

Moralische Maschinen ans Steuer?

Posted on 20. November 2018 by Klaus F. Röhl

Jeder Student wird bereits in der ersten Strafrechtsvorlesung mit dem Notstandsproblem konfrontiert, das heute international als Trolley-Problem geläufig ist: Ein Waggon rollt führerlos auf eine Weiche zu und wird mit einem entgegenkommenden Personenzug zusammenstoßen, wenn nicht der Beobachter im Stellwerk die Weiche stellt. Dann wird der Waggon jedoch auf dem anderen Gleis einen Arbeiter überrollen. Der Handelnde hat also die Wahl, ob er durch sein Eingreifen einen Menschen tötet, um mehrere Menschen zu retten. Wie auch immer er sich entscheidet; er handelt rechtswidrig, denn das Menschenleben als Rechtsgut ist nicht abwägungsfähig. Entscheidet er sich, die Weiche zu stellen, so wird ihm ein übergesetzlicher entschuldigender Notstand zugebilligt. Bleibt er untätig, so handelt er trotzdem im Sinne einer Unterlassung. Offen ist die Frage, ob er handeln muss, um das größere Unheil zu verhindern. Dazu müsste man ihm eine Garantenstellung zumuten. Die verträgt sich aber nicht mit dem Abwägungsverbot. Die Entscheidung bleibt schwierig. Eben deshalb handelt es sich um ein Dilemma.

Dieses Problem stellt sich heute – so jedenfalls scheint es – bei der Programmierung autonomer Fahrzeuge. Sollen bei einer drohenden Kollision eher die Insassen oder die Passagiere anderer Fahrzeuge oder Fußgänger gerettet werden, eher Frauen als Männer, eher Junge als Alte, eher Gesunde als Kranke, eher Menschen als Tiere? Juristisch gibt es eine klare Antwort nur für die Alternative Mensch oder Tier.

Es liegt nahe, in dieser Situation nach Volkes Meinung zu fragen. Das haben Wissenschaftler vom Massachusetts Institute of Technology getan.^[1] Sie haben ins Internet gestellt, was sie eine [Moral Machine](#) nennen. Dort konnten Internetnutzer aus aller Welt für 13 Szenarien ihre Präferenzen angeben. Es soll 40 Millionen Teilnehmer gegeben haben. Dabei zeigten sich weltweit drei Präferenzen: Menschen vor Tieren, Mehrzahl vor Einzahl und jung vor alt. Die Präferenzen unterschieden sich etwas nach der Herkunftsregion. So war im Fernen Osten die Präferenz für das Alter höher, während in Süd- und Mittelamerika Jugend, Frauen und Tiere höher geschätzt wurden. Wenn man die Moralmaschine ans Steuer lässt, fährt sie also einen Schlingerkurs.

Das Problem scheint eine gewisse Ähnlichkeit mit der Verteilungssituation zu haben, die als Triage bekannt ist.^[2] Der Begriff stammt aus der Militär- und

Katastrophenmedizin. Wenn kurzfristig großer Andrang von Verwundeten oder Kranken entsteht, die mit den vorhandenen Mitteln nicht versorgt werden können, so erfolgt eine »Sichtung« und Einteilung in drei Gruppen (daher der Name). Die Schwerstverletzten mit geringen Überlebenschancen werden abgesondert und erhalten allenfalls noch Schmerzlinderung und Trost. Leichtverletzte, die auch später noch behandelt werden können, müssen warten. Die Behandlung konzentriert sich auf diejenigen, deren Versorgung besonders dringlich und zugleich erfolgversprechend ist. Die Situation der Triage ist in der medizinischen Versorgung längst zum Alltagsproblem geworden. Für Dialyse, Organtransplantation oder Intensivmedizin ergeben sich ständig Knappheitssituationen, in denen die Behandlung mehr oder weniger verdeckt nach dem *social worth* der Betroffenen zugeteilt wird. Bei der Triage verfährt man utilitaristisch, das heißt, die juristischen Abwägungsverbote gelten hier nicht. Der soziale Wert der Betroffenen wird quantitativ beurteilt. Das heißt, einer muss vor mehreren zurückstehen; gesund geht vor krank und jung vor alt.

Zu programmieren ist eine Abwägung unter Unsicherheit. Unsicher sind zunächst die empirischen Prämissen. Wer oder was kann mit welcher Wahrscheinlichkeit gerettet werden? Unsicher sind aber auch die normativen Prämissen? Juristen haben für diesen Fall tolle Formeln entwickelt.^[3] Programmierer werden darüber aber wohl eher in Gelächter ausbrechen. Es spricht einiges für die Einschätzung der Ethikkommission des Bundes.^[4] Sie bezweifelt, dass die ethischen Dilemmata, um die es hier geht, überhaupt normierbar und programmierbar seien, und kommt zu dem Schluss:

»Bei unausweichlichen Unfallsituationen ist jede Qualifizierung nach persönlichen Merkmalen (Alter, Geschlecht, körperliche oder geistige Konstitution) strikt untersagt. Eine Aufrechnung von Opfern ist untersagt. Eine allgemeine Programmierung auf eine Minderung der Zahl von Personenschäden kann vertretbar sein. Die an der Erzeugung von Mobilitätsrisiken Beteiligten dürfen Unbeteiligte nicht opfern.«

Weder Trolley-Problem noch Triage passen genau auf den fahrenden Automaten. Hier treffen nämlich Gefährder, Gefährdeter und zur Handlung Aufgerufener zusammen. Sieht man Fahrzeug und Insassen als Einheit, so wäre an einen Notstand zur Selbstrettung nach § 35 StGB zu denken. Dann kommt es darauf an, ob der Handelnde die Gefahr »selbst verursacht« hat. Nackte Kausalität genügt hier zum Ausschluss der Entschuldigung nicht, volles Verschulden, wie früher nach § 54 a. F., ist aber auch nicht erforderlich. Ich bin ziemlich sicher, dass die

Programmierung automatisierter Fahrzeuge darauf hinausläuft, zuerst die Insassen des Fahrzeugs zu retten, koste es was es wolle.

Nachtrag: Susanne Beck, Autonomes Fahren: Herausforderung für das bestehende Rechtssystem, Informatik aktuell, 20. 8. 2018, [online](#). Weitere einschlägige Veröffentlichungen der Hannoveraner Strafrechtsprofessorin werden in deren auch sonst über das Strafrecht hinaus interessanten [Schriftenverzeichnis](#) angeführt.

[1] Edmond Awad u. a., [The Moral Machine Experiment](#), Nature 563, 2018, 59-64.

[2] Volker H. Schmidt, Veralltäglicung der Triage, Zeitschrift für Soziologie 25, 1996, 419-437; Weyma Lübke, Veralltäglicung der Triage?, Ethik in der Medizin 13, 2001, 148-160.

[3] Matthias Klatt/Johannes Schmidt, Abwägung unter Unsicherheit, AöR 137, 2012, 545-591; Justus Quecke/Jan Sturm, Unsicherheit über Abwägung, Rechtstheorie 45, 2014, 113-131.

[4] Bericht der Ethikkommission, [Automatisiertes und Vernetztes Fahren](#), 2017.

Ähnliche Themen

- [Moralische Maschinen 2.0](#)