Abstract The recent progress in the development of autonomous cars has seen ethical questions come to the forefront. In particular, life and death decisions regarding the behavior of self-driving cars in trolley dilemma situations are attracting widespread interest in the recent debate. In this essay we want to ask whether we should implement a mandatory ethics setting (MES) for the whole of society or, whether every driver should have the choice to select his own personal ethics setting (PES). While the consensus view seems to be that people would not be willing to use an automated car that might sacrifice themselves in a dilemma situation, we put forward the claim that an ethics setting that minimizes harm is in the considered interest of everybody. A MES would be in the general interest, since a PES regime would most likely result in a prisoner's dilemma.

The introduction of autonomous cars as well as the development of ever more capable driver assistance systems are moving at a high pace. Big companies like BMW, Mercedes, Ford, GM, Toyota, Nissan, Volvo, Audi and, most prominently, Google are currently working on projects that aim to get humans away from the steering wheel. Tesla has even gone so far as to release an update that enables their cars to drive on autopilot (McHugh 2015).
Although on the one hand, there is – from a normative standpoint – pro tanto good reason to welcome the introduction of autonomous cars, there is no doubt that automated driving also poses new ethical challenges. Self-driving cars – if introduced – will crash eventually and will kill or seriously hurt someone in the process. There has never been a technology that has not failed at one point, and self-driving cars will be no exception. Autonomous cars are highly dependent on software and sensors, which are prone to fail eventually. Yet, even if we assume that a malfunction of the system does not occur, unlucky circumstances might lead to the following situation:

> *Imagine you are sitting in your autonomous car going at a steady pace entering a tunnel. In front of you is a school bus with children on board going at the same pace as you are. In the left lane there is a single car with two passengers overtaking you. For some reason the bus in front of you brakes and your car cannot brake to avoid crashing into the bus. There are three different strategies your car can follow: First, brake and crash into the bus, which will result in the loss of lives on the bus. Second, steer into the passing car on your left – pushing it into the wall, saving your life but killing the other car's two passengers. Third, it can steer itself (and you) into the right hand sidewall of the tunnel, sacrificing you but sparing all other participants' lives.*

In a world without autonomous cars, the tunnel case is a philosophically interesting problem, which is usually discussed in the literature under the rubric of 'trolley problems', but not an ethically relevant "real world" issue. The reason for this is mainly that the driver behind the bus needs to make a split second decision based on very limited information. In such a situation, there is simply no time to form a – what philosophers sometimes call – deliberate judgment and, thus, there are thin grounds for assigning responsibility. In a world with autonomous cars, the case is different. Here, an agent – for instance the driver of a particular car or a regulative agency – essentially needs to tell the car beforehand what it should do in such a case. Or to put it differently: an agent must decide for a specific ethics setting. From a normative perspective, this raises an immediate question, namely: What is the right ethics setting? In this essay, however, we want to deal with another – although related – normative question: Should we collectively mandate a specific ethics setting for the whole of society, or should every driver have the choice to select his own ethics setting?

We argue that the default option in liberal societies to deal with moral disagreement is to partition the moral decision space in order to enable each individual to live according to her own normative ideals and understanding of the good and thus to respect individual autonomy (within limits). Applied to the case of autonomous cars this would allow for a personal ethics setting (PES). However, allowing for a PES, we argued, will likely lead to a situation that has the structure of a prisoner's dilemma. The incentives for the individual will crowd out a moral PES, which is set to value one's own life less than other lives at least under some circumstances, and drive people to choose a selfish PES. The result of this situation, so we argue, is that everybody (the moral as well as the selfish agents) is worse off compared to a mandatory rule that is enforced by a third party. While the consensus view seems to be that people would not be willing to use an automated car that might sacrifice themselves in a dilemma situation, we argued that such a MES is in the considered interest of everybody. Since informal sanctions in anonymous large societies do not possess the force needed to prevent the individual to choose a selfish PES, we advocate for a mandatory rule that aims at minimizing overall harm. State regulation seems to be the most obvious as well as practical way to achieve that. Furthermore, we made the case that the classic trolley problem is conceptually inadequate for discussing the case of ethics settings. To acknowledge the dynamics of an institution like traffic, we argued that it is of the utmost importance to realize that the decision about which ethics setting to choose is strategic in nature, iterated as well as dependent on what position in a dilemma an individual might find herself in.