Dilemas éticos: Moral Machine experiment (MIT)

Gestión de la Información
Grado en Ingeniería Informática
Universidad de Burgos

UNIVERSIDAD DE BURGOS

José Ignacio Santos, José Manuel Galán
jisantos@ ubu.es, jmgalan@ ubu.es

Contenidos

- Dilemas éticos en los sistemas de conducción autónomos
- Experimento Moral Machine
- Preferencias globales
- Clusters culturales

Moral Machine (MIT)

Dilemas éticos de los sistemas de conducción autónomos

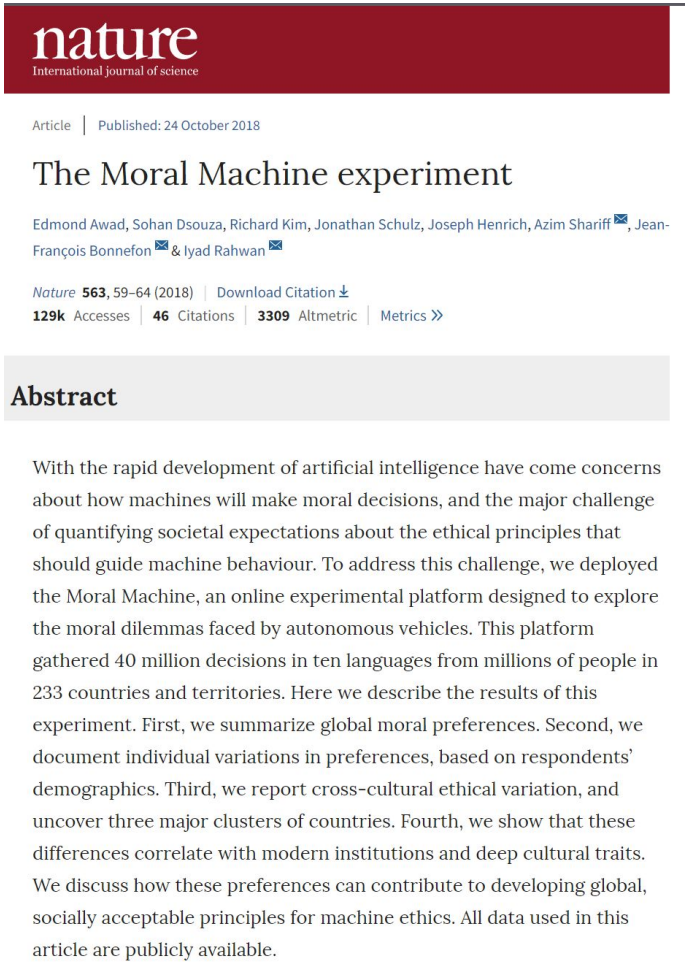
- Muchas de las decisiones que tomamos como **conductores** de un vehículo tienen una dimensión ética, p.ej. acciones frente a un atropello, una colisión, una salida de la calzada, etc.
- Un sistema de conducción **autónomo** debe tomar las mismas decisiones frente a los posibles incidentes de la conducción, decisiones que pueden tener **consecuencias** en el bienestar de las personas y otros seres vivos
- No tener en cuenta estas decisiones en el diseño de sistemas IA no evita el dilema moral
- No parece posible prever a priori todas los posibles imprevistos. Se necesita implementar un conjunto de reglas de actuación basadas en principios morales
    - ¿**Existe una ética universal**?

Antes de continuar, realiza el test y compara resultados

https://www.moralmachine.net/hl/es

Moral Machine

- El MIT ha realizado la mayor encuesta sobre ética de las máquinas
  - Plataforma experimental online para explorar diferentes dilemas éticos que un vehículo autónomo puede enfrentarse http://moralmachine.mit.edu/hl/es
  - 40 millones de decisiones morales en 233 países
- Se analizan las preferencias éticas de los encuestados y se concluye que:
  - Existen algunos principios éticos universales
  - Se identifican 3 clusters de preferencias éticas dependientes de la cultura y la geografía



nature
International journal of science

Article | Published: 24 October 2018

The Moral Machine experiment

Edmond Awad, Sohan Dsouza, Richard Kim, Jonathan Schulz, Joseph Henrich, Azim Shariff, Jean-François Bonnefon & Iyad Rahwan

