Designed to Cooperate: A Kant-Inspired Ethic of Machine-to-Machine Cooperation

Seng W. Loke
Centre for Internet of Things ECOsystems Research and Experimentation (CITECORE)
School of Information Technology, Deakin University, Geelong, Australia
seng.loke@deakin.edu.au

***Abstract*—We envision an increasing presence of devices with agency and autonomous machines in public spaces (e.g., automated vehicles, urban robots and drones) beyond the confines of constrained environments such as a factory floor or research labs. Hence, AI and robotic systems of the future will need to interact with one another, not only in cyber space but also in physical space, and need to behave appropriately in their interactions with one another.**

This position paper highlights an ethic of machine-to-machine cooperation and machine pro-sociality, and argues that machines capable of autonomous sensing, decision-making and action, such as automated vehicles and urban robots, owned and used by different self-interested parties, and having their own agendas (or interests of their owners) should be designed and built to be cooperative in their behaviours, especially if they share public spaces. That is, by design, the machine should first cooperate, and then only consider alternatives if there are problems.

It is argued that being cooperative is not only important for their improved functioning, especially, when they use shared resources (e.g., parking spaces, public roads, curbside space and walkways), but also as a favourable requirement analogous to how humans cooperating with other humans can be advantageous and often viewed favourably. The usefulness of such machine-to-machine cooperation are illustrated via examples including cooperative crowdsourcing, cooperative traffic routing and parking as well as futuristic scenarios involving urban robots for delivery and shopping.

It is argued that just as privacy-by-design and security-by-design are important considerations, in order to yield systems that fulfil ethical requirements, cooperative-by-design should also be an imperative for autonomous systems that are separately owned but co-inhabit the same spaces and use common resources. If a machine using shared public spaces is not cooperative, as one might expect, then it is not only anti-social but not behaving ethically. It is also proposed that certification for urban robots that operate in public could be explored.

I. Cooperation as an Aspect of Behaving Ethically

Busy urban spaces can become places of tremendous competition among users. For example, people effectively "compete" for parking spaces, or for road use to get from one point to another in the fastest possible way, or with the rise of urban robots, urban robots might compete for space for their operations. For example, urban robots from different companies doing deliveries might compete for space on walkways with other urban robots and people. Two or more advertising robots might compete for the attention of a passer-by. Even robots within the same building might compete with each other for resources while performing tasks or get in the way of each other.

In this paper, we mostly use the general meaning of the term "cooperation", i.e., "doing something together" or "working together toward a shared aim."[1] We consider a Kant-inspired view of cooperation later in the paper.

However, many (even if not all) such problems can be alleviated by cooperation. For example, road spaces and parking spaces are limited shared resources in many cities - it was shown that if vehicles could help each other find parking, each vehicle could benefit, or if groups of vehicles cooperate at road junctions in computing their routes, they could all get to their destinations earlier reducing traffic jams [9]. Urban delivery robots could cooperate with each other to ensure safe and efficient use of walkways, instead of trying to out-manoeuvre each other. Vehicles and urban robots could cooperate in sharing information they capture/sense during their journeys, thereby creating a knowledge base for all to benefit from.

Machine-with-machine cooperation in cyber-space via well-defined protocols is happening everyday in the Internet and the Web with the billions of communicating networked devices. Machine-with-machine cooperation (among autonomous machines) in physical space is, however, a consideration only beginning to emerge with the advent of automated vehicles in cities and urban robotics [16], [18]. With an increasing number of (potentially autonomous) machines in public, e.g., automated vehicles and urban robots such as delivery robots on walkways or robots in shopping centres and public spaces, there is then a growing need for such machines to not only be sensitive to and relate to humans in public spaces, behaving in socially appropriate ways, but also to cooperate and relate to each other in public spaces.

It is noted in [11] that urban robots occupy physical space, and hence, "potentially require the rearrangement and redesign of existing spatial entities". This is significant in cities, especially those with high density, where space is a valuable and scarce resource and there are different land use considerations. It is argued in this paper that cooperation is an important means of facilitating efficient use of different types of spaces. Robots in cities are not always welcomed, e.g., the delivery robots that use sidewalks in San Francisco

[1]From: https://www.oxfordlearnersdictionaries.com/definition/american_english/cooperation

were severely restricted in their operations.[2] Hence, deploying machines in public spaces can raise issues of space contention, not only with humans but with each other.

