Ethics, Prosperity and Society: Moral Evaluation Using Virtue Ethics And Utilitarianism

Aditya Hegde, Vibhav Agarwal, Shrisha Rao

IJCAI-PRICAI 2020

Introduction

- Agent Based Modelling
  - Model operations and interactions of agents to understand complex phenomena.
  - Emergent macro-properties from micro-scale agent behaviours.
  - Used for modelling spread of epidemics, population dynamics, financial markets, evacuation during emergencies, etc.

Introduction

- Agent Based Modelling
  - Model operations and interactions of agents to understand complex phenomena.
  - Emergent macro-properties from micro-scale agent behaviours.
  - Used for modelling spread of epidemics, population dynamics, financial markets, evacuation during emergencies, etc.
- Incorporating ethics into Agent Based Modelling
  - Practical implications of ethical theories.
  - Understand and analyse social phenomena and interactions.

Virtue Ethics and Utilitarianism

- Normative Ethics: Branch of ethics that discusses when an action is right or wrong.
- Virtue Ethics
  - Emphasises the inherent moral nature of actions.
- Utilitarianism
  - Emphasises the betterment of society as a whole.

Our Work

- Framework for modelling **ethical decision making** as well as **evaluation of agent behaviour**.
  - **Moral Interactions** capture ethical decision making.
  - Evaluation of agent behaviour using **virtue ethics** and **utilitarianism**.
- Virtue agents
  - Parametrised by level of ethics.
  - Behaviour depends on the agent’s level of ethics.

Our Work

- Simulations
  - Are unethical agents more prosperous?
  - How does societal bias towards positive and negative actions impact agent prosperity?
  - How does the ethical composition of agent population affect society as a whole?

Prior Work

- One of the first instances of using ethics in computer simulations is the work of Danielson [Danielson, 1992].

- Ethics in Agent Based Modelling

  - Evaluation of agent behaviour using ethics [Korb et al., 2010; Cointe et al., 2016].

  - Ethical decision making [Wiegel and van den Berg, 2009; Gaudou et al., 2014].

[Cointe et al., 2016] Nicolas Cointe, Grégory Bonnet, and Olivier Boissier. Ethical judgment of agents' behaviors in multi-agent systems. AAMAS '16, page 1106–1114, Richland, SC, 2016.

[Danielson, 1992] Peter Danielson. Artificial Morality: Virtuous Robots for Virtual Games. Routledge, 1992.

[Korb et al., 2010] Kevin B. Korb, Ann E. Nicholson, and Owen Woodberry. Evolving Ethics: The New Science of Good and Evil. Imprint Academic, 2010.

[Gaudou *et al.*, 2014] Benoit Gaudou, Emiliano Lorini, and Eunate Mayor. Moral Guilt: An Agent-Based Model Analysis. In Advances in Social Simulation, Advances in Intelligent Systems and Computing, pages 95–106, Berlin, Heidelberg, 2014. Springer.

[Wiegel and van den Berg, 2009] Vincent Wiegel and Jan van den Berg. Combining Moral Theory, Modal Logic and Mas to Create Well-Behaving Artificial Agents. *International Journal of Social Robotics*, 1(3):233–242, August 2009.

Framework and Virtue Agents

Framework

- Cellular automaton
- Each iteration
  - Every agent performs an interaction with one of its neighbours.
  - Random order every iteration.
- Interactions governed by agent strategies and parameters.

Framework

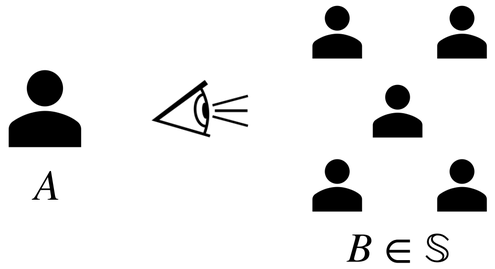
- Cellular automaton
- Each iteration
  - Every agent performs an interaction with one of its neighbours.
  - Random order every iteration.
- Interactions governed by agent strategies and parameters.
- $\mathbb{S}$: Set of all agents.
- $\mathcal{N}(A)$: Neighbours of agent A.

Agent Parameters - Resource

- Agent's prosperity in society.
- r_A: Agent A's resource.
- All agents start with the same resource value.
- Changes through interactions.

Agent Parameters - Opinion

- $\Psi_A(B)$: A's opinion of B.
  - B is any agent in the simulation.
  - Between 0 and 1.
  - A's perception of B's ethicality.

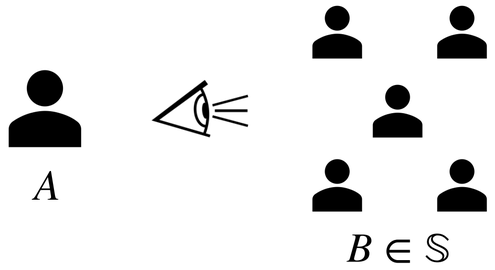


Opinion

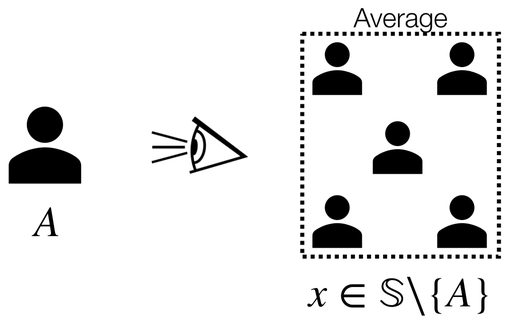
Agent Parameters - Opinion

- $\Psi_A(B)$: A's opinion of B.
  - B is any agent in the simulation.
  - Between 0 and 1.
  - A's perception of B's ethicality.
- Reputation: Average opinion of A across all agents.
  - How ethical is A perceived to be in general.

$$\frac{\sum_{x \in \mathbb{S} \setminus \{A\}} \Psi_x(A)}{\mathbb{S} \ - 1}$$



Opinion



Reputation

Virtue Agents

- Several well-known agent strategies like Tit For Tat (TFT), Suspicious TFT, Grim Trigger, etc.
  - No straightforward way to instantiate agents with ethical and unethical behaviour.