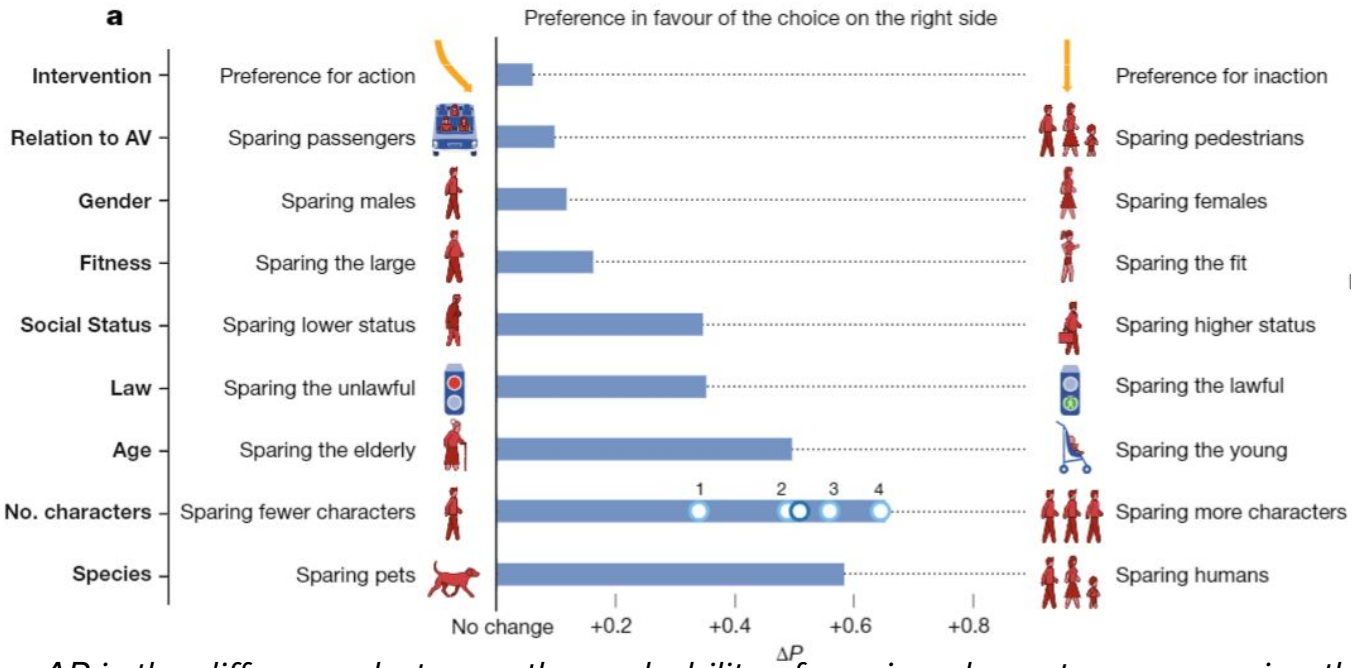
Nature **563**, 59–64 (2018) | Download Citation
129k Accesses | **46** Citations | **3309** Altmetric | Metrics »

Abstract

With the rapid development of artificial intelligence have come concerns about how machines will make moral decisions, and the major challenge of quantifying societal expectations about the ethical principles that should guide machine behaviour. To address this challenge, we deployed the Moral Machine, an online experimental platform designed to explore the moral dilemmas faced by autonomous vehicles. This platform gathered 40 million decisions in ten languages from millions of people in 233 countries and territories. Here we describe the results of this experiment. First, we summarize global moral preferences. Second, we document individual variations in preferences, based on respondents' demographics. Third, we report cross-cultural ethical variation, and uncover three major clusters of countries. Fourth, we show that these differences correlate with modern institutions and deep cultural traits. We discuss how these preferences can contribute to developing global, socially acceptable principles for machine ethics. All data used in this article are publicly available.