Robots that cooperate have been studied in swarm robotics and Internet-based robotics, in particular, where such robots are under the same authority or controller or owner. But one could envision separately owned autonomous machines that occupy physical spaces in the city and use shared urban resources such as public roads, shared walkways and public spaces, each with its own (or its owner's) goals, interests and agendas. Regulations can help govern the behaviour of such machines with respect to shared public resources, but, as is argued here, being aware of each other and cooperating with each other, even helping out each other, becomes an imperative, not only for their harmonious operation, but also to improve their workings, as we will illustrate further.

This paper proposes the idea that machines that use shared resources (e.g., roads and public spaces in the city) should be designed to be cooperative with each other, not only with humans, and that such machines are behaving ethically when "first try to cooperate" is their "ethos", by design.

II. Lessons from Game Theory

Towards this direction, one can consider the notion of the *Kantian cooperative machine*, based on the concept of Kantian cooperation [15], that puts the question of "us" as a priority in its decision-making, i.e. if each machine reason in a Kantian manner and ask the same question: "what strategy(action) would I want all of us to play(take)?" rather than the question "what is the best strategy(action) for me to play(take), given others' strategies(actions)?". This notion is not entirely new, considering the Kantian machine discussed in [14], where a machine should "act only according to that maxim whereby you can at the same time will that it should become a universal law", i.e., "Kant tells the moral agent to test each maxim (or plan of action) as though it were a candidate for a universalized rule." There are examples from game theory, e.g., the Stag Hunt game, that illustrate in theory the value of cooperation for achieving Pareto optimality.

In addition, a default general rule built into a machine such as "cooperate when you can and where appropriate" seems itself to be a good candidate for one such universal rule.

The inspiration from Kant is in two ways: (i) cooperation itself is a good candidate for an universalisable maxim (i.e., "*cooperate*" could be willed as *a strategy for everyone*), and (ii) cooperation can be viewed as a strategy of universalisation (i.e. that is, "to cooperate" is simply *to consider what strategy everyone ought to take*, where the strategy here depends on the application or domain being considered).

[2]https://www.bbc.com/news/technology-42265048, though there seems more freedom for deployment in more recent times: https://www.aitrends.com/robotics/last-mile-delivery-robots-making-a-comeback-after-initial-bans/

III. Lessons from Urban Applications: Machines Cooperating on Shared Urban Resources

One can consider a range of scenarios for machines that are intended to navigate and move autonomously using shared public infrastructure.

- **Automated Vehicles Cooperating in Routing and Parking.** Vehicle-to-vehicle communications (e.g., cooperative-ITS) have been well researched and there are emerging standards (e.g., ETSI[3] and SAE[4]) for formats of messages and protocols for cars to communicate with one another in a variety of contexts, not only to issue warnings for safety, e.g., as a car emerges from a parking space, at intersections or at merging lanes, but also to help vehicles cooperate, e.g., to aid platooning. More recent work is also looking at higher levels of cooperation, instead of simple message exchanges, when finding car parking, or routing to avoid congestion. Two companies are trialing of fleets of robotaxis to serve a particular urban areas, though such taxis may not cooperate in the direct vehicle-to-vehicle communication sense, they are coordinated centrally.[5]
- **Urban Robots Cooperating.** Urban robotics have been proposed not only for delivery [18], but also for cleaning streets and waste management. In public retail spaces such as shopping centres, airports, and hotels, there have been deployment of such robots and trials. Police robots have also been deployed in Singapore.[6] There are also security robots playing the role of security guards.[7] The work in [1] proposes the use of swarms of urban robots to collect wastes in the city - such robots can utilise bike lanes. Such robots can cooperate with each other or via a central controller in their tasks.
  As such robots occupy valuable urban public real estate, there is a need for their cooperation and coordination - cleaning robots, robots helping to carry shopping, policing robots and delivery robots are, if they proliferate, enough, to cause congestion on sidewalks and public spaces. Also, as mentioned, some types of urban robots are not necessarily welcomed in places predominantly for humans. Delivery robots have encountered issues in urban environments, as they physically occupy spaces. Similarly, security robots have been bullied and forced off the street.[8] Hence, there is a need for coordination and/or cooperation of different types of robots if they are to occupy public spaces traditionally just for humans - perhaps their use of certain areas need to be coordinated

