8/14/2016
Editor:
science_letters@aaas.org

Dear Editor:

Re: "The social dilemma of autonomous vehicles", Bonnefon et al., 2016, *Science* [1]

The authors pose a "social dilemma" for autonomous vehicles (AVs) with scenarios they describe as "unlikely" while relying on surveys (static self-reports) to make predictions about human preferences to these AV decisions. But we have known for decades that preferences self-reported by humans often mis-align with human behavior ([4];[9];[6]). For example, reported in *Science News* [3], 90% of female partners self-reported compliance with a drug regimen to prevent transmission from their HIV mates, indicating drug failure. But before rejecting the drug, blood samples collected at the same time as the self-reports were compared to discover that compliance by the females was only 30%, giving new life to the drug. "There was a profound discordance between what they told us … and what we measured," infectious disease specialist Jeanne Marrazzo said.

As two other examples, Nate Silver, the renowned political forecaster [2], declared a crises with polling last year after failing to predict the outcome of five national and international contests. Tetlock and Gardner [8] claimed that "forecasting ... is a skill that can be cultivated." Their webpage titled "Good Judgment" displayed the first question for their hand-picked superforecasters: "Will a majority of voters in Britain's upcoming referendum elect to remain in the European Union?" Despite giving only a 23% chance that the British would leave the EU [5], these superforecasters failed to predict Brexit in 2016.

At our AAAI symposium at Stanford in March [7], we constrained self-reported surveys with dynamic interdependence to tackle these more likely ethical scenarios: When four AVs approach an intersection with one AV "aware" its human driver is impaired, should the AVs coordinate with each other to protect their human occupants? Should we as a society allow a robot pilot of a team to take control when the robot becomes "aware" of an impending suicide by the airliner's human copilot? Should a robot take command of a USS submarine prepared for rapid ascent to prevent the submarine from hitting a Japanese tour boat?

Respectfully,

W.F. Lawless, Augusta, GA 30901

References:

1. Bonnefon, J.F., Shariff, A. & Rahwan, I. (2016), The social dilemma of autonomous vehicles, *Science*, 352(6293): 1573-1576.
2. Byers, D. (2015, 5/8), "Nate Silver: Polls are failing us", *Politico*, from http://www.politico.com/blogs/media/2015/05/nate-silver-polls-are-failing-us-206799.html
3. Cohen, J. (2013), Human Nature Sinks HIV Prevention Trial, Science, 351: 1160, from http://www.sciencemag.org/news/2013/03/human-nature-sinks-hiv-prevention-trial
4. Kelley, H.H. (1991), Lewin, situations, and interdependence, Journal of Social Issues **47**: 211-233.
5. Kennedy, S. (2016, 6/25), "Superforecasters See 23% Brexit Chance as Economy Wins Out", Bloomberg, from http://www.bloomberg.com/news/articles/2016-05-18/superforecasters-see-24-chance-of-brexit-as-economy-wins-out
6. Lawless, W.F. (2016), "Preventing (another) Lubitz: The thermodynamics of teams and emotion", in Harald Atmanspacher, Thomas Filk and Emmanuel Pothos (Eds.), Quantum Interactions. LNCS 9535, Springer International Switzerland, pp. 207-215.

7. Mittu, R., Taylor, G., Sofge, D. & Lawless, W.F. (2016), Organizers: AI and the mitigation of human error: Anomalies, team metrics and thermodynamics. AAAI-2016 Symposium at Stanford; see https://www.aaai.org/Symposia/Spring/sss16symposia.php#ss01.\
8. Tetlock, P.E. & Gardner, D. (2015), Superforecasting: The Art and Science of Prediction, Crown.
9. Zell, E. & Krizan, Z. (2014), Do People Have Insight Into Their Abilities? A Metasynthesis? Perspectives on Psychological Science 9(2): 111-125.