Virtue Agents

- Several well-known agent strategies like Tit For Tat (TFT), Suspicious TFT, Grim Trigger, etc.
  - No straightforward way to instantiate agents with ethical and unethical behaviour.
- Virtue Agents are parametrised by **level of ethics**, ϵ where $0 \leq \epsilon \leq 1$.
  - Behaviour depends on agent's level of ethics.

Virtue Agents

- Several well-known agent strategies like Tit For Tat (TFT), Suspicious TFT, Grim Trigger, etc.
  - No straightforward way to instantiate agents with ethical and unethical behaviour.
- Virtue Agents are parametrised by **level of ethics**, ϵ where $0 \leq \epsilon \leq 1$.
  - Behaviour depends on agent's level of ethics.
- Make use of opinion in their interactions.
  - Motivation: Our actions are based on social perception [Smith, 1982].
  - Opinion is interpreted as the perceived ethicality of an agent.

[Smith, 1982] Christopher Upham Murray Smith. Evolution and the problem of mind: Part 1. Herbert Spencer. Journal of the History of Biology, 15(1):55–88, 1982.

Agent Interactions

- Two types of interactions
  - Continuous Prisoner’s Dilemma (CPD)
  - Moral Interactions.

Agent Interactions

- Two types of interactions
  - Continuous Prisoner’s Dilemma (CPD)
  - Moral Interactions.
- Prisoner’s Dilemma
  - Agent’s opt to either cooperate or defect.

Agent Interactions

- Two types of interactions
  - Continuous Prisoner's Dilemma (CPD)
  - Moral Interactions.
- Prisoner's Dilemma
  - Agent's opt to either cooperate or defect.

Prisoner's Dilemma Matrix

	B Cooperates	B Defects
A Cooperates	-1, -1	-3, 0
A Defects	0, -3	-2, -2

Continuous Prisoner's Dilemma

- Similar to the Iterated Prisoner's Dilemma
  - Cooperation levels between 0 and 1 instead of complete defection or cooperation.
  - Payoffs are scaled based on the cooperation levels.
- Agents interact with a random neighbour.
- Donation game payoff matrix.
  - Trade of goods between A and B.
  - Payoffs can be positive or negative.

Donation Game Payoff Matrix

	B Cooperates	B Defects
A Cooperates	$\alpha - \beta, \alpha - \beta$	$-\beta, \alpha$
A Defects	$\alpha, -\beta$	0, 0

Continuous Prisoner's Dilemma

- Similar to the Iterated Prisoner's Dilemma
  - Cooperation levels between 0 and 1 instead of complete defection or cooperation.
  - Payoffs are scaled based on the cooperation levels.
- Agents interact with a random neighbour.
- Donation game payoff matrix.
  - Trade of goods between A and B.
  - Payoffs can be positive or negative.

Donation Game Payoff Matrix

	B Cooperates	B Defects
A Cooperates	$\alpha - \beta, \alpha - \beta$	$-\beta, \alpha$
A Defects	$\alpha, -\beta$	0, 0

Continuous Prisoner's Dilemma

- Similar to the Iterated Prisoner's Dilemma
  - Cooperation levels between 0 and 1 instead of complete defection or cooperation.
  - Payoffs are scaled based on the cooperation levels.
- Agents interact with a random neighbour.
- Donation game payoff matrix.
  - Trade of goods between A and B.
  - Payoffs can be positive or negative.

Donation Game Payoff Matrix

	B Cooperates	B Defects
A Cooperates	$\alpha - \beta, \alpha - \beta$	$-\beta, \alpha$
A Defects	$\alpha, -\beta$	0, 0

Virtue Agent CPD Strategy

- Interactions are influenced by those around us, especially those who we hold in high regard [Moussaïd *et al.*, 2013; Campbell-Meiklejohn *et al.*, 2010].

- Virtue agents aggregate neighbour's opinion when outputting cooperation level.

  - It weighs the opinion of a neighbour proportional to the perceived ethicality of its neighbour.

[Moussaïd *et al.*, 2013] Mehdi Moussaïd, Juliane E. Kämmer, Pantelis Pipergias Analytis, and Hansjörg Neth. Social influence and the collective dynamics of opinion formation. *PLOS ONE*, 8(11):1–8, 11 2013.

[Campbell-Meiklejohn *et al.*, 2010] Daniel K. Campbell-Meiklejohn, Dominik R. Bach, Andreas Roepstorff, Raymond J. Dolan, and Chris D. Frith. How the opinion of others affects our valuation of objects. *Current Biology*, 20(13):1165–1170, 2010.

Virtue Agent CPD Strategy

- Interactions are influenced by those around us, especially those who we hold in high regard [Moussaïd *et al.*, 2013; Campbell-Meiklejohn *et al.*, 2010].
- Virtue agents aggregate neighbour's opinion when outputting cooperation level.
  - It weighs the opinion of a neighbour proportional to the perceived ethicality of its neighbour.
- A is performing a CPD interaction with B

[Moussaïd *et al.*, 2013] Mehdi Moussaïd, Juliane E. Kämmer, Pantelis Pipergias Analytis, and Hansjörg Neth. Social influence and the collective dynamics of opinion formation. *PLOS ONE*, 8(11):1–8, 11 2013.

[Campbell-Meiklejohn *et al.*, 2010] Daniel K. Campbell-Meiklejohn, Dominik R. Bach, Andreas Roepstorff, Raymond J. Dolan, and Chris D. Frith. How the opinion of others affects our valuation of objects. *Current Biology*, 20(13):1165–1170, 2010.

