Rational or emotional? Alan Turing's "Heads in the Sand" Objection and the discussion of Autonomous Vehicles

Karol Wapniarski[1,2][0009−0001−4877−9370] and
Paweł Łupkowski[1,3][0000−0002−5335−2988]

[1] Adam Mickiewicz University, Poznań, Poland
[2] karwap@st.amu.edu.pl
[3] pawel.lupkowski@amu.edu.pl

Abstract

Autonomous Vehicles (AV) have been a vividly discussed topic from 2016 to 2018. The rapid growth of AI technology in previous years has enabled cars to drive themselves without needing a driver. Instead of receiving a glorious welcome, AVs have raised many concerns among academics and the general public, with most of them relating to what the car should do in the event of an unavoidable accident and various ethical considerations surrounding it. This paper aims to provide an account of the AV discussion as seen through the lenses of Alan Turing's "Heads in the Sand" Objection, arguing that the fundamental nature of these discussions is not rational but emotional, and thus, from the argumentative point of view, they are both unresolvable and irrelevant when it comes to deriving far-fetched philosophical conclusions.

The "Heads in the Sand" Objection was presented by Alan Turing in his 1950 paper [11, p. 444] as one of the objections that can be raised against the development of AI. Turing phrases and comments on this objection as follows:

> The consequences of machines thinking would be too dreadful. Let us hope and believe that they cannot do so. [...] This argument is seldom expressed quite so openly as in the form above. But it affects most of us who think about it at all. [...] I do not think that this argument is sufficiently substantial to require refutation.

The Objection thus claims that the consequences of having machine thinking would be too dreadful, and thus, it is better for us to believe it to be impossible. Turing response is that as the argument is based on and appeals to emotions, it needs not to be refuted. However, as such, it "affects most of us who think about it at all", and Turing [11, p. 444] also identifies the source of this affection: "We like to believe that Man is in some subtle way superior to the rest of creation. It is best if he can be shown to be necessarily superior, for then there is no danger of him losing his commanding position". It is therefore not questioned that the Objection can play a role as a background thought and should be present, but it

is said that it should not guide our theoretical line of thinking, being categorically different from it.

In this paper, we take a closer look at the discussion concerning Autonomous Vehicles which has been visibly active in the years 2016-2018. We aim to show that the debate can be treated as an example, or an extended version, of the "Heads in the Sand" Objection. Our main thesis is that the AV discussion is emotional in nature: that the emotional aspects, and not the rational ones, are what drive the public, whether in academia or not. We show and analyze portions of both academic [1,2,4,6–8,12] and popular [3,5,9,10] literature and show that even in the case of academia, the emotionally based philosophical considerations remain unresolved (being the "Trolley Problem revisited" and remaining unresolvable even based on a social contract [1]) and are finally reduced to more pragmatic issues, comprising law regulations [2], economic rationality [2], and technical improvements [12]. We then propose to look at the present discussions concerning Chat-GPT as a validation of this view, maintaining that with the Chat-GPT coming, the discussion of AV, as an emotional one, was abandoned and turned to the ethical problems posed by the chatbot. Here, we stress that in the case of the chatbot, Turing's comment claiming that the ultimate source of fear of AI comes from humans being scared of losing their cognitively superior position becomes especially valid. We conclude by proposing that the identified formal structure of unresolvable argumentation-reductionism-abandonment will be again iterated with the Chat-GPT discussion itself, with some even never AI technology being developed and introduced.

References

1. Bonnefon, J.-F., Shariff, A., & Rahwan, I.: The social dilemma of autonomous vehicles. *Science* **352**, 1573–1576 (2016)
2. Casey, B.: Amoral machines, or: How roboticists can learn to stop worrying and love the law. *Northwestern University Law Review* **111**, 1347–1366 (2017)
3. Donde, J.: Self-Driving Cars Will Kill People. Who Decides Who Dies? *WIRED*, 21 September 2017. https://www.wired.com/story/self-driving-cars-will-kill-people-who-decides-who-dies/
4. Gogoll, J. & Müller, J.-F.: Autonomous cars: In favor of mandatory ethics setting. *Science and Engineering Ethics* **23**, 681–700 (2016)
5. Hern, A.: Self-driving cars don't care about your moral dilemmas. *The Guardian*, 22 August 2016. https://www.theguardian.com/technology/2016/aug/22/self-driving-cars-moral-dilemmas
6. Keeling, G.: Why trolley problems matter for the ethics of automated vehicles. *Science and Engineering Ethics* **26**, 293–307 (2020)
7. Lawlor, R.: The Ethics of Automated Vehicles: Why Self-Driving Cars Should not Swerve in Dilemma Cases. *Res Publica* **28**, 193–216 (2021)
8. Markoff, P.J.: *Machines of Loving Grace: The Quest for Common Ground Between Humans and Robots.* Ecco Press (2016)
9. Marshal, A.: Lawyers, Not Ethicists, Will Solve The Robocar "Trolley Problem". *WIRED*, 28 May 2017. https://www.wired.com/2017/05/autonomous-vehicles-trolley-problem/

10. Metz, C.: Self-Driving Cars Will Teach Themselves to Save Lives – But Also Take Them. *WIRED*, 9 June 2016. https://www.wired.com/2016/06/self-driving-cars-will-power-kill-wont-conscience/
11. Turing, A.: Computing machinery and intelligence. *Mind* **59**, 433–460 (1950)
12. Wendel, W.B.: Economic Rationality and Ethical Values in Design-Defect Analysis: The Trolley Problem and Autonomous Vehicles. *California Western Law Review* **55**, 129–163 (2018)