[3]https://www.etsi.org/technologies/automotive-intelligent-transport

[4]http://www.sae.org/servlets/pressRoom?OBJECT_TYPE=PressReleases&PAGE=showRelease&RELEASE_ID=3343

[5]https://techcrunch.com/2020/08/16/autox-launches-its-robotaxi-service-in-shanghai-competing-with-didis-pilot-program/

[6]E.g., https://www.straitstimes.com/singapore/transport/robot-traffic-cop-spotted-at-changi-airport,https://mothership.sg/2020/05/police-matar-robots/

[7]https://www.knightscope.com

[8]E.g./, see https://www.dezeen.com/2017/12/13/k5-knightscope-security-robot-sfspca-san-francisco-bullied-off-street/

or the robots should cooperate to ensure certain constraints on public spaces are satisfied, while performing their functions.

- **Machines Contributing to Crowdsourced Knowledge Bases.** Crowdsourcing involves aggregating contributions from many to solve problems. In recent times, the idea of crowdsourcing for sensor data has given rise to the concept of crowdsensing, where contributions of mobile sensor data (including geo-tagged pictures, videos, location and other context data) are aggregated to create a overall map or knowledge base of a particular phenomenon. For example, vehicles could collaborate to upload videos to provide a map of where cherry blossoms are in a city [10]. There are many other uses of crowdsensing, from crowdsourcing in order to create maps of 4G/5G bandwidth variations to creating maps of where the crowds are in the city, traffic variations, or temperature [3], [4], [13]. One can imagine machines, whether automated vehicles or urban robots, that roam the streets for their own purposes, but still being able to provide data towards building real-time urban maps for different purposes, i.e. such machines become contributors towards real-time collective urban knowledge. Such crowdsensing can help inform urban design, e.g., as noted in [19].

The examples above motivates cooperation, or at least, coordinated actions among machines.

IV. Lessons from Multiagent Research: Evolving Cooperation and Learning from Each Other

There is already interesting work on multiagent reinforcement learning where agents learn to optimize not only for their own "self-interest" but the welfare of a group, or to achieve socially optimal outcomes, e.g., [20]. Allowing agents to influence each other during learning can lead to better collective outcomes [7]. Social learning is important to enable agents to learn from each other and to improve themselves, e.g. [12], [6]. Even autonomous vehicles could beneficially learn from other autonomous vehicles [21].

Machines that use pubic spaces should perhaps have basic ability to learn from each other, and to pick up behavioural "cues" from each other, appropriate in a given situation.

V. Certified Pro-social and Cooperative

How can we ensure that robots deployed to roam our city streets have been designed and built to be cooperative and pro-social? Is there a way to certify that a given robot will be cooperative, at least within the scope of its core functions and operational domain?

For example, when a delivery robot encounters a cleaning robot, they will make way for each other (perhaps according to some pre-programmed manner, and based on some regulatory operational requirements of the local city government), robots from different manufacturers (and belonging to different owners) know how to move in a single file when journeying through a narrower part of a walkway (even if this means, to an extent, increasing their own travel time) allowing space for pedestrians, robots (separately owned and each with its own agenda) can cooperate in their routing and movement so as not to cause congestion on walkways and public spaces, urban robots might warn each other of hazards encountered, and when a robot breaks down or gets into a dangerous position, other robots come to its aid (which can be in different ways depending on the situation, the type of fault, and the robots' capabilities).

