The ethical implications of driverless cars disrupting the motor industry

In recent years, driverless cars have moved from an idea firmly planted in science fiction into something much more real and achievable by modern technology. With Google’s self-driving cars having travelled 700,000 miles as of August 2014 (Gomes, 2014) and fully driverless cars anticipated to be brought onto the UK market within the 2020s (Gov, 2015), driverless cars could be set to take over the motor industry. However, with any such advanced technology there are a significant number of ethical issues which will potentially hinder the progress of these vehicles making it to market and this paper aims to explore such issues.

Possibly one of the most publicly discussed issues with regards to these cars is when they have to be ‘programmed to kill’. By such a statement, what I mean is that in certain, serious situations, a driverless car may be required to make a decision which could consequently end in the death of at least one human being. How these cars come to such decisions is crucial in their feasibility as a marketable product.

A well-known thought experiment which covers these issues is the Trolley Problem (Thomson,1985). The problem is described as follows: Imagine you are driving a trolley and you come to a point where the track forks in two directions. In one direction you find 5 people, unable to move out of the way in time, and on the other side only 1 person in the same position. You find your brakes are broken so you are guaranteed to kill at least 1 person. Is it the correct decision to turn the trolley onto the side with 1 person? This problem is clearly applicable to the aforementioned programming of these cars. Most people would take the utilitarian approach in this scenario; opt to kill the lone person as it maximises the total lives saved. An argument can be made against such an approach though, as has been discussed by Ian Chapman (2015). He brings up the issue of a slightly more subtle application of the trolley problem. Suppose a driverless car has a decision between hitting (and potentially injuring) one of two motorcyclists on the road: the first motorcyclist is wearing a helmet whereas the second is not. If the car were to act according to utilitarianism then it would hit the cyclist wearing the helmet, which could seem questionable to some as, if taken as a general rule, would suggest that people attempting to put themselves out of harm’s way, are in fact put directly at harm in such conditions. [PLEASE WRITE SOMETHING HERE REGARDING HOW THIS COULD BE SENSIBLY DEALT WITH – MORE RESEARCH COULD BE REQUIRED]

If driverless cars were to become fully autonomous then one can view them as an independent decision maker. If we assume that the car doesn’t treat any *human* life more important than another (a different case may be applicable for animals etc.) then the car could potentially decide to kill the driver (or more accurately in this case, passenger) if it appeared to be the best option. At this stage, it is notable that this may not be the same decision as a human driver would make so it is worth questioning: if we are questioning the safety and decision making of driverless cars, shouldn’t we question human decision making before using it as a standard for autonomous vehicles?