La morale d'hier et l'automobile autonome de demain

La voiture autonome pose un problème éthique. Le problème n'est pas tout à fait nouveau, mais il a bien quelque chose d'inédit. Jusque-là on ne connaissait que le choix humain, le dilemme moral et, depuis assez peu, l'aide à la décision.
Ici c'est autre chose. Il s'agit de programmer une machine à choisir en situation critique. Ce n'est plus l'homme (l'individu) qui choisit, mais l'homme (plus l'individu) qui doit choisir comment la machine doit choisir à sa place et pour lui. Il appartient à l'homme de choisir comment programmer le choix qui le dispensera à l'avenir de choisir. C'est le dilemme moral de la programmation de l'automobile autonome.
Ou si la situation n'est pas complètement inconnue, c'est paradoxalement à la structuration primitivement religieuse de « l'établissement humain » (Marcel Gauchet[1]) qu'il faut la rapprocher. Selon lui, la société humaine est entrée dans l'histoire en lui tournant le dos, par la soumission à l'hétéronomie (religion) : elle aurait choisi pour se posséder de se déposséder collectivement à l'égard de la loi imprescriptible du passé et de la tradition. Mais la société humaine, à l'autre bout, pourrait sortir de l'histoire en déléguant le principe d'autonomie qui a porté l'avènement de la démocratie des Modernes à travers les facteurs politique, historique, juridique[2], à la machine "intelligente" (IA). Fin donc d'un cycle historique : l'automobile, symbole de l'autonomie du sujet moderne, devient objet autonome à la place du sujet.

Du choix au dilemme

La question du dilemme est une question classique de la philosophie morale. Le dilemme, c'est le problème du choix : que dois-je choisir ? Mais avant même la valeur morale d'un choix (bien ou mal), il y a le choix même comme condition de la morale. On passe de l'esthétique à l'éthique (ou morale) en passant de "et" à "ou". Une vie esthétique, explique le philosophe danois Kierkegaard[3], se déroule sur le plan du "et" : la beauté n'est jamais exclusive. Je peux aimer sans contradiction une fugue de Bach et la cathédrale de Reims, et tout ce qu'on voudra d'autre : l'esthétique c'est le règne du "et" (A

[1] *Le désenchantement du monde*, 1985.
[2] M. Gauchet, *L'avènement de la démocratie*, t.1-4.
[3] *Ou bien … ou bien* (ou *L'Alternative*, 1843).

et B et C et …) Passer à la morale c'est, au contraire, dépasser la simple conjonction des préférences en se mesurant à une dimension nouvelle, l'alternative : ou bien A… ou bien B.
Si je choisis A alors :
(1) je fais passer A du possible au réel ;
(2) je ne choisis pas B.
(1) est la condition de la responsabilité — les conséquences de mon choix A me sont imputables, au moins dans les limites de ce que je savais et pouvais anticiper comme conséquences de A ;
(2) est la condition de l'irréversibilité — à la limite, sans doute, je peux encore choisir B, mais il n'annulera pas le choix A et viendra après lui (B après et d'après A n'est pas le même que B quand il était opposé à A).
Autrement dit, existentiellement le choix approfondit tragiquement la vie de l'individu : si je choisis A, alors il est impossible que non A, maintenant et pour toujours. Ou il sera toujours vrai que A, si j'ai choisi A. Ethiquement, l'existence se laisse décrire comme une suite ouverte de choix irréversibles (A puis B, puis C). L'existence de l'individu est ainsi l'histoire (irréversible) de ses choix (irréversibles) en situation. On peut donc dire, à ce premier niveau, que l'alternative (ou bien… ou bien) est la condition formelle ou générale de la morale et que l'alternative révèle le tragique de l'existence (responsabilité et irréversibilité).
Mais le dilemme moral est quelque chose de plus déterminé que l'alternative (comme condition formelle de la morale). Il soumet nos intuitions et nos théories éthiques à une sorte de test. On peut même dire que le dilemme est crucial en morale. Pour être consistante, une théorie morale doit-elle inclure ou exclure la possibilité du dilemme moral ? Si elle exclut le dilemme, elle paraît ne pas prendre en compte la réalité de la vie morale, de fait exposée à des conflits de devoirs[4] ; si elle l'inclut, sans le résoudre, elle ne répond pas à sa fonction de fonder le choix moral en certitude. Donc si la théorie morale exclut le dilemme elle perd de vue la réalité de l'action : si elle l'inclut, elle risque de se nier comme théorie.
Mais on peut aussi, tout à l'inverse, relativiser l'importance du dilemme en morale, au point d'en nier la réalité. C'est d'ailleurs le lieu commun des théories morales, chez Platon pour l'Antiquité, Leibniz, Kant (éthique kantienne de type « déontologique ») et Stuart Mill (éthique utilitariste de type conséquentialiste) pour la philosophie moderne.
Faut-il rendre une arme à son propriétaire s'il est devenu fou[5] ? Rendre à son propriétaire l'objet qui lui appartient est un devoir. Défendre la sécurité en est un autre. Conflit des devoirs. Si le propriétaire n'était pas fou, l'obligation de rendre l'arme s'imposerait directement. Mais si celui-là est devenu fou,