Virtue Agent CPD Strategy

- Interactions are influenced by those around us, especially those who we hold in high regard [Moussaïd *et al.*, 2013; Campbell-Meiklejohn *et al.*, 2010].
- Virtue agents aggregate neighbour's opinion when outputting cooperation level.
  - It weighs the opinion of a neighbour proportional to the perceived ethicality of its neighbour.
- A is performing a CPD interaction with B
  - It asks each neighbour $x \in \mathcal{N}(A)$ what it thinks about B and receives $\psi_x(B)$.

[Moussaïd *et al.*, 2013] Mehdi Moussaïd, Juliane E. Kämmer, Pantelis Pipergias Analytis, and Hansjörg Neth. Social influence and the collective dynamics of opinion formation. *PLOS ONE*, 8(11):1–8, 11 2013.

[Campbell-Meiklejohn *et al.*, 2010] Daniel K. Campbell-Meiklejohn, Dominik R. Bach, Andreas Roepstorff, Raymond J. Dolan, and Chris D. Frith. How the opinion of others affects our valuation of objects. *Current Biology*, 20(13):1165–1170, 2010.

Virtue Agent CPD Strategy

- Interactions are influenced by those around us, especially those who we hold in high regard [Moussaïd *et al.*, 2013; Campbell-Meiklejohn *et al.*, 2010].
- Virtue agents aggregate neighbour's opinion when outputting cooperation level.
  - It weighs the opinion of a neighbour proportional to the perceived ethicality of its neighbour.
- A is performing a CPD interaction with B
  - It asks each neighbour $x \in \mathcal{N}(A)$ what it thinks about B and receives $\psi_x(B)$.
  - It scales $\psi_x(B)$ according to $\psi_A(x)$.

[Moussaïd *et al.*, 2013] Mehdi Moussaïd, Juliane E. Kämmer, Pantelis Pipergias Analytis, and Hansjörg Neth. Social influence and the collective dynamics of opinion formation. *PLOS ONE*, 8(11):1–8, 11 2013.

[Campbell-Meiklejohn *et al.*, 2010] Daniel K. Campbell-Meiklejohn, Dominik R. Bach, Andreas Roepstorff, Raymond J. Dolan, and Chris D. Frith. How the opinion of others affects our valuation of objects. *Current Biology*, 20(13):1165–1170, 2010.

Virtue Agent CPD Strategy

- Interactions are influenced by those around us, especially those who we hold in high regard [Moussaïd *et al.*, 2013; Campbell-Meiklejohn *et al.*, 2010].
- Virtue agents aggregate neighbour's opinion when outputting cooperation level.
  - It weighs the opinion of a neighbour proportional to the perceived ethicality of its neighbour.
- A is performing a CPD interaction with B
  - It asks each neighbour $x \in \mathcal{N}(A)$ what it thinks about B and receives $\psi_x(B)$.
  - It scales $\psi_x(B)$ according to $\psi_A(x)$.

$$c_A = \frac{\sum_{x \in \mathcal{N}(A) \setminus B} w_A(x) \Psi_x(B)}{\sum_{x \in \mathcal{N}(A) \setminus B} w_A(x)} \qquad w_A(x) = \mathscr{H}_{1,1}(\Psi_A(x))$$

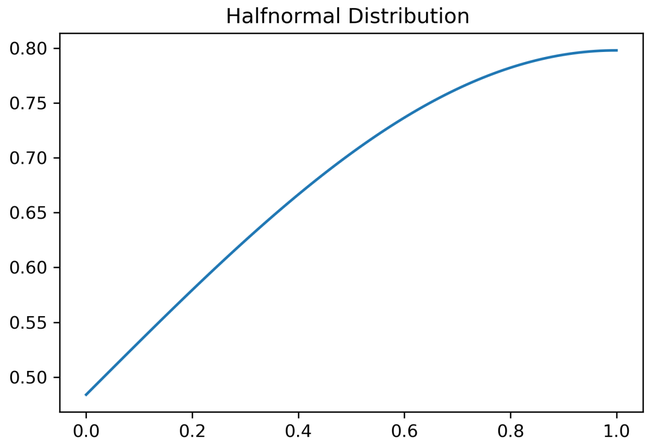
[Moussaïd *et al.*, 2013] Mehdi Moussaïd, Juliane E. Kämmer, Pantelis Pipergias Analytis, and Hansjörg Neth. Social influence and the collective dynamics of opinion formation. *PLOS ONE*, 8(11):1–8, 11 2013.

[Campbell-Meiklejohn *et al.*, 2010] Daniel K. Campbell-Meiklejohn, Dominik R. Bach, Andreas Roepstorff, Raymond J. Dolan, and Chris D. Frith. How the opinion of others affects our valuation of objects. *Current Biology*, 20(13):1165–1170, 2010.

Virtue Agent CPD Strategy

- Interactions are influenced by those around us, especially those who we hold in high regard [Moussaïd *et al.*, 2013; Campbell-Meiklejohn *et al.*, 2010].
- Virtue agents aggregate neighbour's opinion when outputting cooperation level.
  - It weighs the opinion of a neighbour proportional to the perceived ethicality of its neighbour.
- A is performing a CPD interaction with B
  - It asks each neighbour $x \in \mathcal{N}(A)$ what it thinks about B and receives $\psi_x(B)$.
  - It scales $\psi_x(B)$ according to $\psi_A(x)$.

$$c_A = \frac{\sum_{x \in \mathcal{N}(A) \setminus B} w_A(x) \Psi_x(B)}{\sum_{x \in \mathcal{N}(A) \setminus B} w_A(x)} \qquad w_A(x) = \mathscr{H}_{1,1}(\Psi_A(x))$$



[Moussaïd *et al.*, 2013] Mehdi Moussaïd, Juliane E. Kämmer, Pantelis Pipergias Analytis, and Hansjörg Neth. Social influence and the collective dynamics of opinion formation. *PLOS ONE*, 8(11):1–8, 11 2013.

