with random credence and Dempster credence opted for lying only without pushing the large man when the credence in deontology was less than 20%. This demonstrates the consistency of the agent using simplex credence.
- In the doomsday environment with both voting methods, all three agents showed similar preferences. However, the agent using random credence with variance voting changed its preference to do

nothing from switch when the credence in deontology was around 95% and the number of people on the track was less than 5. Where as, the agents using Dempster credence and simplex credence were consistent in choosing switch from start to end. From the performance analysis of the agents, it is evident that using simplex credence instead of random credence boosted the agent's performance in saving a greater number of people in most environments.

Using simplex credence consistently increased the agent's efficiency in saving more people, with no signs of integration issues affecting the agent's performance.

6 CONCLUSION

We propose two different approaches to deal with the problem of random credence. In the first approach, we combine random credence with the Dempster-Shafer technique to stabilize the agent's performance. In the second approach, we introduce a novel framework that generates credence from moral theories using the simplex algorithm and the Dempster-Shafer technique collectively. From the preferences of agents in different environments, we observe that agents using Dempster and simplex credence consistently choose the option that saves more people. This can be seen in the results, where the agent using simplex credence incurs considerably less penalty compared to the agent using Dempster credence, indicating a preference for saving more people. From the results it is evident that the choice preference of agent using Dempster credence is uniform where the agent using simplex preferred the choice which saved more number of people. In future work, we plan to incorporate the dynamics of the agent's environment into the framework to calculate credence.

LIMITATIONS

One limitation of our study is that we observed the selection of only one theory by the agents, reflecting a utilitarian behavior. While this aligns with the simplicity of our model, it may not capture the complexity of real-world ethical decision-making, which often involves considering multiple ethical theories simultaneously. Another limitation is that the agent's behavior using simplex credence changed when the voting method was switched from Nash voting to variance voting this is observed mainly in double and guard environment. This behavior change suggests that the choice of voting method can influence the agent's decision-making process. Additionally, we found that the agent's behavior with Nash voting was generally better than with variance voting.

REFERENCES

[1] Kareem Othman. Exploring the implications of autonomous vehicles: a comprehensive review. *Innovative Infrastructure Solutions*, 7(165):165, March 2022.

[2] Enrica Papa and António Ferreira. Sustainable accessibility and the implementation of automated vehicles: Identifying critical decisions. *Urban Science*, 2:5, 01 2018.

[3] Hong Wang, Amir Khajepour, Dongpu Cao, and Teng Liu. Ethical decision making in autonomous vehicles: Challenges and research progress. *IEEE Intelligent Transportation Systems Magazine*, 14(1):6–17, 2022.

[4] Joel Lehman Adrien Ecoffet. Reinforcement learning under moral uncertainty. 2021.

[5] Raphael Max Alexander Kriebitz and Christoph Lütge. The german act on autonomous driving: Why ethics still matters. *Philosophy Technology*, 35(29), April 2022.

[6] de Moura Nelson Chauvier Stéphane Chatila Raja Dogan Ebru Evans, Katherine. Ethical decision making in autonomous vehicles: The av ethics project. *Science and Engineering Ethics*, 26, December 2020.

[7] Nelson De Moura, Raja Chatila, Katherine Evans, Stéphane Chauvier, and Ebru Dogan. Ethical decision making for autonomous vehicles. pages 2006–2013, 2020.

[8] Szilárd Aradi. Survey of deep reinforcement learning for motion planning of autonomous vehicles. *IEEE Transactions on Intelligent Transportation Systems*, 23(2):740–759, 2022.

[9] Bykvist Krister Ord Toby MacAskill, Michael. *Moral uncertainty*. Oxford University Press, 2020.

[10] Elumalai P Sangeetha V, Thirisangu K. Dual simplex method based solution for a fuzzy transportation problem. *Journal of Physics: Conference Series*, 1947(1):012017, June 2021.

[11] KARI SENTZ and SCOTT FERSON. Combination of evidence in dempster-shafer theory. Apr 2002.

[12] Ron Shamir. The efficiency of the simplex method: a survey. *Management Science*, 33(3):301–334, Mar 1987.

[13] Kyle Bogosian. Implementation of moral uncertainty in intelligent machines. *Minds Mach.*, 27(4):591–608, Dec 2017.

[14] Noushin Mehdipour Radboud Duintjer Tebbens Amitai Y. Bin-Nun, Patricia Derler. How should autonomous vehicles drive? policy, methodological, and social considerations for designing a driver. *Humanities and Social Sciences Communications*, August 2022.

[15] Prafulla Dhariwal Alec Radford Oleg Klimov John Schulman, Filip Wolski. Proximal policy optimization algorithms. 2017.

[16] Barto Andrew G Sutton, Richard S. *Reinforcement learning: An introduction*. MIT press, Cambridge, MA, 2nd. edition, 2018.

[17] The SciPy community. Linear programming optimization using tableau-based simplex method, Mar 2024.

[18] Joel Lehman Adrien Ecoffet. Normative uncertainty, reinforcement learning under moral uncertainty, Mar 2021.