

TRIBUNE

Voiture autonome : par quoi serez-vous choqués ?

Par [Jean-François Bonnefon, docteur en psychologie cognitive et directeur de recherches CNRS à la Toulouse School of Economics](#) — 19 octobre 2017 à 18:36 (mis à jour à 19:08)



Au Congrès international des transports intelligents, en octobre 2015, à Bordeaux. Photo Rodolphe Escher

En cas d'accident inévitable au volant, vaut-il mieux sacrifier la vie des passagers ou celle des piétons. C'est sur ce type de dilemme que le simulateur Moral Machine enquête.

Voiture autonome : par quoi serez-vous choqués ?

Quels risques êtes-vous prêt à accepter à bord d'un véhicule **autonome**?

Pour vous-même, les autres passagers et les passants ? Préférez-vous qu'une voiture renverse cinq piétons d'âge mûr plutôt qu'un bébé ? Un homme qu'une femme ? Votre voiture devrait-elle sacrifier votre vie pour sauver une femme enceinte, même si elle traverse au rouge ? Nous travaillons avec des chercheurs du MIT, de Harvard, et de l'université de Californie (1) sur le simulateur Moral Machine (moralmachine.mit.edu) qui

permet de collecter les préférences morales des gens dans ces situations où une voiture autonome doit choisir entre plusieurs accidents inévitables. Disponible en dix langues, il rassemble près de 4 millions d'utilisateurs et a déjà permis de récolter 37 millions de réponses dans une centaine de pays (2).

Les dilemmes éthiques de nos simulations ne sont pas entièrement nouveaux : ils se rapportent au fameux dilemme du tramway, expérience de pensée décrite dès 1967. Elle consiste à imaginer ce que devrait faire une personne qui a la possibilité de détourner de sa route un tramway, pour l'empêcher de tuer un groupe de cinq individus, mais ne peut que le rediriger vers quelqu'un d'autre qui mourra alors à leur place. Beaucoup de philosophes considèrent qu'il est impossible de faire un choix moralement irréprochable dans cette situation. Cette volonté de ne pas choisir est légitime tant que l'on reste dans une expérience de pensée. Mais le dilemme du tramway va se poser quand les voitures autonomes seront déployées sur les routes, et nous ne pourrons plus alors nous permettre de ne pas choisir.

Le cas du véhicule autonome est unique en intelligence artificielle en ce qui concerne les décisions à caractère moral. D'abord, parce qu'avec les voitures autonomes nous confions à une machine une décision que nous n'avons nous-mêmes jamais prise de façon réfléchie : en cas d'accident, nos réactions relevaient jusqu'ici de réflexes et d'impulsions imprévisibles. Ensuite, parce que les autres domaines (comme le médical ou le judiciaire) permettent un temps de réflexion avant d'accepter ou non le verdict de la machine, ce qu'une situation d'accident ne permet pas.

Le projet Moral Machine ne prétend pas déterminer ce qui est éthique ou moral. Mais il nous semble qu'avant de légiférer et mettre ces voitures sur les routes, il convient pour les pouvoirs publics et les constructeurs de savoir quelles sont les solutions les plus acceptables socialement aux yeux de la population. Ainsi, nous savons déjà grâce à nos premiers résultats qu'en cas d'accident inévitable, un consensus moral se forme pour minimiser le nombre de morts ou de blessés, qu'ils soient piétons ou passagers. Mais lorsque l'on demande aux utilisateurs quelle voiture ils achèteraient, ils préfèrent une voiture qui privilégie leur propre sécurité en tant que passager, aux dépens de celle des piétons.