[Campbell-Meiklejohn *et al.*, 2010] Daniel K. Campbell-Meiklejohn, Dominik R. Bach, Andreas Roepstorff, Raymond J. Dolan, and Chris D. Frith. How the opinion of others affects our valuation of objects. *Current Biology*, 20(13):1165–1170, 2010.

Moral Interactions

- Few social interactions in real life involve ethical choices. These often have higher stakes [Kidder, 2009].
  - CPD models “normal” social interactions.
  - Moral Interactions incorporate ethical decision making.
- θ : Probability to perform a moral interaction instead of CPD.

[Kidder, 2009] Rushworth Moulton Kidder. How Good People Make Tough Choices Rev Ed: Resolving the Dilemmas of Ethical Living. HarperCollins, November 2009.

Moral Interactions

- Few social interactions in real life involve ethical choices. These often have higher stakes [Kidder, 2009].
  - CPD models “normal” social interactions.
  - Moral Interactions incorporate ethical decision making.
- θ : Probability to perform a moral interaction instead of CPD.
- Interacting agent steals or donates to a target agent.
  - Choice of theft or donation.
  - Choice of target agent from neighbours.

[Kidder, 2009] Rushworth Moulton Kidder. How Good People Make Tough Choices Rev Ed: Resolving the Dilemmas of Ethical Living. HarperCollins, November 2009.

Moral Interactions

- Few social interactions in real life involve ethical choices. These often have higher stakes [Kidder, 2009].
  - CPD models “normal” social interactions.
  - Moral Interactions incorporate ethical decision making.
- θ : Probability to perform a moral interaction instead of CPD.
- Interacting agent steals or donates to a target agent.
  - Choice of theft or donation.
  - Choice of target agent from neighbours.
- A donates to B: Transfer of δ_d units of resource from A to B.
- A steals from B: Transfer of δ_t units of resource from B to A.

[Kidder, 2009] Rushworth Moulton Kidder. How Good People Make Tough Choices Rev Ed: Resolving the Dilemmas of Ethical Living. HarperCollins, November 2009.

Virtue Agent Moral Interaction Strategy

- Theft vs Donation
  - Ethical agents are expected to donate.
  - Virtue agents opt for donation with probability ϵ.

Virtue Agent Moral Interaction Strategy

- Theft vs Donation
  - Ethical agents are expected to donate.
  - Virtue agents opt for donation with probability ϵ.
- Motivation for Moral Interaction strategy
  - People like to make gifts which they believe will make a tangible difference; to targets they like [Cryder and Loewenstein, 2011].
  - Criminals often focus on targets that they consider more lucrative [Vandeviver and Bernasco, 2019].

[Cryder and Loewenstein, 2011] Cynthia Cryder and George Loewenstein. The critical link between tangibility and generosity. In Society for Judgment and Decision Making series. The science of giving: Experimental approaches to the study of charity, pages 237–251. Psychology Press, 2011.

[Vandeviver and Bernasco, 2019] Christophe Vandeviver and Wim Bernasco. “location, location, location”: Effects of neighborhood and house attributes on burglars’ target selection. *Journal of Quantitative Criminology*, pages 1–43, 2019.

Virtue Agent Moral Interaction Strategy

- Theft vs Donation
  - Ethical agents are expected to donate.
  - Virtue agents opt for donation with probability ϵ.
- Motivation for Moral Interaction strategy
  - People like to make gifts which they believe will make a tangible difference; to targets they like [Cryder and Loewenstein, 2011].
  - Criminals often focus on targets that they consider more lucrative [Vandeviver and Bernasco, 2019].
- Donation target should have low resource and high opinion while the opposite is true for theft targets.
  - Donation target: Agent with maximum opinion to resource ratio.
  - Theft target: Agent with minimum opinion to resource ratio.

[Cryder and Loewenstein, 2011] Cynthia Cryder and George Loewenstein. The critical link between tangibility and generosity. In Society for Judgment and Decision Making series. The science of giving: Experimental approaches to the study of charity, pages 237–251. Psychology Press, 2011.

[Vandeviver and Bernasco, 2019] Christophe Vandeviver and Wim Bernasco. “location, location, location”: Effects of neighborhood and house attributes on burglars’ target selection. *Journal of Quantitative Criminology*, pages 1–43, 2019.

Opinion Updates

- Agents evaluate behaviour of interacting agents.
  - Change in opinion.
- Virtue Ethics
  - Inherent moral nature of actions.
  - Higher cooperation levels and acts of donation.
- Utilitarianism
  - Acts that increases the global utility are considered to be ethical.
  - Global utility: Sum of resource of all agents.

Opinion Updates - CPD

- Only interacting agents, A and B, perform updates.
- Sum of payoffs s is change in global utility.
- A updates its opinion of B
  - $\psi_A(B)$ is increased by ω_v if cooperation level of B is greater than λ_v and decreased by ω_v otherwise.
  - $\psi_A(B)$ is increased by ω_u if $s > \lambda_u$ and decreased by ω_u otherwise.
- Identical updates by B.

Opinion Updates - Moral Interaction

- Broadcast: γ fraction of agents update their opinion of interacting agent A.
- The agents x which receive the broadcast
  - Increase $\psi_x(A)$ by ω_d if A performed a donation.
  - Decrease $\psi_x(A)$ by ω_t if A performed a theft.

