



La RICHARD HANLEY

La métaphysique
de *Star-Trek*

Bayard Éditions

PHILOSOPHIE

La métaphysique de Star Trek

Par Richard Handley

Traduit par Camille Cantoni-Fort

Introduction

L'entreprise philosophique

L'espace, l'ultime frontière. C'est lui que traverse le vaisseau spatial l'Entreprise. Sa mission de cinq ans: chercher une vie nouvelle, de nouvelles civilisations, et se rendre hardiment là où nul homme n'est jamais allé.

Le capitaine James Tiberius Kirk

Voilà de belles perspectives d'exploration! li ne s'agit plus de tracer le plan des étoiles, ou d'étudier des nébuleuses, mais de dresser la liste des possibilités d'existence inconnues.

Q à Picard, dans l'épisode « Les meilleures choses ... »

Philosophie signifie littéralement «l'amour de la sagesse», mais cette définition ne vous apprendra sans doute pas grand-chose. À mon sens, mieux vaut considérer la philosophie comme une activité, une entreprise qui consiste, sommairement, à se mettre en quête de sagesse. La mission d'une entreprise philosophique ressemble donc assez à celle du vaisseau spatial qui porte ce nom. La métaphysique - littéralement, « ce qui vient après la physique » - est cette branche de la philosophie qui explore la nature de l'existence.

Et pourquoi l'explorer ? Parce que les voyages forment l'esprit.

Lorsque nous visitons d'autres lieux, d'autres peuples, nous croisons souvent des modes de vie différents, et cela nous incite à réexaminer les nôtres. Si cette observation demeure objective et raisonnable, nous en concluons parfois qu'il existe plus d'un mode de vie, ou que celui des autres n'est pas fait pour nous, ou que nous devrions au contraire l'adopter. Le voyage ne consiste pas simplement à découvrir d'autres lieux ou d'autres peuples: c'est aussi un voyage en nous-mêmes, une mise à l'épreuve de nos propres limites. Bien sûr, tous les voyageurs ne sont pas objectifs et raisonnables - certains croient déjà savoir comment vivre, et ne vont à la rencontre des tenants d'une autre pensée que pour les convertir. Mais, en règle générale, ce sont les voyageurs qui sondent leur propre existence, et non les pantouflards qui ne voient pas plus loin que leur foyer et leurs petites habitudes. Le voyage forme l'esprit parce qu'il donne à sentir la réalité d'autres lieux - il permet au voyageur (momentanément

du moins) de se détacher d'un foyer confortable qui lui est devenu familier, et d'explorer un contexte différent. Les voyageurs le savent, la meilleure solution est encore de vivre loin du pays: c'est une occasion de s'immerger dans un lieu inconnu. La nature humaine nous porte à désirer ce qui nous est familier et à nous plaindre de ce qui ne l'est pas, et il faut du temps pour surmonter ces préjugés et percevoir un cadre nouveau dans toute son authenticité.

Star Trek nous propose une vision optimiste, celle d'un avenir où nous aurions entièrement exploré la Terre et tiré les leçons de cette exploration. Au XXIV^e siècle, notre planète s'est dotée d'un gouvernement mondial, pacifiste et éclairé, qui a mis fin à toutes les discriminations de race, de sexe et de genre. Bien évidemment, l'exploration a reculé ses frontières : elle s'étend désormais à l'espace planétaire. Tout comme les premiers explorateurs terrestres ont découvert d'autres formes de vie et d'autres civilisations humaines, les explorateurs de l'espace découvrent des formes de vie et des civilisations extraterrestres. Ces explorations leur ont ouvert l'esprit.

La mission du vaisseau spatial l'Entreprise se heurte néanmoins à une limite incontournable : il ne saurait découvrir quelque chose d'absent. Il se peut qu'il manque ainsi d'importantes alternatives aux modes de vie et de pensée admis, pour la simple raison qu'on ne les a jamais prises en considération. C'est ici que la philosophie fait son entrée : les philosophes, eux, tentent d'examiner toutes les possibilités. Si l'Entreprise traverse l'espace extérieur, l'entreprise philosophique traverse un espace logique en visitant divers mondes possibles, à savoir des ensembles de représentations dotés d'une cohérence interne. Elle prend le temps nécessaire pour se faire une idée de chaque monde. Sa mission constante: aller hardiment là où peu d'esprits humains sont allés. Et parfois rentrer à la maison avec une expérience enrichie, puisque son objectif premier est l'autoexploration, l'exploration d'un espace intérieur.

C'est là une activité qui ne va pas de soi. Non seulement il est quelquefois ardu de dégager les possibilités diverses, mais il est encore plus difficile de suspendre ses croyances et habitudes assez longtemps pour tester sérieusement ces nouveaux mondes. Et pourtant, c'est là la tâche des philosophes, et leur apprentissage, s'il est sérieux, les libère progressivement de leurs propres préjugés. Dans ce but, il leur arrive de faire usage de l'expérience mentale - une expérience menée simplement en pensée. Il faut un certain talent pour concevoir ces expériences, pour imaginer les situations expérimentales avec assez de réalisme, et c'est le rôle de la fiction. La bonne fiction, et notamment la bonne fiction visuelle, sollicite nos facultés d'imagination. Voilà en quoi Star Trek « fait de la philosophie ». Nombre des situations présentées dans cette série reflètent les réflexions des philosophes, et, en observant un scénario de Star Trek dans ses diverses péripéties, nous sommes contraints d'accueillir des idées que nous aurions refoulées par préjugé. Durant une cinquantaine de minutes, nous vivons dans un monde différent que nous mettons à l'épreuve.

