

# Does Automation Alter Moral Appraisals of (In)Action: The Case of Autonomous Vehicles

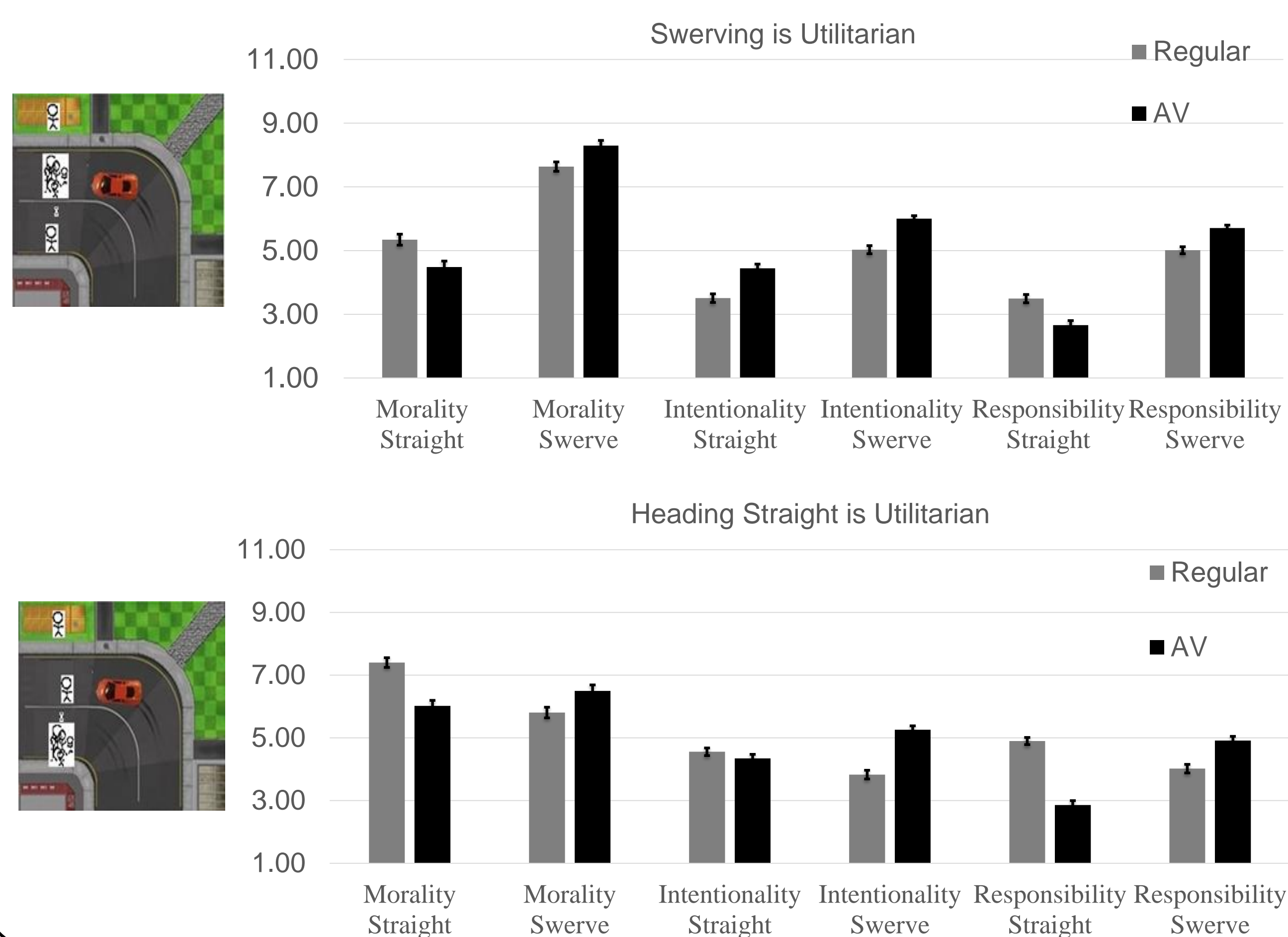
Amos Schurr & Simone Moran

Guilford Glazer Faculty of Business and Management Ben-Gurion University of the Negev

**Introduction:** Automation is taking an increasing role in our lives as is apparent in developments of automated vehicles (AVs), medical decision support systems, aviation, etc. Notwithstanding, to date, it is rarely the case that machines autonomously decide without some degree of human supervision. We thus focus on human-machine interactions, and ask: do people evaluate individuals' decisions and (in)actions differently, depending on whether they are performed in the presence (versus absence) of automated defaults? We suggest that due to the existence of externally pre-determined defaults in Automated systems (e.g., AVs), but not in non-automated ones (e.g., regular vehicles), taking action is more salient in the former and will thus play a greater role in determining judgements of decision makers. Namely, we predict and indeed find a greater impact of deontological considerations such as whether outcomes were reached by means of taking or not taking action, and the inferred intentions behind the behaviors, when judging decision makers who perform in the presence of automated defaults (e.g., AV drivers) than in their absence (e.g., regular car drivers).