Opinion Updates - Moral Interaction

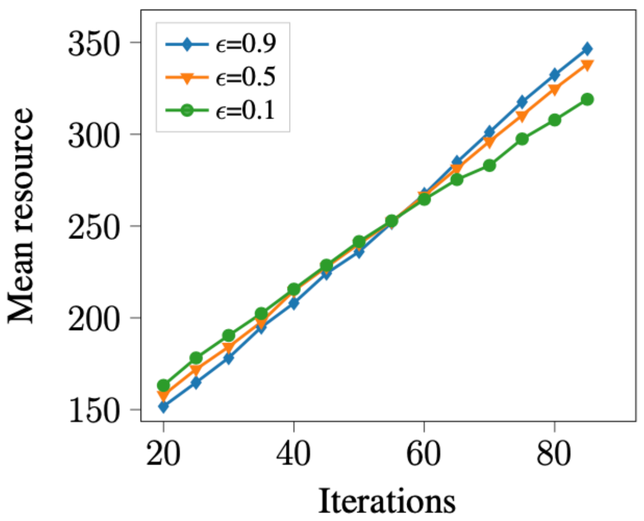
- Broadcast: γ fraction of agents update their opinion of interacting agent A.
- The agents x which receive the broadcast
  - Increase $\psi_x(A)$ by ω_d if A performed a donation.
  - Decrease $\psi_x(A)$ by ω_t if A performed a theft.
- ω_d and ω_t determine society's bias towards ethical and unethical actions
  - $\omega_d < \omega_t$: Negativity bias
  - $\omega_d > \omega_t$: Positivity bias
  - $\omega_d = \omega_t$: No bias

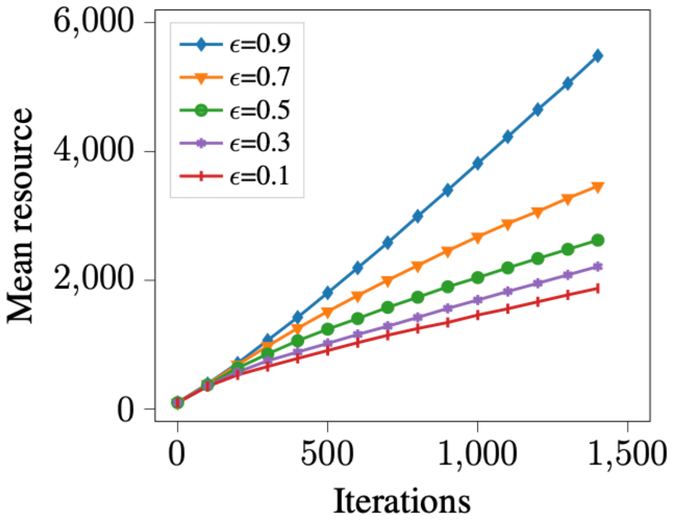
Experiments

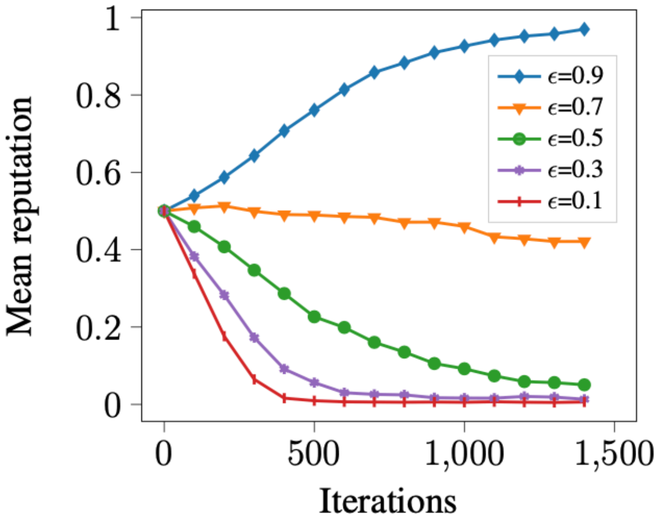
Experiment Setup

- Analyse emergent trends through simulations.
- Simulations consist of virtue agents with different levels of ethics.
  - 50 agents for a given value of ϵ.
  - All agents randomly arranged on the grid.
  - 1500 iterations.
- Moral interactions are fewer and have high stakes
  - $\theta = 0.05 \ll 1$
  - $\omega_d, \omega_t \gg \omega_v, \omega_u$