[4] Socrate (Platon, *Criton*) doit-il suivre le conseil de son ami et échapper grâce à son aide à une condamnation à mort injuste ou doit-il obéir à la loi qui l'a condamné mais qui l'a d'abord élevé ?
[5] Platon, *République* I, 331c.

c'est le devoir de sécurité qui l'emporte. Dans ce cas le dilemme est soluble. Il suffit de hiérarchiser les devoirs selon un ordre de priorité (sécurité collective>propriété individuelle) ou de tenir des circonstances particulières (l'événement de la folie). Le dilemme entre les deux devoirs était un faux dilemme.

Pour Thomas d'Aquin également, ce qui passe pour un dilemme moral résulte en réalité d'une mauvaise délibération de l'agent moral qui a mal ordonné les moyens pour la fin visée ou les devoirs relatifs entre eux (dilemme *secundum quid*). Et donc une théorie morale qui reconnaît l'existence de dilemmes en soi (*simpliciter*) est fausse.

Mais on peut avancer un autre argument pour supprimer le dilemme. Il n'y a peut-être jamais égalité entre deux devoirs et donc jamais non plus aucun dilemme, faute d'identité entre les objets des devoirs ou du même devoir. Soit A et B en train de se noyer, A et B étant jumeaux. Apparemment le devoir de sauver A = le devoir de sauver B, puisque A = B. Mais en application du principe leibnizien de l'identité des indiscernables[6], la position de l'observateur par rapport à A et à B à elle seule induit une différence. Si je sauve A plutôt que B de la noyade c'est que j'avais une raison de le faire : par exemple, je suis plus proche de A, ou A est plus proche de la rive, ou A a encore la tête hors de l'eau, etc. Donc c'est un raisonnement rétrospectif mais illusoire qui me fait croire que j'avais le choix égal entre A ou B ou le devoir égal de choisir de sauver A ou de sauver B.

Enfin, on peut considérer qu'à chaque fois qu'il y a dilemme, le critère moral n'a pas été appliqué correctement. Soit le cas de mon ami recherché par un homme qui veut le tuer et qui se réfugie chez moi[7]. L'assassin frappe à ma porte et me demande si cet ami est dans ma maison. J'ai le choix entre ne pas mentir et livrer mon ami à son assassin ou sauver mon ami au prix d'un mensonge. Y a-t-il dilemme ? Non car soit je considère que la vie d'un homme (devoir 1) et qui se trouve être mon ami (devoir 2) prévaut sur tout, même sur le devoir de vérité ; soit, au contraire, comme le préconise Kant, je considère que le devoir de vérité ou de véracité est inconditionnel alors que l'autre est conditionnel. Et même si je ne livre pas mon ami à son assassin, je sais que n'ai pas agi comme le devoir moral l'exige(r)ait. Donc pour Kant, il n'y a pas de conflit des devoirs parce qu'il n'y a qu'un seul devoir, l'impératif catégorique, qui implique de toujours dire la vérité. Et si néanmoins je n'agis pas par devoir mais contre le devoir, je sais que j'ai fait un choix immoral : j'ai préféré ou fait passer des intérêts particuliers avant le respect pour le devoir, j'ai lésé l'humanité (universel) en voulant sauver mon ami (particulier). Donc il n'y a jamais eu de dilemme moral.

---

[6] Ce principe énonce que si deux objets sont indiscernables, c'est-à-dire si tout ce qui est vrai de l'un est vrai de l'autre, alors ils sont identiques, ce qui a pour conséquence qu'il est impossible d'avoir deux individus qualitativement identiques (A = B) mais numériquement distincts (A et B).

[7] Kant, *Sur un prétendu droit de mentir par humanité*, 1797.