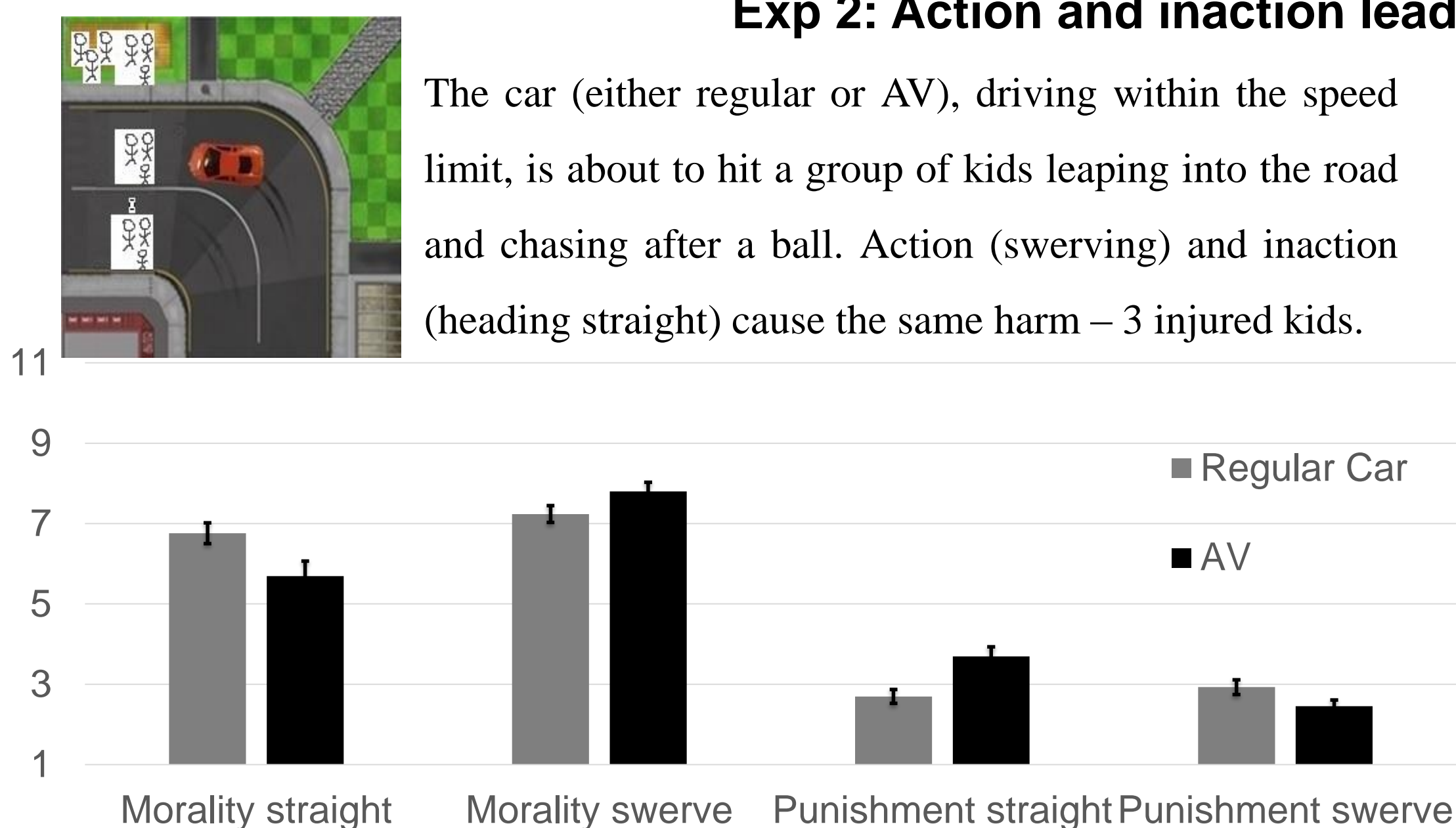
**Methods:** In two experiments, participants were presented with trolley-like scenarios illustrating an accident about to occur. We vary car type to be either regular or autonomous. We compare peoples' judgments of the morality, intentionality and attributed responsibility of a driver who swerves versus continues straight (Exp. 1), and test the consequences of these judgmental patterns in terms of deserved punishment (Exp 2).

## Exp 1: Judging morality, intentionality and responsibility (N=722)



Across scenarios when estimating the morality of the regular car driver (gray bars), participants follow a utilitarian rule of minimizing harm – Swerving is judged more moral when 5 are ahead, continuing straight is more moral when only 1 is ahead. A similar pattern is apparent for judgments of intentionality and assuming responsibility – hitting a single kid is judged as taking greater responsibility over the situation and as more intentional. When estimating the AV driver (black bars), independent of what is utilitarian (across scenarios), swerving is always deemed more moral than heading straight. Also, when comparing appraisals of AV drivers to regular car drivers, AV drivers are credited more for taking action. This pattern is also apparent in ratings of intentionality and taking responsibility over the situation. Acting is judged more intentional and as taking greater responsibility over the situation when driving AVs.

## Exp 2: Action and inaction lead to same outcome (N=353)



The car (either regular or AV), driving within the speed limit, is about to hit a group of kids leaping into the road and chasing after a ball. Action (swerving) and inaction (heading straight) cause the same harm – 3 injured kids.

When estimating the morality and deserved punishment of the regular car driver (gray bars), participants did not differentiate between swerving (action) and heading straight (in action). Lack of a difference implies that participants were driven by consequentialism. In contrast, when estimating the morality and deserved punishment of the AV driver (black bars) participants differentiated between action and inaction crediting action over inaction. This finding implies that judging AV drivers was significantly driven by deontological considerations. Also when comparing appraisals of AV to regular car drivers, AV drivers were credited more than regular car drivers for taking action.

## Conclusions:

We use different criteria to evaluate drivers in regular cars versus drivers in autonomous vehicles.

- Regular car - Consequential criterion
- Avs - Action/inaction criterion matters as well
  - AV (vs. regular car) driver who acts is deemed more moral
  - AV driver who does not act is deemed less moral
- Different criteria may lead to different punishments for identical outcomes and actions.