Comparing Agent Resource Across Time

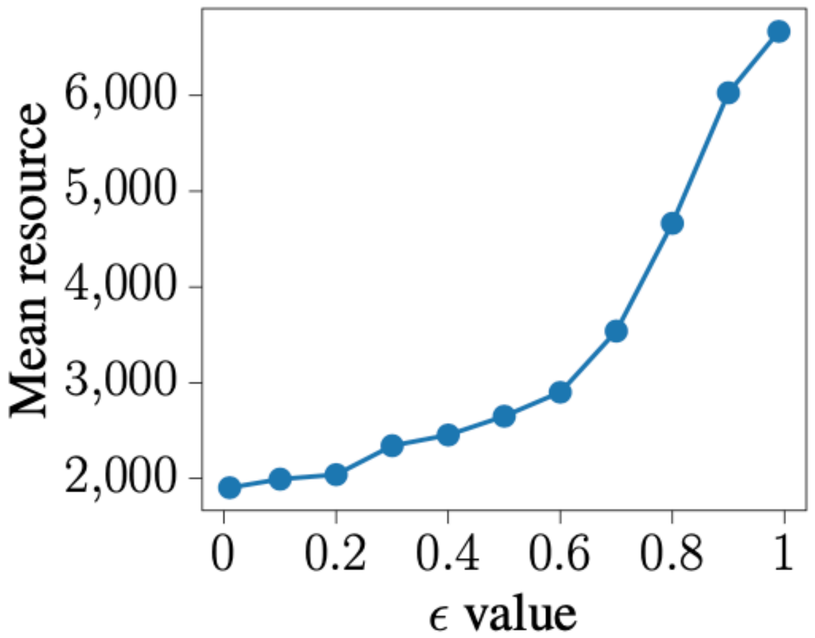


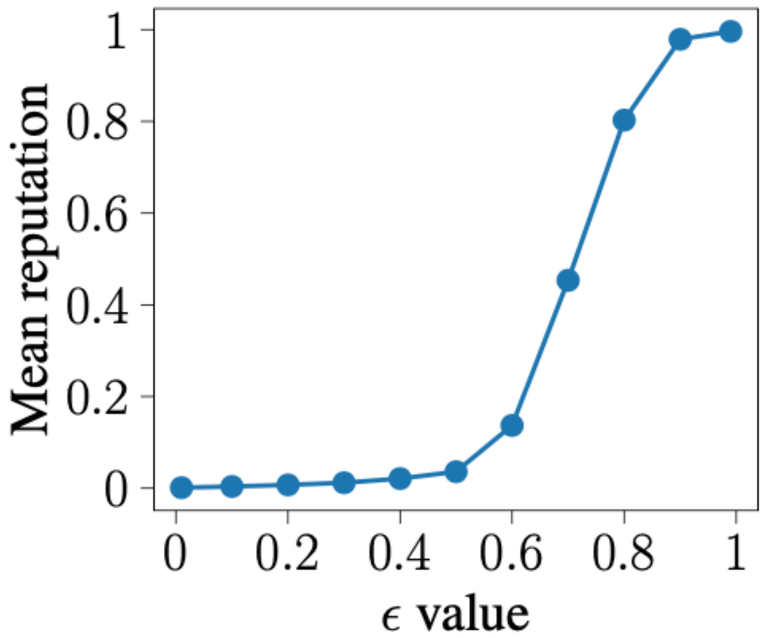




- Unethical agents have higher resources initially but have significantly lower resources in the long run.

Effects of Ethics on Resource in the Long Run

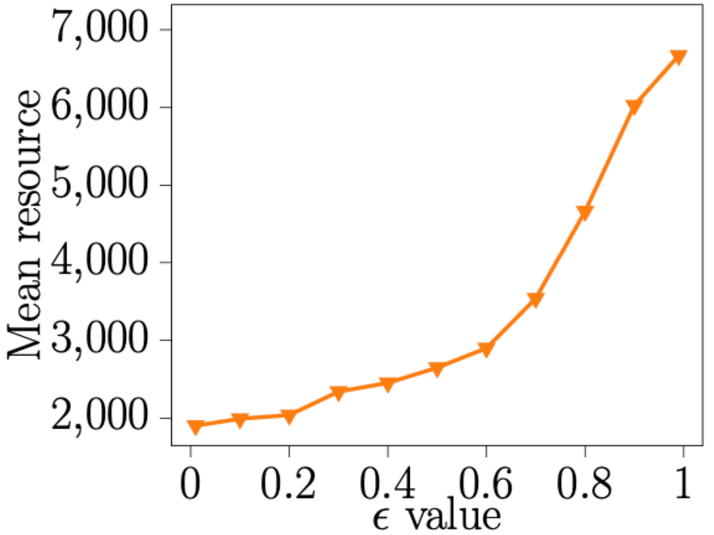


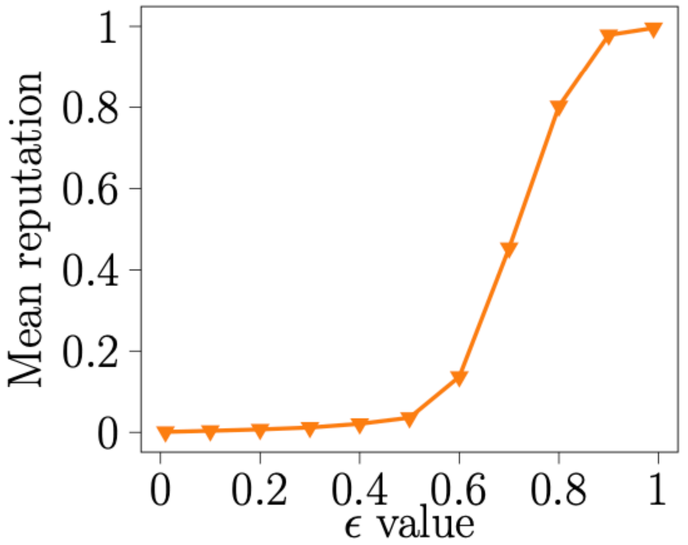


- Resource and reputation steeply increases with ethics.

Bias in Society - Negativity Bias

- High rate of change in the high-ethics range.
  - Ethical agents have large incentive to be more ethical to increase their reputation as well as resource.
- Lower rate of change in the low ethics range.
  - Unethical agents might become more unethical for short-term gains.
- Societies with negativity bias show divergent trends.