Cette préférence peut amener des comportements peu rationnels. Ainsi, en tant qu'acheteur potentiel d'une voiture autonome, nous pouvons nous

montrer trop sensibles aux risques relatifs (c'est-à-dire la façon dont les risques sont distribués entre nous et les autres en cas d'accident). Et, à l'inverse, nous oublions la notion de risque absolu (c'est-à-dire la probabilité qu'un accident survienne en premier lieu). Or, ce risque absolu pourrait être jusqu'à dix fois inférieur pour une voiture autonome que pour un conducteur humain (3). Cette difficulté à intégrer les deux types de risque, et surtout leur importance respective, est problématique. Des consommateurs pourraient être rebutés par un risque relatif plus élevé (c'est-à-dire le fait que la voiture ne les protège pas toujours en cas de dilemme) sans réaliser que la voiture autonome diminue énormément la probabilité qu'une telle situation se produise !

A cette difficulté, s'ajoute celle de maintenir la confiance des utilisateurs dans un système autonome alors que ce système commettra inmanquablement des erreurs. En effet, les gens perdent plus vite confiance si une machine dotée d'intelligence artificielle fait une erreur, même petite, alors qu'ils pardonnent une erreur commise par un autre être humain. Nous pardonnons leurs erreurs aux humains car nous supposons qu'ils s'efforcent de ne pas les reproduire. Mais si un algorithme fait une erreur, alors la méfiance s'installe sur la façon dont il est programmé. Cette sévérité n'est pas toujours fondée (car les algorithmes peuvent apprendre de leurs erreurs et s'améliorer au fil de leur utilisation), mais elle sera un paramètre important à considérer avant la mise en circulation des voitures autonomes.

Même si nous parvenons à clarifier dans l'esprit du public les notions de risque relatif et absolu, et à dissiper certaines idées fausses sur les capacités des algorithmes, le débat sur l'éthique des voitures autonomes sera loin d'être tranché. Ainsi, le ministère des Transports allemand a présenté en septembre un rapport réalisé par des experts en intelligence artificielle, en droit et en philosophie, concluant qu'il serait moralement inacceptable pour une voiture de choisir qui elle doit sauver sur la base de caractéristiques telles que l'âge ou le sexe. Interdiction, donc, de sauver un enfant au prix de la vie d'un adulte. Fort bien. Notre rôle n'est pas de débattre des mérites éthiques de cette recommandation. Mais notre projet Moral Machine permet de comprendre ce à quoi les gens s'attendent et ce qui risque de les scandaliser. Et au vu de nos résultats, il est probable que l'opinion soit scandalisée le jour où un enfant sera sacrifié pour ne pas risquer la vie d'un adulte. Il faut comprendre que même si des décisions ont été prises sur des

critères éthiques respectables, elles conduiront à certains accidents qui généreront un outrage public et qui compromettront l'acceptabilité et l'adoption des voitures autonomes. Il s'agit donc pour nous d'anticiper ces mouvements d'opinions et d'en mesurer les risques, et non de peser sur les législateurs afin qu'ils collent à la volonté des gens telle qu'elle est mesurée avec notre simulateur.

(1) Edmond Awad, Jean-François Bonnefon, Sohan Dsouza, Joe Henrich, Richard Kim, Iyad Rahwan, Azim Shariff, Jonathan Schulz. Le projet Moral Machine est né de notre première publication sur le sujet : «The Social Dilemma of Autonomous Vehicles», *Science*, de Bonnefon, Shariff et Rahwan (2016).

(2) Selon les projections les plus optimistes, la voiture autonome déployée à large échelle permettrait d'éviter 90 % des accidents car 90 % des accidents actuels sont liés à une erreur humaine (conducteur fatigué, en état d'ébriété, qui ne respecte pas le code de la route, etc.).

«Algorithm Aversion : People Erroneously Avoid Algorithms After Seeing Them Err.» *Journal of Experimental Psychology : General*, de Berkeley J. Dietvorst, Joseph P. Simmons, et Cade Massey (2014).

(3) «Psychological Roadblocks to the Adoption of Self-Driving Vehicles», *Nature Human Behaviour*, de Shariff, Bonnefon et Rahwan (2017).

[Jean-François Bonnefon docteur en psychologie cognitive et directeur de recherches CNRS à la Toulouse School of Economics](#)