Observer la philosophie à travers le prisme de Star Trek, voilà la stratégie que

j'ai adoptée avec succès auprès de mes étudiants. Dans cet ouvrage, toutefois, j'ai inversé les choses : j'observe Star Trek à travers le prisme philosophique. Je ne viens pas louer cette fiction, ni l'ensevelir. Le canon de Star Trek offre matière à réflexion: la première série (abrégée désormais en PS), La nouvelle génération (NG), Profondeur Espace 9 (PE9), Le voyageur (V), et les films longue durée. Ce qui m'intéresse, c'est de montrer comment un éclairage philosophique peut accroître votre compréhension de Star Trek, puisque cette série est par nature philosophique.

Encore deux brèves remarques sur les divers courants philosophiques. En ce qui me concerne, j'appartiens à l'école «anglo-américaine» de philosophie, dite école analytique, et accepte de plus la pensée naturaliste, stipulant que la philosophie est et devrait être rattachée aux sciences naturelles, puisque ces deux domaines recourent à la fois à la raison et à l'imagination empirique. C'est pourquoi le prisme philosophique adopté ici, qui insiste tout particulièrement sur les processus et moyens technologiques dépeints dans Star Trek, n'est pas représentatif de la philosophie dans son ensemble. Par exemple, certains philosophes considèrent qu'il est impossible d'avoir une approche objective et scientifique lorsqu'on étudie la nature de l'esprit, qu'il faut plutôt partir de l'intérieur, de l'expérience mentale en ce qu'elle présente de subjectif et de phénoménologique. D'autres philosophes s'intéresseraient sans doute au mysticisme révélé par les Vulcains, les Bajoriens, ou le commandant américano-indien Chatokay. D'autres se pencheraient sur une mythologie de Star Trek comme «texte» ou « icône culturelle». Ce n'est pas là mon objet. Par ailleurs, l'approche analytique elle-même a suscité des désaccords sur plus d'un point philosophique, et mon opinion relève parfois d'une minorité. Je présenterai et défendrai souvent la position qui me paraît la plus valable, mais qui ne doit pas être considérée comme représentative de l'opinion philosophique générale. Cela dit, j'examinerai aussi les perspectives majoritaires, quitte à les rejeter ensuite.

Quant à la métaphysique, elle peut être plus ou moins pure ou appliquée. Star Trek pose des questions de métaphysique pure, comme dans « Là où nul n'est allé », où le jeune Wesley Crusher se demande si la pensée est au fondement de toute existence. Cet ouvrage, toutefois, se concentre sur la métaphysique appliquée - celle qui concerne la vie quotidienne de tout un chacun. Il se veut relativement abordable: ainsi, j'ai évité de le surcharger de notes. D'où un ouvrage sur les idées plutôt que sur les philosophes ou sur l'histoire de la philosophie, où j'ai essayé de m'en tenir à des perspectives contemporaines, notamment celles qui portent sur les frontières mouvantes qui séparent la science et la métaphysique. Chaque fois que c'était nécessaire, j'ai mentionné les philosophes auxquels on doit telle position ou tel argument. Les lecteurs désireux d'étendre leur recherche pourront consulter la bibliographie placée en fin de volume. Ce texte, je l'espère, est une introduction utile aux débats contemporains sur la place de l'humanité dans le monde, débats qui sont eux-mêmes au cœur de Star Trek.

Première partie

**Une nouvelle vie,
de nouvelles civilisations**

Chapitre UN

Suspects numéro un

Identifier de nouvelles formes de vie, quelles qu'elles soient, est une des missions principales du vaisseau.

Jean-Luc Picard, « La qualité de la vie » (NG)

Toute fiction parle de la condition humaine, et Star Trek ne fait pas exception. Ce qu'il y a d'unique dans cette fiction populaire, c'est l'extension qu'elle donne aux préoccupations humaines. Dans chacun des épisodes, les personnages les plus intéressants sont le plus souvent ceux qui nous ressemblent, sans être tout à fait comme nous. Au lieu de traiter l'étrange et le différent comme quelque chose de redoutable - qu'il faut détruire - Star Trek nous incite à penser la vie d'un point de vue non humain. La série remplit donc deux respectables fonctions : elle nous apprend quelque chose sur nous-mêmes, et soumet à la conscience populaire (et par là remet en question) la conviction commune qui accorde aux membres de l'espèce dite Homo sapiens une position privilégiée dans le cosmos. Cette conviction est-elle justifiée, ou est-elle pure hybris? Quelles sont, si nous en avons, nos obligations envers ceux qui nous ressemblent sans être comme nous? Un principe moral comme la règle d'or s'étend-il à nos rapports avec les non-humains? À quel point une espèce doit-elle différer des humains pour que la règle d'or cesse de s'appliquer ?