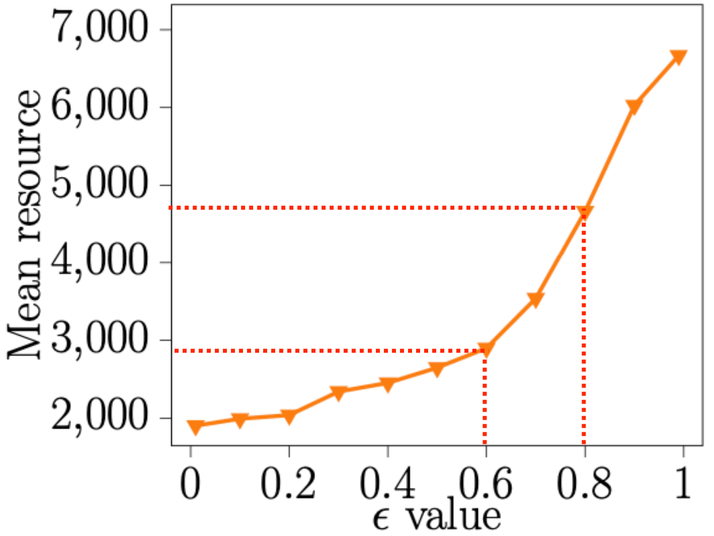


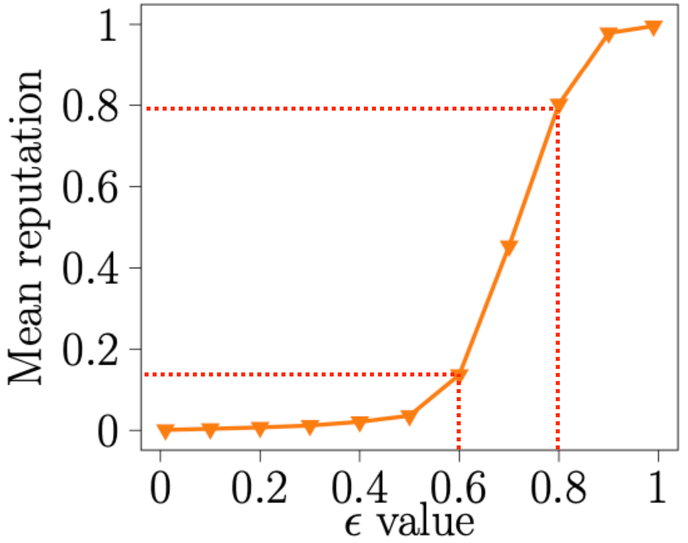


ω_d=0.02 || ω_t=0.05

Bias in Society - Negativity Bias

- High rate of change in the high-ethics range.
  - Ethical agents have large incentive to be more ethical to increase their reputation as well as resource.
- Lower rate of change in the low ethics range.
  - Unethical agents might become more unethical for short-term gains.
- Societies with negativity bias show divergent trends.

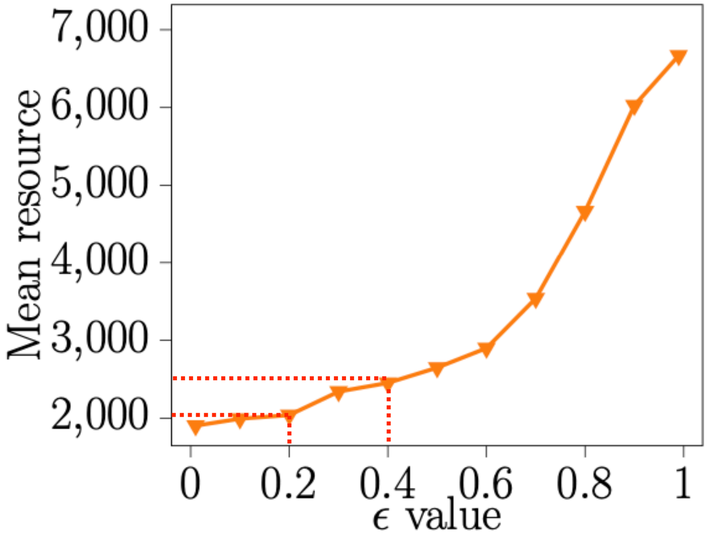


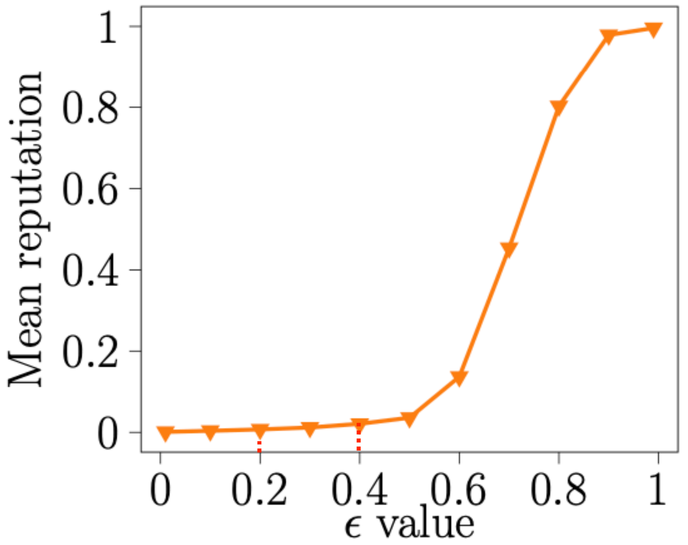


ω_d=0.02 || ω_t=0.05

Bias in Society - Negativity Bias

- High rate of change in the high-ethics range.
  - Ethical agents have large incentive to be more ethical to increase their reputation as well as resource.
- Lower rate of change in the low ethics range.
  - Unethical agents might become more unethical for short-term gains.
- Societies with negativity bias show divergent trends.