Le dilemme du tramway

Le dilemme moral a retrouvé un certain crédit dans la philosophie morale contemporaine[8], autour du dilemme dit du tramway. Le dilemme du tramway a été imaginé par la philosophe britannique Philippe Foot en 1967 (« The Problem of Abortion and the Doctrine of the Double Effect », *Virtues and Vices*, Oxford, Basil Blackwell, 1978). C'est une expérience de pensée, d'habitude présenté sous cette forme : une personne peut effectuer un geste qui bénéficiera à un groupe de personnes A, mais, ce faisant, nuira à une personne B ; dans ces circonstances, est-il moral pour la personne d'effectuer ce geste ?

« Imaginez qu'un juge se retrouve face à des émeutiers qui demandent qu'un coupable soit trouvé pour un certain crime et, si cela n'est pas fait, qui menacent de se venger de manière sanglante sur une partie spécifique de la communauté. Le coupable réel étant inconnu, le juge se retrouve avec pour seule solution de condamner à mort un innocent pour prévenir le bain de sang. Imaginons parallèlement un autre exemple où le pilote d'un avion qui est sur le point de s'écraser se doit de choisir un point de chute dans une aire plus ou moins habitée. Afin de rendre les deux situations le plus semblables possibles, imaginons plutôt le conducteur d'un tramway hors de contrôle qui se doit de choisir sa course entre deux voies possibles : cinq hommes travaillent sur l'une et un homme est situé sur l'autre. La voie prise par le tram entraînera automatiquement la mort des personnes qui s'y trouvent. Dans le cas des émeutiers, ces derniers ont en otage cinq personnes, ce qui fait en sorte que les deux situations amènent le sacrifice d'une vie pour en sauver cinq ».

On peut considérer qu'un acte moral est rationnel s'il peut s'appuyer sur une raison, si possible universalisable. Dans le cas du dilemme du tramway et de ses variantes, il paraît plus rationnel de sacrifier un pour sauver plusieurs. Ce critère est celui de la morale ou de l'éthique utilitariste. L'arbitrage est conforme à la maximisation de l'intérêt (de la fonction d'utilité) ou, ce qui revient au même, à la minimisation de la souffrance. Il est moral de réduire la quantité de souffrance et/ou d'augmenter la quantité d'utilité. La force de l'utilitarisme est de résoudre le problème du dilemme moral. L'intérêt ou l'utilité est le critère consistant de la morale. Ainsi l'automobile autonome sera morale en suivant l'utilitarisme — qui est l'éthique dominante dans le monde anglo-saxon ! En situation critique, elle sera programmée pour maximiser l'intérêt et minimiser la souffrance : froisser de la tôle pour épargner les passants, risquer de blesser une personne plutôt que plusieurs.
Mais le critère utilitariste est-il incontestable ? Faut-il programmer l'automobile selon le critère de la morale utilitariste — mais le critère d'une

[8] Il y a une littérature abondante en philosophie contemporaine, notamment anglo-saxonne, sur le dilemme, par exemple McConnell "Moral Dilemmas and Consistency in Ethics", *Moral Dilemmas*, New York, Oxford, Oxford University Press, 1987, et Williams (Bernard), "Ethical Consistency 1965 dans *La fortune morale*, Paris, Presses Universitaires de France, 1994.
- Williams (Bernard), "Consistency and Realism", *Proceedings of the Aristotelian Society*, 1966, réimprimé dans *Problems of the Self: Philosophical Papers*, Cambridge University Press, 1973.

autre éthique (déontologisme ou éthique des vertus) peut-il être programmé ? Car on n'a pas manqué de critiquer l'expérience de pensée du tramway.
1) sur un plan théorique, on peut contester la réduction de la morale à la psychologie, ou du moins l'abstraction de l'expérience de pensée, son caractère peu réaliste qui n'est alors d'aucun enseignement pour constituer un critère normatif de l'action ;
2) certaines variantes du dilemme font apparaître des choses surprenantes, comme celle de l'obèse proposée par Judith Jarvis Thomson[9] :

« Imaginez une situation — que j'appellerai « Homme obèse » — où vous êtes sur un pont sous lequel va passer un tramway hors de contrôle se dirigeant vers cinq ouvriers situés de l'autre côté du pont. Que faites-vous ? Étant un expert en tramways, vous savez qu'une manière sûre d'en arrêter un hors de contrôle est de placer un objet très lourd sur son chemin. Mais où en trouver un ? Au moment des événements, il y a un homme obèse, vraiment très obèse, à côté de vous sur le pont. Il est penché au-dessus du chemin pour regarder le tramway. Tout ce que vous avez à faire est de lui donner une petite poussée pour qu'il tombe sur les rails et bloque le tramway dans sa course. Devriez-vous donner cette poussée ? Tous ceux à qui j'ai posé cette question m'ont répondu non. Mais pourquoi ? ».

