

## **Lesionar a otros o lesionarme a mí, un dilema para nuestros vehículos autónomos.**

**Acerca del autor:** Dr. Miguel Angel Luzuriaga Moreno. Egresado de la licenciatura en Administración de Empresas, generación 2002. Investigador de Post Doctorado en la Universidad de Ciencias Aplicadas Neu-Ulm en Alemania.

Los vehículos autónomos (mejor conocidos como AVs por sus siglas en inglés: *Autonomous Vehicles*) pronto serán una realidad en nuestros caminos. El monstruo de la tecnología Google, junto con las automotrices más importantes a nivel mundial están invirtiendo una gran cantidad de talento y capital para diseñar vehículos fiables capaces de transportar mercancías y personas sin necesidad de conductores. Sin duda, los nuevos AVs serán mucho más seguros que los clásicos dirigidos por humanos. Los conductores automáticos nunca serán víctimas de la somnolencia o distracción, no consumirán alcohol ni drogas, y controlarán aspectos importantes de su entorno, como las condiciones meteorológicas y la presencia o ausencia de peatones y otros vehículos. Dado que los AVs estarán siempre alerta, y siempre que los posibles errores y fallas del software se reduzcan al mínimo, la sustitución de conductores humanos por estos nuevos dispositivos permitirá una reducción significativa en accidentes de tráfico y pérdidas humanas. Aparentemente los AVs nos ayudarán a ahorrar dinero, tiempo y dolor. Sin embargo, los científicos e ingenieros que diseñan y proyectan estos sistemas todavía tienen que superar importantes desafíos. Los obstáculos a superar no sólo son de índole técnico, sino también legales y filosóficos. Estos últimos se relacionan con el comportamiento de los AVs ante accidentes inevitables. No olvidemos que los AVs son “criaturas” humanas y por tanto está dentro de las posibilidades el involucramiento en accidentes. El problema es que el comportamiento del AV esencialmente está predeterminado, ya que está sujeto a reglas de programación. No es el conductor humano (que de hecho no existe) pero si los ingenieros quienes tienen que decidir cómo se comportan los AVs en las situaciones de tráfico. Los ingenieros deben decidir de antemano si el objetivo más importante es proteger a los pasajeros por todos los medios o bien minimizar el número total de afectados, o cualquier objetivo alternativo. Esta no es una cuestión técnica o de ingeniería, sino un problema ético el cual plantea importantes dilemas morales relacionados con la distribución de daños en la sociedad. Incluso los gobiernos están ya preocupados por este tema. En Alemania, un comité de ética nacional para la conducción automatizada designado por el gobierno ha establecido que:

*“La decisión que debe tomarse es si la concesión de licencias de sistemas de conducción es éticamente justificable, e incluso, si es imperativo hacerlo. Si estos sistemas tienen licencia, y es evidente que esto está sucediendo a nivel internacional - todo recae en las condiciones en las que se utilizan y la forma en que estos sistemas están diseñados” (Luetge, 2017).*

Imaginemos por un momento un accidente inevitable: ¿cómo debe actuar el AV? Quizás deba proteger la vida de los pasajeros hasta cierto punto. Muchas personas estarían de acuerdo con esto. Pero ¿qué pasa con los demás involucrados? ¿Debería el AV sacrificar la vida de (digamos) un peatón para salvar la vida de varios pasajeros? En otras palabras: ¿deberíamos tener un AV programado para minimizar el número total de víctimas? Para algunos esto es lo ideal, pero no para todos, esta parece ser una