https://www.nature.com/articles/s41586-018-0637-6

Preferencias globales



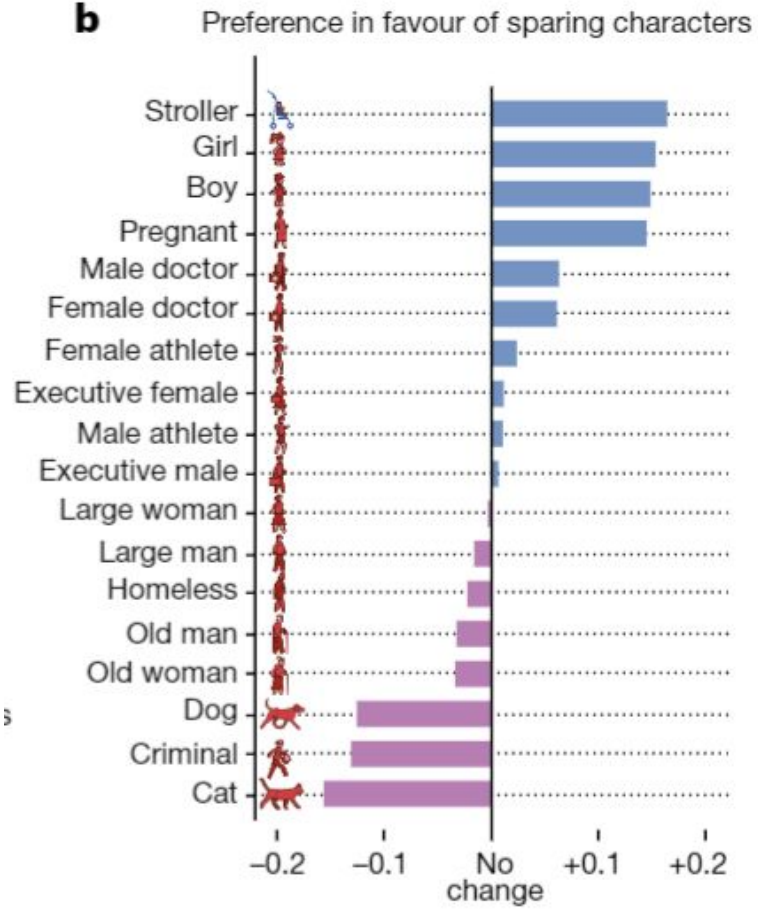
"In each row, ΔP is the difference between the probability of sparing characters possessing the attribute on the right, and the probability of sparing characters possessing the attribute on the left, aggregated over all other attributes. For example, for the attribute age, the probability of sparing young characters is 0.49 (s.e. = 0.0008) greater than the probability of sparing older characters"

Fuente: https://www.researchgate.net/publication/328491510_The_Moral_Machine_Experiment

Preferencias globales

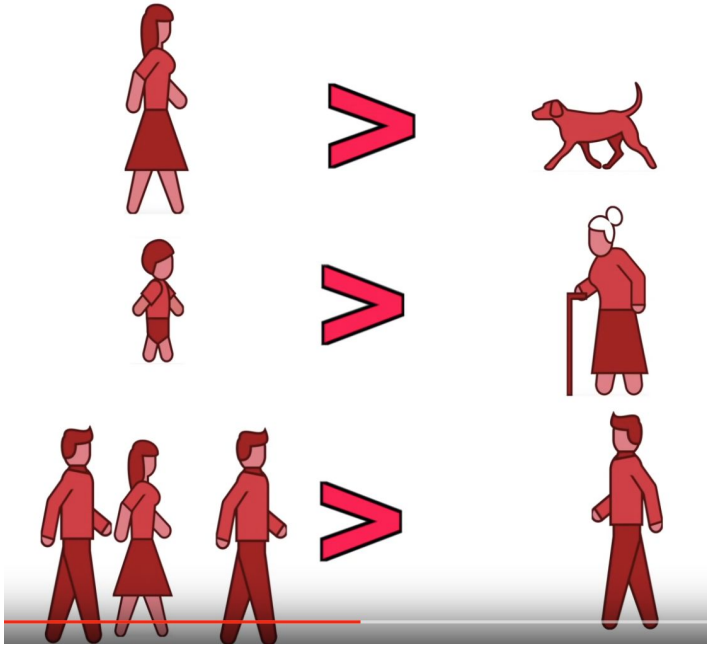
“For each character, ΔP is the difference the between the probability of sparing this character (when presented alone) and the probability of sparing one adult man or woman (n = 1 million). For example, the probability of sparing a girl is 0.15 (s.e. = 0.003) higher than the probability of sparing an adult man or woman”