Or, 50% des individus interrogés qui approuvent le sacrifice de 1 individu pour 5 (la version originale du dilemme), n'approuvent pourtant pas le geste dans cette variante. Et de manière encore plus inattendue, selon des études plus récentes menées par Joshua Green, professeur de psychologie de Harvard[10], si on propose de pousser l'innocent (obèse) de sa passerelle non pas avec l'épaule ou la main mais avec une perche, beaucoup plus de personnes se déclarent prêtent à faire mourir l'inconnu — comme s'il était plus grave et moralement signifiant de pousser l'innocent avec la main qu'avec un instrument. De ces variantes du dilemme du tramway, on peut tirer la conclusion que le sujet moral éprouve bien une tension entre des principes moraux : la maximisation du bien, minimisation du mal d'un côté, l'interdit de tuer de l'autre. Et il semblerait que la morale emprunte deux circuits neuronaux différents, selon qu'elle suit l'émotion (que Green nomme le mode « automatique », inconscient, peut-être sélectionné par l'évolution) qui est la voie la plus courte, la plus générale, ou la raison critique, consciente qui calcule les conséquences (le mode « manuel ») de l'action. Donc le calcul utilitariste ne peut passer ni en droit ni en fait pour l'action morale. On obtiendrait encore d'autres démentis avec des variables supplémentaires : si l'individu isolé est un proche, et les autres des inconnus ; si l'individu isolé est un immense savant et les autres des hommes ordinaires ; si l'individu est un enfant, les autres des vieillards, ou si l'individu est une victime et les autres des criminels…
Mais évidemment, ces éléments sont non pertinents dans le dilemme, où précisément la situation exclut toute différence qualitative entre les individus

[9] *The Trolley Problem*, 1985.
[10] *Tribus morales*, Markus Haller, 2017.

et inclut au contraire que 1 = 1. L'automobile autonome est dans la même situation de ne pas faire acception des personnes. Donc il sera moral de la programmer selon le critère de la morale utilitariste. Le passage à l'automobile autonome peut encore de s'autoriser de l'utilitarisme s'il est vrai que, selon les projections les plus optimistes, son déploiement à large échelle permettrait d'éviter 90 % des accidents qui sont dû à une erreur humaine : conducteur fatigué, sous l'emprise de l'alcool ou d'une drogue, qui ne respecte pas le code de la route[11]. Autrement dit, c'est la faillibilité humaine qui est le problème plus que l'automobile autonome. La voiture autonome peut être le moyen technologique de réduire la quantité de souffrance (accidents) dans la circulation automobile.

Le dilemme moral de l'automobile autonome

Mais cette conclusion ne supprime pas tout problème éthique pour la voiture autonome. Il faut souligner la singularité éthique de l'automobile autonome. On peut retenir trois arguments :

1) Le dilemme moral de l'automobile autonome n'est pas si éloigné que cela du dilemme du tramway. L'automobile "doit"-elle sacrifier systématiquement le passant plutôt que le conducteur, ou le conducteur même si le passant ou l'autre véhicule est en infraction … ? Il y a pourtant une différence majeure. Le dilemme du tramway était une expérience de pensée pour laquelle on pouvait refuser de choisir au nom de l'abstraction et de la fiction du scénario — le sujet moral se refugie encore dans le choix de ne pas choisir —, alors qu'avec l'automobile autonome il est requis de proposer une procédure de choix. L'automobile autonome contraint de choisir une programmation éthique.

2) L'automobile autonome illustre une nouvelle fois la conséquence éthique du progrès technoscientifique. En un sens, le monde technoscientifique n'est pas moins éthique mais (toujours) plus éthique que l'ancien monde, puisque devient objet de responsabilité ou de choix éthiques ce qui ne l'avait jamais été parce que la nature ou le hasard se chargeait de résoudre les problèmes à la place de l'homme — avant le premier respirateur artificiel, la question de l'acharnement thérapeutique ne se posait pas. En l'occurrence ce qui est singulier avec l'automobile autonome, y compris par rapport à d'application de l'IA à d'autres domaines (juridique, chirurgical), c'est de confier à un programme-machine une décision que nous n'avons jamais prise de manière réfléchie. En situation critique, un conducteur réagit comme il peut, par réflexe en cherchant à éviter un obstacle et à sauver sa vie. L'action ne souffre pas