Star Trek nous oblige à méditer ces questions en examinant la limite entre l'humain et le non-humain sous plusieurs angles, notamment à travers le choix des personnages principaux. Dans la première série, le second de l'Entreprise est le célèbre M. Spock, mi-humain, mi-Vulcanien. Spock semble quasi méprisant envers l'humanité, et fait des efforts continuels pour réprimer ses penchants humains au profit de son côté vulcanien. Lorsque le médecin du vaisseau, Leonard McCoy, affirme dans « Cour martiale » que Spock est l'individu le plus insensible qu'il ait jamais rencontré, ce dernier réplique : « Oh ! merci, docteur. » Et lorsqu'à la fin de « Démon dans le noir », au cours de la rituelle pause-détente sur la passerelle, Kirk dit : « Je vous soupçonne de devenir plus humain chaque jour », Spock répond : « Capitaine, je ne vois pas pourquoi je resterais ici à me faire insulter. »

Un contrepoint au préjugé de Spock est présenté à travers trois personnages centraux de la série suivante. Dans La nouvelle génération, le comportement de

l'androïde Data mime largement celui des humains, et si nous en croyons Data (mais le faut-il? c'est une question que nous étudierons au chapitre 3), il aspire essentiellement à devenir humain. Dans Profondeur Espace 9, le chef de la sécurité, Odo, peut changer de forme : après avoir passé tout son temps parmi les humains, il a adopté leur apparence extérieure et se sent manifestement plus à l'aise avec eux que parmi les siens, qui règnent en tyrans sur la majeure partie du quadrant Gamma. Dans Le voyageur, un « hologramme médical d'urgence » est chargé d'un service permanent après que le médecin du vaisseau a été tué, et, bien qu'il ne puisse quitter le holodeck, on le voit peu à peu aspirer à l'humanité. Mais quelle est la nature de l'humanité, et pourquoi peut-on désirer être humain?

Logique contre émotion

Seule la discipline de la logique a sauvé ma planète de l'extinction.

Spock à Bélé, « Ce sera là votre dernier champ de bataille » (PS)

Je ne regrette rien.

Spock au lieutenant-commandant Data, « Unification II » (NG)

Qu'a donc Spock contre l'humanité? Ceci seulement que les humains sont trop soumis à l'influence de leurs émotions, et pas assez à celle de la logique. Dans leur antique passé, les Vulcariens étaient si agressivement émotionnels qu'ils passaient tout leur temps à se battre, ce qui, en toute probabilité, aurait dû les amener à périr dans une guerre civile perpétuelle. La solution adoptée par les Vulcariens fut d'éradiquer leurs émotions grâce à un entraînement rigoureux, et à se soumettre aux dictats de la logique (nonobstant le Pon farr - leur rut septennal). Le Dr McCoy, à l'inverse, cède souvent à son émotion, et affirme dans de nombreux épisodes qu'il rejette l'approche logicienne de l'existence privilégiée par Spock. McCoy oppose à la froideur de Spock des réactions émotionnelles et instinctives, tout en adoptant cependant un air non moins supérieur. Le capitaine Kirk représente peut-être la quintessence de l'humain, en ce qu'il fait preuve de logique et d'émotion, mais il a des moments de sympathie profonde avec McCoy : « Votre logique est parfois agaçante au plus haut point », fait-il remarquer à Spock dans « Demain est hier ».

Les non-humains semblent largement d'accord avec Kirk et McCoy pour dire qu'il vaut la peine de se laisser envahir par l'émotion ou la déraison. En voici quelques exemples rapides. Dans « Ou tout autre nom » (PS), les arrogants Kelviens adoptent une forme humaine pour prendre le pouvoir sur l'Entreprise : ils en viennent rapidement à apprécier certains des aspects les plus vils de l'humanité, comme le sexe et l'alcool. Dans « Indices » (NG), les Paxiens, bien connus pour leur xénophobie, décident de détruire l'Entreprise-D et son équipage pour protéger le secret de leur existence, mais, impressionnés par la détermination humaine à survivre, ils finissent par changer d'avis. Q, cet être immortel et omnipotent venu d'une autre dimension de

l'existence, tourmente l'équipage de l'Entreprise-S dans divers épisodes de La, nouvelle génération, précisément parce qu'il trouve l'humanité illogique, donc infiniment fascinante.