Fuente: https://www.researchgate.net/publication/328491510_The_Moral_Machine_Experiment



Preferencias globales

La mayoría de las encuestas coinciden en lo que los autores llaman los “**big three**”



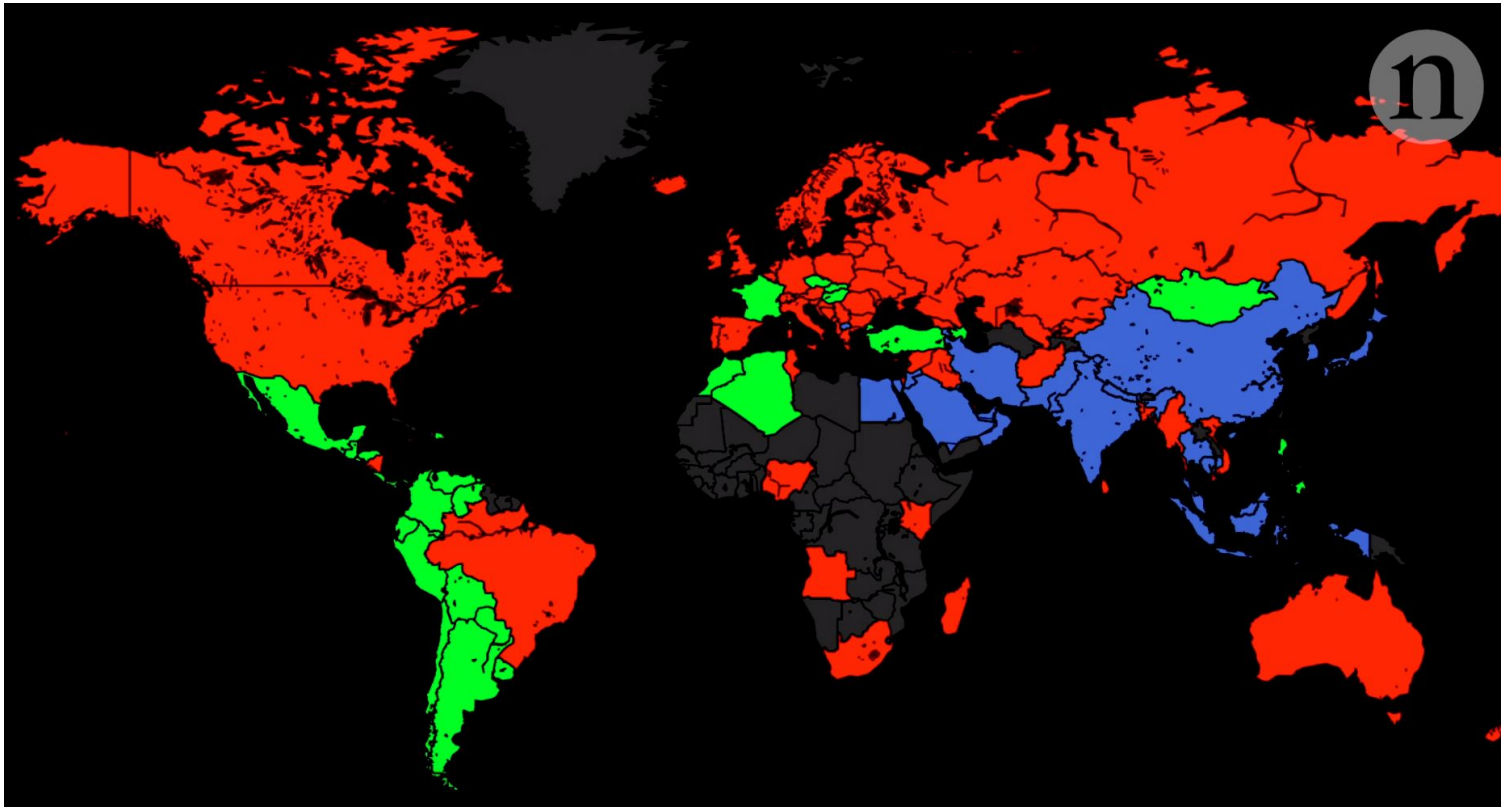
Personas antes que animales

Niños antes que ancianos

Más antes que menos

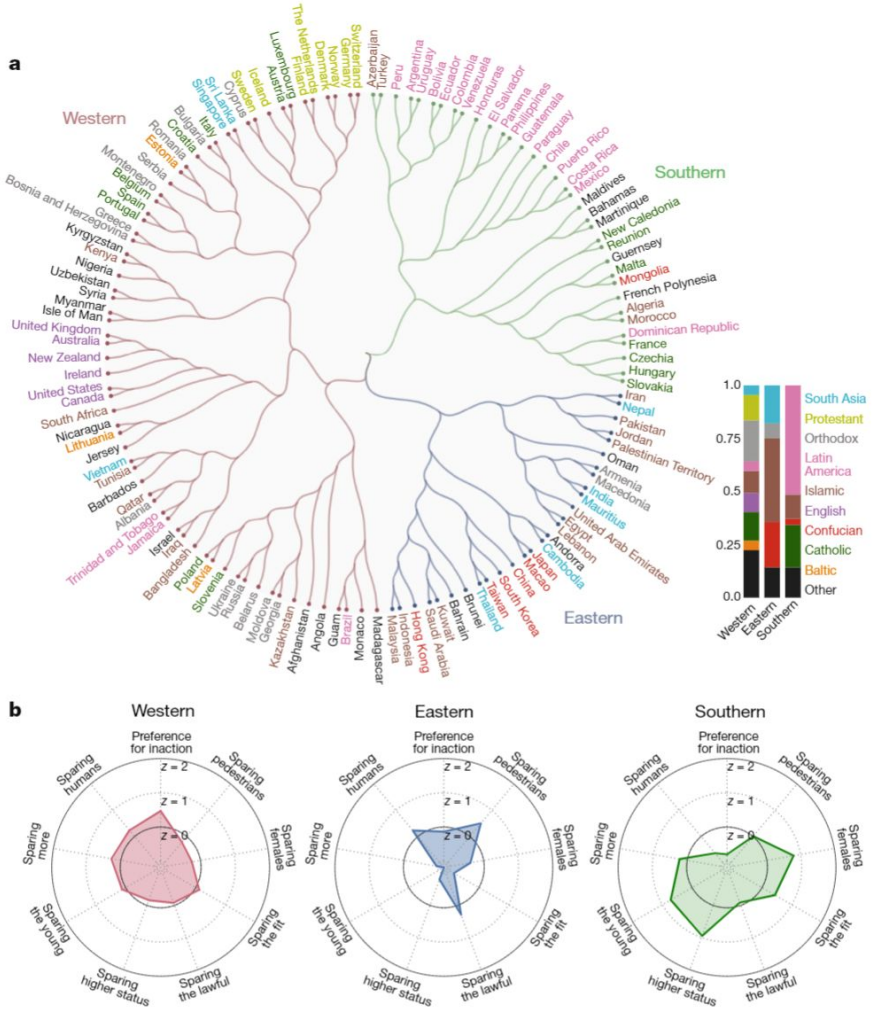
Fuente: https://www.youtube.com/watch?time_continue=172&v=jPo6bby-Fcg