Even for automated vehicles to exchange context messages, e.g., for safety or platooning, standards are needed and a range of tests required to validate that they conform to particular basic functionalities. Certification of automated vehicles has been discussed,[9] and could involve the automated vehicle passing a "(self-)driving test" under different operating situations or scenarios. Testing of automated vehicles via simulations has also been considered (e.g., see [17], [2]).[10] Testing and validation for pro-sociality and cooperative behaviours for robots could also be carried out, at least to an extent, in a similar way, though it is not clear how complete one can be in terms of the range of testing situations - it is likely not possible to be exhaustive, i.e. to anticipate all possible scenarios that the robots might find itself in.[11]

VI. Conclusion

As machines interact with humans and other machines in shared public spaces, analogous to how humans live together, machines should be designed to (learn to) "live" with each other, and not only humans. Machines can be designed to be cooperative, according to their capabilities, not only to when performing their core functions, but also able to learn from each other (up to what a machine can reasonably be programmed to learn) to behave appropriately in specific contexts. It would be ethical to design such machines to be cooperative, and in being cooperative, as appropriate to the situation, is one aspect of the machines behaving ethically. There could be myriad solutions - for example, there could be a city platform from which robots operating within that city or neighbourhood download "cooperation rules" that enable the robots to cooperate in a range of situations. And to be allowed to operate in open shared urban spaces, the robot should somehow first be certified "pro-social and cooperative".

Much future work remains, in order to create the pro-social and cooperative robots that operate in public spaces shared with other robots and humans, not only the mechanisms, but also the validation of such robots. Work is also needed to determine what would be the basic cooperative requirements on specific types of robots that operate in public. Also, being predisposed to cooperate is not without risks (e.g., even as the Stag Hunt game would illustrate) since the other party might not cooperate (due to fault or malice) and the one

[9] For example, see [8], and https://unece.org/DAM/trans/doc/2019/wp29grva/GRVA-02-09e.pdf

[10] See https://www.ansys.com/applications/autonomous-vehicle-validation and https://www.dlr.de/content/en/articles/news/2021/02/20210430_the-set-level-project.html

[11] A large database of scenarios and situations for automated driving systems testing is being considered: https://www.safetypool.ai

trying to cooperate might end up disadvantaged, or might be taken advantage of in the situation. There is also the issue of the difference between rational cooperation and ethical cooperation - we have outlined scenarios where it is rational to cooperate, and perhaps from a utilitarian point of view, it would be ethical to cooperate. However, in some situations, it might be ethical to cooperate but costly for the individuals to do so. Future work is required to address these issues.

Also, future work can perhaps draw lessons for machine-to-machine cooperation by considering work from human-to-human cooperation, e.g., the work on *morality-as-cooperation* might provide a pathway towards developing moral machines; to quote from [5]: "...morality as cooperation predicts that people will regard specific types of cooperative behaviour— behaviour that solves some problem of cooperation—as morally good." That is, machines that cooperate can in some way be viewed as moral machines - further thought is required for this.

It is also unclear if the Kantian notion of cooperation here, which asks the question "what strategy would I want all of us to play?", can be employed in human-to-machine cooperation, involving a collection of machines and humans. A machine might not have the ability to answer this question adequately. Future work can address this question.

Acknowledgements

The author would like to thank the anonymous reviewers for the many insightful comments and helping to improve the paper.

References

[1] A. L. Alfeo, E. C. Ferrer, Y. L. Carrillo, A. Grignard, L. A. Pastor, D. T. Sleeper, M. G. C. A. Cimino, B. Lepri, G. Vaglini, K. Larson, M. Dorigo, and A. Pentland. Urban swarms: A new approach for autonomous waste management. In *2019 International Conference on Robotics and Automation (ICRA)*, pages 4233–4240, 2019.

[2] Vimal Rau Aparow, Apratim Choudary, Giridharan Kulandaivelu, Thomas Webster, Justin Dauwels, and Niels de Boer. A comprehensive simulation platform for testing autonomous vehicles in 3d virtual environment. In *2019 IEEE 5th International Conference on Mechatronics System and Robots (ICMSR)*, pages 115–119, 2019.

[3] Rim Ben Messaoud. *Towards efficient mobile crowdsensing assignment and uploading schemes*. Theses, Université Paris-Est, July 2017.

[4] A. Capponi, C. Fiandrino, B. Kantarci, L. Foschini, D. Kliazovich, and P. Bouvry. A survey on mobile crowdsensing systems: Challenges, solutions, and opportunities. *IEEE Communications Surveys Tutorials*, 21(3):2419–2465, 2019.