Dans la série originale, nous avons certes l'impression générale qu'il vaut la peine d'avoir des émotions violentes. Dans « Ce côté du paradis », l'équipage tout entier de l'Entreprise succombe à l'influence de spores végétales provoquant le délire, et c'est seulement lorsqu'il cède à une fureur sans bornes que Kirk y échappe. Dans « Les années mortelles », la délégation qui visite Gamma Hydra IV est tout entière contaminée par un virus accélérant le vieillissement, à l'exception de l'enseigne Pavel Chekov. Alors que Kirk et McCoy sont proches de la mort, on découvre que Chekov n'est pas affecté pour la simple raison qu'il a éprouvé une grande peur lors de son séjour sur la planète. La réaction de l'adrénaline a tué le virus, et McCoy élabore un antidote à partir de ce constat. Revenons sur terre: nous expliquons cette réceptivité à une émotion telle que la peur comme une capacité d'adaptation liée à notre évolution: la peur empêche une créature vulnérable d'adopter une attitude suicidaire en cas de danger; plus généralement, l'adrénaline prépare l'homme au combat ou à la fuite. (Il est plus sage, toutefois, de ne pas invoquer systématiquement l'adaptation pour expliquer l'origine de chacune de nos caractéristiques : souvent, tout ce que nous savons d'une telle caractéristique est qu'elle n'a pas fait l'objet d'un choix préalable.)

Spock et McCoy incarnent deux traditions philosophiques diamétralement opposées sur le rapport entre raison (la faculté mentale qui produit et évalue les croyances) et passion (la faculté mentale qui produit les désirs et les émotions). Selon les philosophes grecs de l'Antiquité Platon et Aristote, la condition humaine se définit par une lutte entre raison et passion - où la raison devrait l'emporter. Selon l'autre tradition, étroitement associée au philosophe écossais du xviii^e siècle David Hume, il n'y a pas de conflit du tout, et la passion est prééminente. Ce sont les désirs et les émotions qui motivent le comportement, et, pour citer la célèbre formule de Hume, « la raison est, et ne devrait qu'être, l'esclave des passions ». Sommairement, la raison peut vous dire comment obtenir ce que vous désirez, mais ne peut vous dire quel est l'objet de ce désir.

Qui a raison? Spock et les Grecs, ou McCoy et Hume? Je suis enclin à penser que la vision la plus défendable de la nature humaine est quelque part entre les deux. On est tenté de considérer les passions - émotions et désirs - comme des états que nous subissons. Nous sommes en quelque sorte les victimes de la passion, alors que la raison, l'application délibérée de la logique dans la formation et l'évaluation des opinions, dépend entièrement de notre volonté. Mais je doute que les choses soient si simples. Supposons que je vous dise (et vous me croyez) que l'université de Monash est à Victoria, et que Victoria se trouve en Australie. Comment ne pas en conclure que l'université de Monash se trouve en Australie? Voici un encouragement : je vous donne un million de dollars si vous vous persuadez tout à fait sincèrement que l'université de Monash est à Victoria, que Victoria est en Australie, mais que l'université de Monash n'est pas en Australie. Sans aucun doute, la logique dit qu'il est dans votre intérêt de

faire ce que je demande. (J'attends que vous ayez essayé... ouf ! Encore un million d'économisé !) Raison et passion échappent largement à un contrôle délibéré, mais toutes deux peuvent être cultivées jusqu'à un certain point. On peut exercer ses capacités logiques, comme les Vulcaniens, mais aussi ses désirs et ses émotions.

User de la logique pour tenter de contrôler ses émotions, comme le fait Spock, a donc un sens. Mais supprimer entièrement les émotions ? En aurions-nous au contraire besoin pour motiver notre action, comme l'affirme Hume ? Mon sentiment est que Hume a raison, au moins en partie. Les émotions peuvent donner la clef d'un puzzle : elles peuvent contribuer à la capacité humaine ordinaire de résoudre ce que les théoriciens de la connaissance appellent le problème du cadre. Pour accomplir une action, un agent cognitif - moi, par exemple - doit choisir entre diverses options. La réflexion rationnelle semble poser comme condition minimale l'examen des conséquences probables d'un nombre stupéfiant de possibilités. Cela vaut pour les plus simples tâches, par exemple quitter mon ordinateur pour rejoindre le frigo. Devrai-je me mettre debout ? Lever mon bras d'un centimètre ? Siffler ? Chanter ? Contracter mes abdominaux ? Énumérer toutes les actions possibles même dans un futur proche produit une explosion de combinaisons supplémentaires. Qui plus est, pour accomplir cette action je dois prendre conscience de ce que je sais, et tirer de ce savoir des conclusions appropriées à l'accomplissement de ma tâche. Mais voilà : je sais un nombre incalculable de choses, et dois d'une façon ou d'une autre éliminer celles qui n'ont rien à voir avec la question - que nous sommes mercredi, que la lune n'a pas d'atmosphère - sans passer par la méthode brutale consistant à examiner chaque élément d'information en fonction de l'action envisagée. Le problème du cadre recouvre en fait des problèmes nombreux et divers, que nous pouvons résumer en ces termes : un agent cognitif doit savoir à quoi penser sans d'abord réfléchir à ce à quoi il lui faut réfléchir.