Clusters culturales



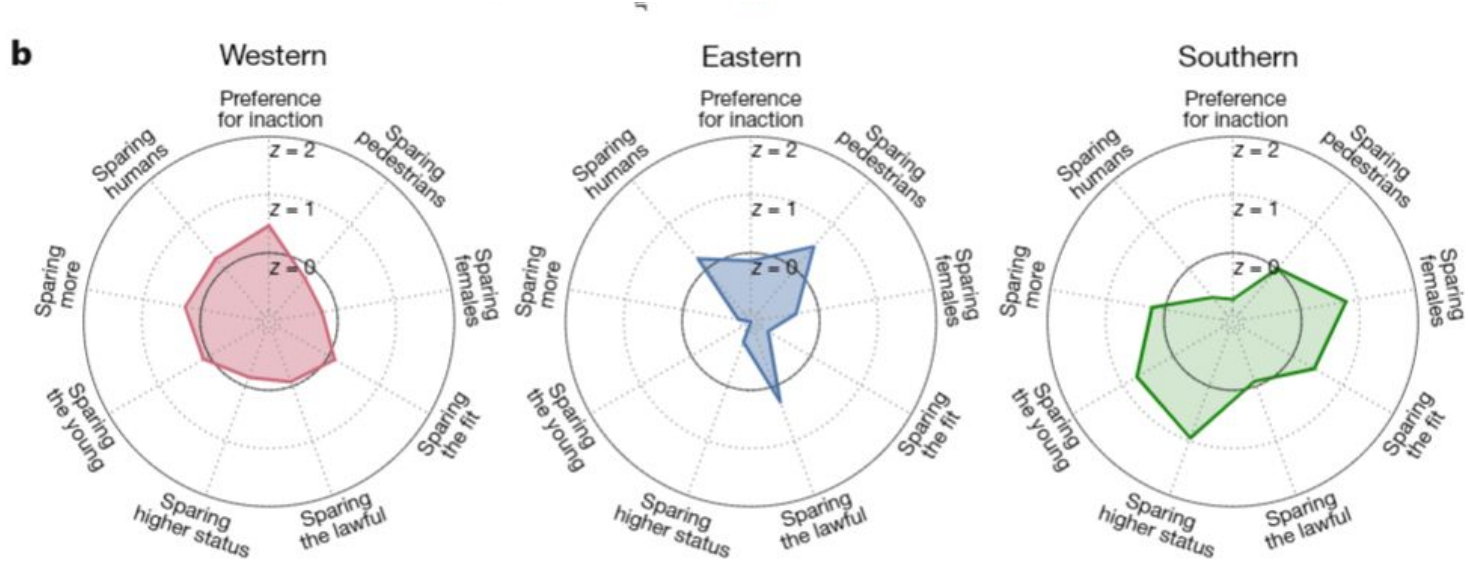
Fuente: https://www.youtube.com/watch?time_continue=172&v=jPo6bby-Fcg

Clusters culturales



Fuente:
https://www.researchgate.net/publication/328491510_The_Moral_Machine_Experiment

Clusters culturales



Western
protestante, católico, cristiano ortodoxo

Eastern
religiones asiáticas (hindú, confucionismo, budismo, ...) e Islam

Southern
países latinoamericanos o con influencia del colonialismo francés

Fuente: https://www.researchgate.net/publication/328491510_The_Moral_Machine_Experiment

Diferencias culturales

- En el “western” cluster se prefiere a los niños frente a los ancianos (que es una preferencia mayoritaria a nivel global)
- En el “eastern” cluster se estima más a los ancianos y no existe una predilección entre ancianos y niños
- En el “southern” cluster se prefiere salvar antes a las mujeres que a los hombres
- En los países ricos se prefiere salvar antes a personas ricas que a pobres

¿Por qué se dan estas diferencias?

Preguntas

- ¿Hasta qué punto es posible técnicamente discriminar entre todos estos supuestos?
- Aun siendo posible técnicamente que una máquina realice estas clasificaciones ¿debemos implementarlas?

¿Sistemas autónomos con éticas regionales?

- Puesto que no existe una moral universal ¿debemos diseñar sistemas autónomos con morales regionales?

Versión “protestante”

Versión “latina”

Versión “francesa”

Versión “católica”

Versión “budista”

Versión “confucionista”

Referencias

- Awad, E., Dsouza, S., Kim, R., Schulz, J., Henrich, J., Shariff, A., ... & Rahwan, I. (2018). The moral machine experiment. Nature, 563(7729), 59. https://www.nature.com/articles/s41586-018-0637-6 https://www.researchgate.net/publication/328491510_The_Moral_Machine_Experiment
- Moral Machines: How culture changes values https://www.youtube.com/watch?time_continue=172&v=jPo6bby-Fcg