[5] Oliver Scott Curry. *Morality as Cooperation: A Problem-Centred Approach*, pages 27–51. Springer International Publishing, Cham, 2016.

[6] Jianye Hao, Ho-Fung Leung, and Zhong Ming. Multiagent reinforcement social learning toward coordination in cooperative multiagent systems. *ACM Trans. Auton. Adapt. Syst.*, 9(4), December 2014.

[7] Natasha Jaques, Angeliki Lazaridou, Edward Hughes, Çaglar Gülçehre, Pedro A. Ortega, DJ Strouse, Joel Z. Leibo, and Nando de Freitas. Social influence as intrinsic motivation for multi-agent deep reinforcement learning. In Kamalika Chaudhuri and Ruslan Salakhutdinov, editors, *Proceedings of the 36th International Conference on Machine Learning, ICML 2019, 9-15 June 2019, Long Beach, California, USA*, volume 97 of *Proceedings of Machine Learning Research*, pages 3040–3049. PMLR, 2019.

[8] L.V. Legashev, T.V. Letuta, P.N. Polezhaev, A.E. Shukhman, and Yu.A. Ushakov. Monitoring, certification and verification of autonomous robots and intelligent systems: Technical and legal approaches. *Procedia Computer Science*, 150:544–551, 2019.

[9] Seng W. Loke. Cooperative automated vehicles: A review of opportunities and challenges in socially intelligent vehicles beyond networking. *IEEE Transactions on Intelligent Vehicles*, 4(4):509–518, 2019.

[10] Shigeya Morishita, Shogo Maenaka, Daichi Nagata, Morihiko Tamai, Keiichi Yasumoto, Toshinobu Fukukura, and Keita Sato. Sakurasensor: Quasi-realtime cherry-lined roads detection through participatory video sensing by cars. In *Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing*, UbiComp '15, page 695–705, New York, NY, USA, 2015. Association for Computing Machinery.

[11] Michael Nagenborg. Urban robotics and responsible urban innovation. *Ethics and Information Technology*, 2018.

[12] Kamal K Ndousse, Douglas Eck, Sergey Levine, and Natasha Jaques. Emergent social learning via multi-agent reinforcement learning. In Marina Meila and Tong Zhang, editors, *Proceedings of the 38th International Conference on Machine Learning*, volume 139 of *Proceedings of Machine Learning Research*, pages 7991–8004. PMLR, 18–24 Jul 2021.

[13] J. Phuttharak and S. W. Loke. A review of mobile crowdsourcing architectures and challenges: Toward crowd-empowered internet-of-things. *IEEE Access*, 7:304–324, 2019.

[14] Thomas M. Powers. Prospects for a kantian machine. *IEEE Intelligent Systems*, 21(4):46–51, July 2006.

[15] John E. Roemer. *How We Cooperate: A Theory of Kantian Optimization*. Yale University Press, 2019.

[16] Pericle Salvini. Urban robotics: Towards responsible innovations for our cities. *Robotics and Autonomous Systems*, 100:278 – 286, 2018.

[17] Hans-Peter Schoener. The role of simulation in development and testing of autonomous vehicles. 09 2017.

[18] Aidan H While, Simon Marvin, and Mateja Kovacic. Urban robotic experimentation: San Francisco, Tokyo and Dubai. *Urban Studies*, 2020. https://doi.org/10.1177

[19] Shili Xiang, Lu Li, Si Min Lo, and Xiaoli Li. People-centric mobile crowdsensing platform for urban design. In Gao Cong, Wen-Chih Peng, Wei Emma Zhang, Chengliang Li, and Aixin Sun, editors, *Advanced Data Mining and Applications*, pages 569–581, Cham, 2017. Springer International Publishing.

[20] Chengwei Zhang, Xiaohong Li, Jianye Hao, Siqi Chen, Karl Tuyls, Wanli Xue, and Zhiyong Feng. Sa-iga: a multiagent reinforcement learning method towards socially optimal outcomes. *Autonomous Agents and Multi-Agent Systems*, 33(4):403–429, 2019.

[21] Jimuyang Zhang and Eshed Ohn-Bar. Learning by watching. *CoRR*, abs/2106.05966, 2021.