Une façon de résoudre le problème du cadre est de diviser la pensée en divers sous-ensembles ou facultés, dont chacune accomplit tout simplement le travail qu'elle est censée fournir. (Lorsque je me mets debout, je n'ai pas à contracter consciemment mes abdominaux - d'une façon ou d'une autre, ça se fait « sans moi ».) Pour le dire sommairement, le problème du cadre est éliminé dans certaines tâches, qui échappent aux griffes de la délibération, et qu'on sous-traite à des processus mentaux spécialement adaptés. Mais la souplesse de comportement des agents cognitifs que nous sommes suggère que la question n'est pas entièrement résolue. Il semble que nous disposions d'un mécanisme plus global pour réduire la visée de notre pensée : c'est peut-être ici qu'interviennent les émotions. Il est bien entendu que les émotions affectent l'attention ; lorsque nous aimons quelqu'un, nous voyons ses vertus et restons aveugles à ses défauts, pour user d'un exemple trivial. Si - et c'est un « si » extrêmement spéculatif - nous n'avions pas d'émotions, nous serions toujours paralysés devant l'immensité des tâches les plus simples.

Cette explication d'une interdépendance entre émotion et logique fait l'objet d'un habile traitement allégorique dans « L'ennemi intérieur » (PS). Une défaillance du téléporteur « scinde » Kirk en ses parts animale et cérébrale. Bien que physiquement

identiques, ces deux individus ont un comportement très différent: le Kirk animal, dépossédé du contrôle de sa raison, essaie de violer un membre féminin de l'équipage. Le Kirk cérébral semble tout d'abord bien adapté, mais il devient clair que, privé de ses émotions et instincts inférieurs, il est faible et indécis, incapable de commander. Chaque moitié est incomplète sans l'autre. L'être humain Kirk a besoin à la fois de logique et d'émotion s'il veut être un agent cognitif accompli.

Spock pourrait convenir que certaines émotions sont utiles. Il y a une distinction plausible entre ce qu'on appelle souvent les émotions supérieures et les émotions inférieures (il est assurément plus juste d'imaginer une répartition des émotions le long d'un spectre avec les émotions inférieures à une extrémité, les émotions supérieures à l'autre, mais nous pouvons ignorer cela). Les émotions inférieures sont enracinées dans le système limbique et le pédoncule cervical - dit cerveau reptilien - alors que les émotions supérieures relèvent du cortex. Nous partageons les émotions inférieures avec bien d'autres animaux - les exemples les plus évidents en sont la crainte et la colère. Les émotions supérieures sont, significativement, absentes chez les autres animaux, y compris l'espoir et le regret. En général, nous associons des sentiments très puissants - ou affects, pour user du terme technique - aux émotions inférieures, alors que les émotions supérieures semblent caractérisées plutôt par le fait qu'elles accompagnent des états cognitifs. Par exemple, si on regrette d'avoir fait quelque chose, l'on croit qu'il aurait mieux valu faire quelque chose d'autre, et l'on s'en veut de ne pas l'avoir fait.

Certes, les Vulcaniens sont mieux dotés en émotions supérieures. On ne saurait douter que Spock se préoccupe sincèrement de ses amis (je ne crois pas que cela participe de sa moitié humaine - les Vulcaniens manifestent ce trait). C'est l'admiration de Spock pour la logique et son espoir de la maîtriser qui le motivent. Lorsque Spock dit à Data qu'il ne « regrette rien » (voir l'exergue de cette section), il ne dit pas que le regret est une émotion qu'il n'éprouve pas - bien au contraire, Spock a bien des regrets, suscités par exemple par ses mauvais rapports avec son père, Sarek -, mais qu'il ne regrette pas particulièrement d'avoir suivi l'enseignement vulcanien. Ce sont seulement les émotions inférieures (auxquelles Kirk, McCoy et les scénaristes se complaisent de temps à autre) que Spock cherche à supprimer.

«Humain» et « non humain» : qu'est-ce que l'humain ?

Colon: « Vous n'êtes pas humain. »

Data: « C'est exact. Je suis un androïde. »

« Les enseignes du commandement » (NG)

D'horribles sacs géants, essentiellement remplis d'eau ...

**Le micro-cerveau, à propos de l'équipage de l'Enterprise-D
dans « Le sol natal » (NG)**

Puisque Spock est à moitié humain génétiquement, n'est-il pas absurde qu'il

dédaigne son humanité? Puisque Data n'appartient assurément pas à l'espèce humaine, n'est-il pas absurde qu'il aspire à être humain? Oui et non. «Humain», au sens que nous donnons ordinairement au mot, a bien trois significations distinctes. Suivant la première, on est humain au sens où on relève du genre Homo. Dans ce sens biologique, Data ne sera jamais humain. Star Trek suggère qu'on y peut inclure certaines espèces extraterrestres, tels les Vulcانيens, les Klingons, les Romuliens, les Ferengi, les Bajoriens et les Cardassiens. Il y a des similarités génétiques dans toutes ces espèces, et dans « La chasse » (NG) l'équipe de l'Entreprise-Den apprend la raison. Une énigme introduite dans le code génétique de diverses espèces humanoïdes conduit l'Entreprise-D vers un message du passé : dans tout le quadrant Alpha, une race antique a laissé sa semence génétique sur diverses planètes, d'où l'origine commune des espèces humanoïdes. Je m'en tiendrai au vocabulaire de Star Trek et utiliserai « humanoïde » pour désigner le sens strictement biologique d'«humain» - celui que les colons cités en exergue ont en tête lorsqu'ils font remarquer que Data n'est pas humain.