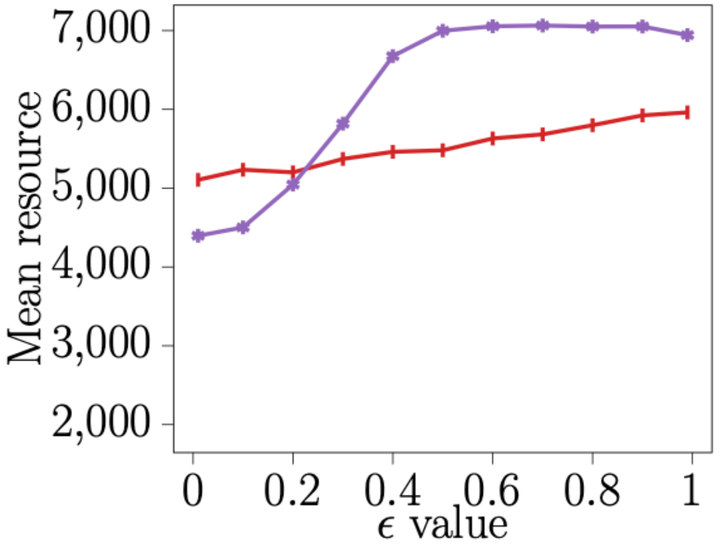


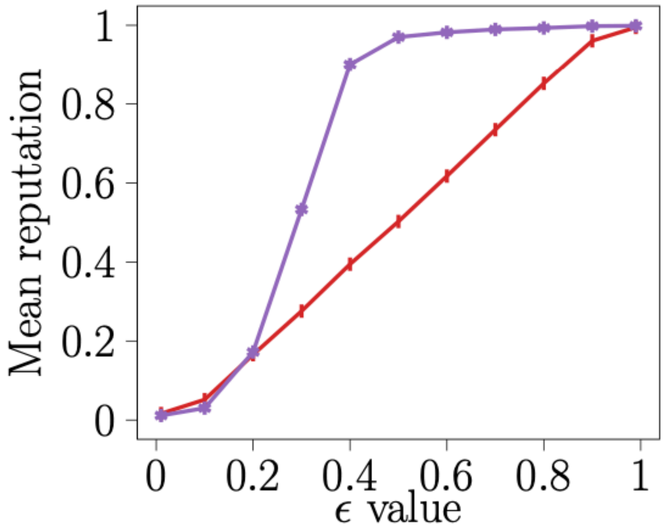


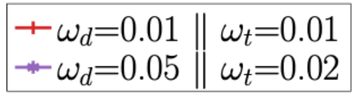
ω_d=0.02 || ω_t=0.05

Bias in Society - Positivity and No Bias

- High rate of change in the low ethics range.
  - Unethical agents have incentive to be more ethical.
- Low rate of change in the high ethics range.
  - Not much incentive for ethical agents to change.
- Societies with positivity bias show divergent trends.
- Societies without bias don't seem to provide strong incentives.







Conclusion

- Other results
  - Even a small population of ethical agents leads to a significant increase in the global utility.
  - Ethics of different agent strategies.
- Rewarding good deeds might provide a stronger incentive for people to be ethical.
  - Our society emphasises penalising unethical behaviour [Galak and Chow, 2019].
  - Prior work shows that rewards are more effective than punishments at securing cooperation [Rand and Nowak 2011; Dreber et al. 2008].
  - Our work shows that rewarding ethical behaviour might provide a stronger incentive for people to be ethical.

[Dreber et al., 2008] Anna Dreber, David G. Rand, Drew Fudenberg, and Martin A. Nowak. Winners don't punish. Nature, 452(7185):348–351, 2008.

[Galak and Chow, 2019] Jeff Galak and Rosalind M. Chow. Compensate a little, but punish a lot: Asymmetric routes to restoring justice. PLOS ONE, 14(1), 2019. https: doi.org/10.1371/journal.pone.0210676.

[Rand and Nowak, 2011] David G. Rand and Martin Andreas Nowak. The evolution of antisocial punishment in optional public goods games. Nature Communications, 2(1):1–7, 2011.

References

- **[Bostrom and Yudkowsky, 2014]** Nick Bostrom and Eliezer Yudkowsky. The ethics of artificial intelligence. In The Cambridge Handbook of Artificial Intelligence, page 316–334. Cambridge University Press, 2014.
- **[Boyles, 2017]** Robert James M. Boyles. Philosophical signposts for artificial moral agent frameworks. Suri, 6(2), 2017.
- **[Campbell-Meiklejohn *et al.*, 2010]** Daniel K. Campbell-Meiklejohn, Dominik R. Bach, Andreas Roepstorff, Raymond J. Dolan, and Chris D. Frith. How the opinion of others affects our valuation of objects. *Current Biology*, 20(13):1165–1170, 2010.
- **[Cointe et al., 2016]** Nicolas Cointe, Grégory Bonnet, and Olivier Boissier. Ethical judgment of agents' behaviors in multi-agent systems. AAMAS '16, page 1106–1114, Richland, SC, 2016.
- **[Cryder and Loewenstein, 2011]** Cynthia Cryder and George Loewenstein. The critical link between tangibility and generosity. In Society for Judgment and Decision Making series. The science of giving: Experimental approaches to the study of charity, pages 237–251. Psychology Press, 2011.
- **[Danielson, 1992]** Peter Danielson. Artificial Morality: Virtuous Robots for Virtual Games. Routledge, 1992.
- **[Dreber et al., 2008]** Anna Dreber, David G. Rand, Drew Fudenberg, and Martin A. Nowak. Winners don't punish. Nature, 452(7185):348–351, 2008.
- **[Galak and Chow, 2019]** Jeff Galak and Rosalind M. Chow. Compensate a little, but punish a lot: Asymmetric routes to restoring justice. PLOS ONE, 14(1), 2019. https: doi.org/10.1371/journal.pone.0210676.
- **[Gaudou *et al.*, 2014]** Benoit Gaudou, Emiliano Lorini, and Eunate Mayor. Moral Guilt: An Agent-Based Model Analysis. In Advances in Social Simulation, Advances in Intelligent Systems and Computing, pages 95–106, Berlin, Heidelberg, 2014. Springer.
- **[Kidder, 2009]** Rushworth Moulton Kidder. How Good People Make Tough Choices Rev Ed: Resolving the Dilemmas of Ethical Living. HarperCollins, November 2009.
- **[Korb et al., 2010]** Kevin B. Korb, Ann E. Nicholson, and Owen Woodberry. Evolving Ethics: The New Science of Good and Evil. Imprint Academic, 2010.
- **[Moussaïd *et al.*, 2013]** Mehdi Moussaïd, Juliane E. Kämmer, Pantelis Pipergias Analytis, and Hansjörg Neth. Social influence and the collective dynamics of opinion formation. *PLOS ONE*, 8(11):1–8, 11 2013.
- **[Rand and Nowak, 2011]** David G. Rand and Martin Andreas Nowak. The evolution of antisocial punishment in optional public goods games. Nature Communications, 2(1):1–7, 2011.
- **[Smith, 1982]** Christopher Upham Murray Smith. Evolution and the problem of mind: Part 1. Herbert Spencer. Journal of the History of Biology, 15(1):55–88, 1982.
- **[Vandeviver and Bernasco, 2019]** Christophe Vandeviver and Wim Bernasco. "location, location, location": Effects of neighborhood and house attributes on burglars' target selection. *Journal of Quantitative Criminology*, pages 1–43, 2019.
- **[Wiegel and van den Berg, 2009]** Vincent Wiegel and Jan van den Berg. Combining Moral Theory, Modal Logic and Mas to Create Well-Behaving Artificial Agents. *International Journal of Social Robotics*, 1(3):233–242, August 2009.