

Ethics and Autonomous Cars

Patrick Lin, Ph.D.

Director, Ethics + Emerging Sciences Group
Associate Professor, Philosophy Department
California Polytechnic State University
San Luis Obispo, California 93407 USA
palin@calpoly.edu

ABSTRACT:

Contrary to popular opinion, but understood well by auto manufacturers, crashes with autonomous cars are inevitable. To operate safely and responsibly, especially in dynamic city environments, “robot cars” will need the capacity to make difficult decisions involving a tradeoff of harms. These are fundamentally decisions about ethics: any engineering design that involves a tradeoff—such as to strike object x instead of object y—requires a value-judgment about the wisdom of the tradeoff, that is, the relative weights of x and y.

This presentation explains why programming a machine to behave ethically is a difficult challenge, but perhaps not unsolvable as a matter of policy. To stress-test design principles related to ethics and identifying possible failure points, I will consider novel scenarios as well as ones inspired by classical dilemmas in ethics, such as the trolley problem.

I will also consider novel solutions, such as using a random-number generator to make morally fraught decisions. Finally, for the sake of completeness, I will look at ethical concerns beyond programming or algorithm ethics, such as issues related to use, abuse, and broader impacts of the technology on society.

This work is informed by my research at Stanford’s School of Engineering and Stanford Law School during the previous academic year, including meetings with industry leaders such as Google, Tesla, Daimler, Nissan, Bosch, State Farm, and many others. Please see my website for more discussion about this subject, especially in the Contributed Papers section: <http://ethics.calpoly.edu/robots.htm>.