On est humain au second sens du terme si on a, à peu près, les mêmes caractéristiques psychologiques que les êtres humains terrestres arrivés à maturité. C'est en ce sens que Spock se bat pour être moins humain et que Data aspire à l'être. C'est en ce sens que Kirk accuse Spock de devenir « de plus en plus humain » et que Picard, dans « La qualité de la vie », définit une décision de Data - mettre en jeu son existence à lui, Picard, pour empêcher la destruction de machines potentiellement intelligentes - comme « la décision la plus humaine » qu'il ait jamais prise. Je marquerai ce second sens psychologique en mettant humain en italique.

Le troisième sens ordinaire d'«humain» est le plus subtil, mais aussi le plus important; c'est le sens moral, ou éthique, du terme. Lorsque les gens parlent de « droits de l'homme », par exemple, c'est dans ce sens moral.

Star Trek nous aide à examiner ce troisième sens, que les scénaristes désignent en parlant de «sensibilité». Aussi longtemps que nous restons enracinés sur terre, il est difficile de voir à quoi rime ce troisième sens, puisque nous sommes habitués à l'idée que les humanoïdes et les humains sont les seules sortes d'individus qui ont (par exemple) le droit à la vie. La Fédération unie des planètes reconnaît non pas les droits de l'homme, mais ceux du « sensible », et il est clair qu'elle étend cette notion de droits aux autres humanoïdes, pour le moins. Il serait aussi condamnable (toutes choses égales par ailleurs) de tuer McCoy que de tuer Spock.

Il nous faut ici abandonner l'usage fixé par Star Trek, puisqu'un individu doué de sensibilité est simplement un individu conscient de ses impressions sensuelles. Si la sensibilité est nécessaire à l'«humanité» au sens moral, elle est probablement insuffisante. Un poulet est sensible, mais je doute que nous incluions jamais les poulets dans notre discussion sur les « droits de l'homme ». Un meilleur terme - celui que les philosophes utilisaient pour cerner ce sens moral de l'«humain» - est celui de personne. La question philosophique importante que Star Trek peut nous aider à résoudre est: qu'est-ce qui constitue la personne? Être humanoïde n'est pas suffisant, puisqu'un cadavre humanoïde n'est pas une personne. Et s'il est indubitable

qu'être humain ne suffit pas à faire une personne, l'exemple de Spock montre que ce n'est pas strictement nécessaire. Mais faut-il être humanoïde pour être une personne? Star Trek, dans sa célèbre quête d'une « vie nouvelle et de civilisations nouvelles », nous montre que la réponse est non.

Vie nouvelle, civilisations nouvelles, et nouvelle morale

*Je ne crois pas qu'il soit justifié de sacrifier
une forme de vie pour une autre.*

Data au commandant William T. Rike, « La qualité de la vie » (NG)

*Si le vaisseau est véritablement une intelligence en voie
d'émergence, il est de notre responsabilité de le traiter
avec autant de respect que toute autre créature.*

Picard aux officiers supérieurs, « Urgence » (NG)

Star Trek examine la possibilité d'une vie nouvelle dans quatre catégories distinctes de la vie extraterrestre, à savoir la vie suscitée par des moyens naturels dans d'autres mondes, la vie artificielle, la vie intelligente et la vie « sensible » (celle des personnes). Bien sûr, ces catégories ne s'excluent pas mutuellement (vous, par exemple, vous êtes vivant(e), intelligent(e), et vous êtes une personne) et il peut ne pas y avoir de vie extraterrestre ou artificielle du tout.

Dans « La qualité de la vie », Data dit au médecin-chef de l'Entreprise-D, le Dr Beverly Crusher: « Je m'interroge sur ce qui s'est passé entre le moment où je n'étais rien de plus qu'un assemblage de pièces ... et le suivant, où je suis devenu vivant. Qu'est-ce qui m'a donné la vie? » Si on met de côté pour l'instant la question de savoir si Data est ou non vivant, nous pouvons nous poser la même question. Qu'est-ce qui nous a donné la vie?

Lorsque les scientifiques parlent de la possibilité d'une vie extraterrestre, ils ont généralement en tête une catégorie purement biologique : le débat récent sur la vie sur Mars est un cas d'espèce. Et bien sûr, pour déterminer si quelque chose est ou non une forme de vie, il faut une conception efficiente de ce qu'est la vie - au sens purement biologique. Les premiers scientifiques pensaient communément que la vie était une animation de la matière ordinaire par un « truc » spécial appelé l'élan vital. Picard adhère à cette idée, à en juger par « Datalore » (NG), un épisode où l'Entreprise-D est menacée par une entité cristalline dont on sait peu de chose, sinon qu'elle a ravagé la planète maternelle de Data. Picard note dans le journal de bord que cette entité « se nourrit insatiablement de la force vitale, qu'on trouve dans les formes de vie ». Toutefois, la science moderne considère la vie comme l'organisation d'une matière privée d'élan, résultant d'une succession d'événements viables. Le Dr Crusher semble y concourir: « La vie est ce qui permet aux plantes et aux animaux de consommer de la nourriture, d'en tirer leur énergie, de grandir, de s'adapter à leur environnement, et de se reproduire. » Dans « Le sol natal » (NG), l'Entreprise-D visite