[11] «Algorithm Aversion : People Erroneously Avoid Algorithms After Seeing Them Err », *Journal of Experimental Psychology*, 2014.

de délai et donc ne laisse pas le loisir au conducteur de choisir entre plusieurs options éthiques. L'accident se présente toujours comme un événement, que chacun négocie comme il peut en situation, et non sous la forme d'un dilemme. Surtout, l'IA oblige à choisir éventuellement entre un voiture autonome exclusivement altruiste (choisissant de sauver systématiquement les piétons par exemple, même au risque de la mort du conducteur et des passagers) ou égoïste (choisissant de sauver systématiquement le conducteur et les passagers au risque de la mort des piétons). Et de fait, les enquêtes d'opinion montrent la préférence forte pour un véhicule qui assure en priorité la sécurité de leur utilisateur et de leur propriétaire — sinon la voiture deviendrait une machine sacrificielle, peut-être alors superlativement éthique mais contre son usager, ce qui en fait un argument de vente douteux.

3) Enfin le cas de l'automobile autonome met aux prises de manière exemplaire le conflit entre l'impératif économique (qui n'est pas éthique) et l'impératif éthique. Du moins l'impératif économique se présente-t-il comme un impératif hypothétique : a) Si tu veux une circulation automobile qui réduise la quantité de risques absolus, tu dois vouloir l'automobile autonome ; b) si tu veux favoriser l'avenir et l'essor de l'industrie automobile, alors tu dois financer la fabrication de l'automobile autonome. Mais doit-on vouloir l'automobile autonome qui priverait, dans le cas d'une autonomisation complète, le sujet de sa responsabilité ?

Mais les réticences éthiques les plus fortes relèvent plutôt d'une éthique de type prudentielle. L'erreur est humaine : on la juge, on la sanctionne et éventuellement on la pardonne. Pour la machine, rien de tel. Qu'est-ce qu'être propriétaire d'un véhicule sans la responsabilité de sa conduite ? En cas d'accident, qui sera(it) responsable ? Surtout, il est illusoire d'opposer la machine et la faillibilité humaine : toute l'histoire technologique est là pour rappeler que l'infaillibilité mécanique est un leurre, même si tout système technique progresse et que l'IA apprend de ses erreurs. On sait surtout que tout système informatique est piratable. Est-il raisonnable de parier sur un système de cyber-sécurité infaillible des programmes d'automobile autonome ? Non seulement le conducteur est dépossédé de sa responsabilité, mais la sécurité qui gère la sécurité de son véhicule lui échappe sans qu'elle échappe pourtant à des individus ou des sociétés malveillants.

Mais finalement le dilemme éthique de l'automobile autonome ne se pose que dans un monde mixte d'automobiles autonomes et d'automobiles non autonomes. Imaginons un réseau routier, un parc automobile où tous les véhicules seraient autonomes : ce serait un monde qui aurait supprimé non seulement l'accident, mais le dilemme moral. Cette expérience de pensée (d'un parc automobile intégralement autonome) supprime l'expérience de pensée du dilemme moral. En effet, le dilemme moral de l'automobile autonome ne se pose que pour un monde technologiquement insuffisant. Le dilemme moral de l'automobile autonome ne serait-il qu'une étape vers la suppression

technologique du dilemme moral, s'il s'avérait que la technoscience est moins la solution dont l'homme a besoin que l'homme n'est le problème-obstacle au progrès de la technoscience ? Le cercle sens-finitude en contient peut-être un autre : morale-faillibilité. Si la technologie supprime la finitude (transhumanisme) ne supprime-t-elle pas la condition du sens ? Si elle supprime la faillibilité, ne supprime-t-elle pas la signification de l'éthique ?

L'automobile autonome est donc, au total, un nouvel objet étrange. Elle satisfait l'individualisme des sociétés contemporaines — on pourrait imaginer une autre politique économique en faveur du transport en commun, en lui appliquant les progrès de l'IA — mais en ôtant à l'individu la responsabilité de sa conduite et peut-être au passage, le plaisir de conduire. Que sera donc une automobile autonome — dans un horizon plus proche de 2050 que de 2025 ? Un deuxième bureau, un salon transitoire où l'on pourra boire autant qu'on voudra puisqu'on n'aura plus à conduire ? Nous gagnerons en sécurité, nous gagnerons du temps, mais peut-être au prix du risque de nos libertés même. L'automobile autonome représentera donc assurément une mutation essentielle dans l'anthropologie de la mobilité. Bienvenue dans le monde de l'autonomobile.

Laurent Cournarie