

Rose Martin, Petko Kusev and Paul van Schaik





UNIVERSITY OF SURREY

HIGHLIGHTS

- Recent research reveals that people do not want to buy the pro-social utilitarian autonomous vehicle (AV) that they judge to be the most moral (Bonnefon et al., 2016).
- We argued that this behavioural phenomenon is fuelled by participants having access to only partial perspective-taking (PT) accessibility in AV crash scenarios.
- We have developed and empirically established a novel theoretical proposal
 PT accessibility, providing people with access to all situation perspectives during a crash (full PT accessibility).
- Accordingly, in three experimental studies, we found that full PT accessibility induced respondents' utilitarian prosocial judgments and purchasing behaviour, and consistent utilitarian preferences across judgment tasks.
- We found that consistent with (and informed by) their moral judgments, participants were more willing to buy, ride and spend more money on prosocial cars over passenger-protective vehicles.
- Importantly, we have argued that the full PT accessibility is a new type of 'veil of ignorance', which is not based on purposely induced self-interest and uneven risk options.

EXTENDED SUMMARY

This research article is published in a world leading and interdisciplinary peer-reviewed journal (Cognition), and offers a new theoretical proposal regarding human moral decisionmaking, as well as a new experimental method for testing the predictions of the proposal.

Some decisions made to improve the safety of one individual can consequently impede the safety of many others. For instance, purchasing an autonomous vehicle designed to prioritise the safety of its passengers (passenger-protective vehicles) could endanger the lives of other drivers and pedestrians. Accordingly, in preparation for unavoidable collisions, autonomous vehicle (AV) manufacturers could program their cars with pro-social utilitarian ethical algorithms that maximise the number of lives saved during a crash. However, recent research reveals that people do not want to buy the pro-social utilitarian AV that they judge to be the most moral (Bonnefon et al., 2016).

In order to address the psychological underpinnings of this phenomenon, we have developed and empirically established a novel theoretical proposal – perspective-taking (PT) accessibility. Specifically, in three experiments, we found that providing participants with access to both situational perspectives (AV buyers

can be passengers or pedestrians) in crash scenarios, eliminated the behavioural inconsistency between their utilitarian (prosocial) judgments of moral appropriateness and nonutilitarian (passenger-protective) purchasing behaviour. Accordingly, the results revealed that full PT accessibility induced respondents' utilitarian prosocial judgments and purchasing behaviour, and consistent utilitarian preferences across judgment tasks. For example, consistent with (and informed by) their moral judgments, participants were more willing to buy, ride and spend more money on prosocial cars over passenger-protective vehicles. Thus, contrary to claims that implementing a utilitarian AV policy might delay public adoption of AVs (Bonnefon et al., 2016; Greene, 2016; Shariff et al., 2017), we provide evidence that people perceive utilitarian AVs as more valuable than passenger-protective models.

Importantly, we have argued that the full PT accessibility is a new type of 'veil of ignorance', which is not based on purposely induced self-interest and uneven risk options (as in Huang, Greene, & Bazerman, 2019), but rather is based on even odds of being a passenger or pedestrian, and therefore with an even 50/50 chance to die/live as passenger or pedestrian. Accordingly, we have concluded that under these circumstances one can measure utilitarian preferences.