une planète où une autre substance cristalline manifeste un comportement étrangement proche de la vie. Le Dr Crusher suggère que la maturation, le développement et la reproduction peuvent fonder toute définition de la vie, « organique ou inorganique ». Par la suite, Data fait remarquer à propos de la substance cristalline que « seule la vie peut se reproduire, docteur. Elle est vivante ! »

Dans cet épisode, nous avons donc deux revendications explicites d'un rapport entre l'activité de reproduction et la vie. Selon le Dr Crusher, la capacité de se reproduire est nécessaire à la vie, et selon Data elle suffit à l'établir. Mais ces deux propositions sont incorrectes. Tout d'abord, il y a sur terre bien des animaux hybrides incapables de se reproduire et qui n'en sont pas moins vivants. Et de simples programmes informatiques peuvent se reproduire, mais ne sauraient nous paraître vivants pour autant.

La tentative du Dr Crusher dans « La qualité de la vie » est plus prometteuse, puisqu'il s'agit de ce que les philosophes appellent une définition « ostensive ». Au lieu de tenter d'énoncer les conditions nécessaires et suffisantes à la vie, Crusher souligne certaines caractéristiques des plantes et des animaux et affirme que la vie est ce qui sous-tend ces caractéristiques. Corrects ou non, les deux genres de définitions avancées par le Dr Crusher laissent ouverte la possibilité que la vie puisse être produite artificiellement. Les deux soulèvent donc la question de savoir si des entités comme les exocomps dans « La qualité de la vie » - ces petites machines conçues à diverses fins technologiques, et capables d'altérer leurs propres circuits - valent comme formes de vie. Très ironiquement, le grand défenseur des exocomps est Data, lui-même un artifice, une entité conçue et manufacturée. Mais après tout, pourquoi irait-on se demander si les exocomps peuvent être à juste titre considérés comme des êtres vivants ? La raison en est, ou du moins Star Trek nous invite à le croire, que le fait que quelque chose soit vivant ou non fait une différence dans la façon dont nous devons le traiter.

Dans « La qualité de la vie », Will Riker doit décider s'il va confier aux exocomps une mission où ils seront détruits, pour sauver les vies de Jean-Luc Picard et Geordi La Forge, l'ingénieur en chef de l'Enterprise-D. Si les exocomps sont de simples machines, pas de problème. Mais si ce sont des êtres vivants, ce serait un tort de sacrifier leur vie pour celle des deux êtres humains. C'est le point de vue, asséné par Data et cité en exergue de cette section, qui semble faire partie du message de l'épisode.

L'équipage de l'Enterprise-D semble en effet avoir adopté ce que les philosophes appellent une morale vitaliste. Selon celle-ci, les entités de l'univers à prendre moralement en considération sont, dans leur seule intégralité, les entités vivantes. Qui plus est, toutes les formes de vie sont d'importance morale égale, et il serait mal de sacrifier l'une pour sauver l'autre. Voici qui résonne comme l'expression populaire « La vie, c'est sacré », confirmée par l'expression biblique « Tu ne tueras pas », mais je suggère qu'il ne s'agit pas de la même chose, et que ni l'une ni l'autre ne sont des morales très plausibles. Si on prend les restrictions d'une morale vitaliste de façon absolument littérale, on aurait tort de déraciner et de cuire une plante pour la

manger, de tuer un moustique dans sa chambre à coucher, de dévier son véhicule pour éviter de percuter un enfant si c'est pour tuer un poulet, etc. (encore que ce genre de morale autorise sans doute à tuer d'autres êtres vivants si c'est pour se défendre).

Sans aucun doute, cette prise en considération morale va au-delà des humanoïdes. Des animaux tels que les chiens et les loirs sont dignes de cette attention ; toutes choses égales par ailleurs, on ne devrait pas les torturer, par exemple. Mais la morale vitaliste n'est pas convaincante. On a du mal à voir ce qui est si moralement important dans cette délimitation entre vie et non-vie, et toutes les formes de vie ne semblent pas d'une même importance morale. Pourquoi faire tout ce tintouin dans « La qualité de la vie » ? Pourquoi Data est-il prêt à sacrifier les vies de Picard et LaForge pour des machines qui ne sont qu'éventuellement vivantes ? Peut-être la réponse est-elle que cette catégorie est la troisième de celles mentionnées au début de cette section - celle de la vie intelligente. Observons un moment la vie terrestre : d'ordinaire nous faisons une distinction morale entre la vie intelligente et la vie non intelligente. D'un côté chiens et loirs, de l'autre aulnes et orchidées. Une raison en est que les créatures intelligentes sur terre sont capables de souffrir, et que, si une créature est capable de souffrir, il est condamnable de lui infliger une souffrance gratuite.