decisión muy compleja. De hecho, es difícil llegar a un consenso sobre esta cuestión. Hace muchos años, cuando nadie podía ni siquiera soñar con la existencia de los AVs, los filósofos morales diseñaron un experimento mental denominado "*El Problema del Carrito*", el cual sintetiza las principales características de este dilema moral. En su versión más simple, un tranvía descontrolado se dirige hacia una serie de personas que están trabajando sobre las vías y las matarán a todas, a menos que alguien tire de una palanca y desvíe el carro hacia una vía lateral. En este caso, sólo una persona sería la sacrificada, el conductor. ¿Qué haríamos ante esta situación? Este dilema y sus variaciones fueron objeto de animadas discusiones entre los filósofos quienes no pudieron llegar a una solución aceptable para todos. Inicialmente, dicho dilema fue estudiado de forma experimental por Philippa Foot en 1967 y desde entonces ha sido objeto de controversia. Este dilema clásico de la literatura lo hemos revivido experimentalmente bajo el nuevo disfraz de los vehículos autónomos. Nuestra intención fundamental es la de investigar, usando métodos científicos, cómo la gente actuaría ante este dilema si manejara un automóvil tradicional *versus* cómo programaría *-ex ante-* un AV para enfrentar el mismo problema. En nuestro experimento, introducimos además una serie de variables que no habían sido investigadas en la literatura relacionada al tema. Por ejemplo, las consecuencias de nuestras decisiones en el mundo real están rodeadas de incertidumbre, y por lo general es difícil predecir con precisión lo que sucederá después de que se ha tomado una decisión. Así que introducimos la incertidumbre, así como la posibilidad de tomar algunas decisiones adicionales que pueden influir en las probabilidades ligadas a un accidente, como la velocidad del vehículo antes de un impacto. La gravedad de los daños y lesiones después de un choque están estrechamente relacionados con la velocidad de conducción. Los peatones, por ejemplo, tienen un 90% de posibilidades de sobrevivencia cuando son golpeados por un automóvil que viaja a 30 km/h, pero las posibilidades bajan más del 50% cuando la velocidad es de 45 km/h.

Economistas del comportamiento han concluido que la toma de decisiones puede ser irracional, y esa irracionalidad puede llegar a ser predecible. ¿Cuál es entonces la racionalidad o la irracionalidad predecible detrás de un problema ético, en particular, cuando hay componentes económicos importantes en juego? Debido a la importancia de los costos económicos durante un incidente, incorporamos también este factor a nuestro experimento. Esto es, asignamos incentivos económicos reales a las consecuencias de las decisiones tomadas por cada uno de los participantes de nuestro estudio. En total, recolectamos las decisiones de 150 participantes y pagamos a cada uno de ellos un promedio de 14 EUR. Cerca de la mitad resolvieron este dilema conduciendo por sí mismos un simulador de manejo programado específicamente para simular el "*Problema del Carrito*" (ver Fig. 1). En tanto el grupo restante programó un (hipotético) AV para enfrentar el mismo dilema (ver Fig. 2). Utilizando modelos probabilísticos encontramos que el programar un AV hace a los individuos más propensos a sacrificarse a sí mismos para así evitar impactar a un peatón. En contraste, aquellos que conducen un auto tradicional son más propensos a sacrificar al peatón para no salir afectados. Paradójicamente, encontramos que esta diferencia en el comportamiento humano se debe más a la aversión de conductores y pasajeros a tomar riesgos en donde se vean afectados económicamente, que a las creencias morales de lo que debe o no debe hacerse ante un dilema de tal magnitud.

El marco teórico, diseño experimental, resultados detallados, así como una amplia discusión sobre las implicaciones de nuestra investigación se encuentran publicados en la reconocida revista científica *Review of Behavioral Economics* la cual es líder internacional en la divulgación científica en temas de economía conductual. Dicho artículo forma parte de un proyecto de investigación conjunto entre el Centro de Investigación Allgäu (Dr. Miguel Luzuriaga), la Universidad Complutense de Madrid (Prof. Antonio Heras) y la Universidad de Ciencias Aplicadas Neu-Ulm (Prof. Oliver Kunze):

*Miguel Luzuriaga, Antonio Heras and Oliver Kunze (2020), "Hurting Others versus Hurting Myself, a Dilemma for Our Autonomous Vehicle", Review of Behavioral Economics: Vol. 7: No. 1, pp 1-30.*



Fig.1: Experimento con simulador de manejo (Simu Tech GmbH, Bremen, 2017) de un vehículo tradicional: transmisión automática, triple pantalla, velocidad máxima 220 km/h, tres espejos retrovisores, indicadores de laterales, pedales y volante incluidos.



Fig. 2: Experimento de programación de un AV a través del software z-Tree (Fischbacher, 2007).

## Referencias:

- [1] Foot, P. 1967. "The problem of abortion and the doctrine of double effect". Oxford Review. 5: 5–15.
- [2] Luetge, C. 2017. "The german ethics code for automated and connected driving". Philosophy and Technology. 30: 547–558. doi: 10.1007/s13347-017-0284-0.
- [3] Fischbacher, U. 2007. "z-Tree: Zurich toolbox for ready-made economic experiments". Experimental Economics. 10: 171–178. doi: 10.1007/s10683-006-9159-4.

**Tags:** Miguel Luzuriaga, vehículos autónomos, AVs, simuladores, futuro, economía, seguridad, peatón, Egresados UDLAP