Les philosophes nomment psychocentrique la morale qui affirme d'abord que seule la vie intelligente, dans son ensemble, est à prendre moralement en considération, ensuite que toutes les créatures intelligentes sont d'importance morale égale. La citation de Picard placée en exergue de cette section semble relever exactement de cette opinion. Mais si la première revendication du psychocentrisme est viable, ce n'est pas le cas pour la seconde. Un être humain parvenu à maturité est d'une importance morale supérieure à un poulet ordinaire, même si tous deux sont intelligents et dignes de considération morale. Pourquoi ? Sommairement, parce qu'un être humain parvenu à maturité a une attitude mentale orientée vers l'avenir : il a des idées sur son avenir, et des préférences sur ce qu'il doit être - préférences qui peuvent être frustrées de diverses façons. Il a ces préférences parce qu'il se voit comme une entité qui perdure dans le temps. S'il est mal, toutes choses égales par ailleurs, d'infliger une grave douleur à un poulet, il est mal, toutes choses égales par ailleurs, de frustrer sévèrement les préférences d'un être humain. Mais si le poulet a tout intérêt à éviter la douleur et la souffrance, il n'a aucune préférence d'avenir susceptible d'être frustrée.

Relions ce point à notre discussion sur l'humanité : un être humain parvenu à maturité est une personne, qu'on doit prendre en considération moralement en ce qu'il a une idée de lui-même et de ses préférences pour l'avenir. Une réflexion rapide révèle qu'il ne s'agit pas d'une coïncidence. C'est pourquoi les philosophes considèrent qu'une personne est un individu doué d'une idée de soi comme entité perdurant dans le temps. Les personnes sont plus dignes d'attention moralement que les « nonpersonnes » intelligentes, comme les poulets justement, parce qu'ils ont des préférences plus étendues et plus profondes. Certains philosophes ont préconisé une morale centrée sur la personne, suivant laquelle seules les personnes, dans leur

ensemble, sont à prendre en considération moralement, et toutes les personnes sont d'une importance morale égale. Je ne suis pas plus d'accord avec cette morale qu'avec les précédentes. Certes toutes les personnes sont d'une importance morale égale, mais la morale centrée sur la personne implique qu'infliger une souffrance extrême et gratuite à une non-personne, un poulet par exemple, ne fait aucun tort au poulet. Selon la morale centrée sur la personne, si cet acte est condamnable, c'est qu'il fait du tort aux personnes - en troublant d'autres personnes, ou en traumatisant le bourreau. J'avancerais plutôt une morale où seules les entités intelligentes dans leur ensemble seraient dignes de considération morale et où seul l'ensemble des personnes serait d'importance morale égale. C'est là une extension de la morale centrée sur la personne : une partie de la considération morale due aux personnes s'étend aux non-personnes.

Cette extension reflète après tout avec le message de l'épisode « La qualité de la vie ». Data et LaForge se livrent à une simulation expérimentale où un exocomp est menacé de destruction; ils espèrent que celui-ci détectera le danger et tentera de l'éviter. Mais ils testent bien plus que la simple intelligence -ils essaient de voir si l'exocomp raisonne sur l'évolution de sa «vie». Ce genre de comportement en situation réelle avait, à l'origine, convaincu Data que les exocomps étaient vivants: un exocomp s'était court-circuité au lieu d'obéir à un ordre qui aurait entraîné sa propre destruction. Il s'avère que Data a raison: non seulement l'exocomp réussit le test, mais il le perce à jour. Lorsque les exocomps réalisent le danger couru par Picard et LaForge, tous trois montent une expédition de rescousse collective, et l'un d'eux est détruit au cours de l'opération.

La morale que j'ai dégagée est relativement conservatrice - au sens où les êtres humains sont au centre - mais pas totalement. Les êtres humains ne jouent pas un rôle essentiel dans une morale centrée sur la personne, puisque la question de savoir si les êtres humains parvenus à maturité sont des personnes est tout à fait contingente. Si notre espèce avait évolué différemment, nous aurions pu ne pas appartenir au cercle des individus dont l'importance relève de leur valeur morale. Certains critiques affirmeront qu'une morale centrée sur la personne est trop conservatrice, néanmoins. Ils avancent une opinion radicale qui polarise la considération morale non pas sur les individus, mais sur de larges systèmes écologiques comme la biosphère terrestre - la Terre elle-même et son atmosphère. Parmi les propriétés remarquables des grands écosystèmes, si l'on en croit cette vision « écologique en profondeur », figure la diversité biologique. Mais l'écologie des profondeurs connaît d'importantes difficultés, notamment quand il s'agit de définir la diversité. Si la diversité se définit par le nombre des espèces, un système écologique qui ne comporterait que des scarabées, mais par millions d'espèces différentes, serait plus varié qu'une forêt contenant un demi-million d'espèces. De plus, même si tout le monde peut donner une définition valable de la diversité, pour sauvegarder la vision de l'écologie des profondeurs il faut expliquer ce que la diversité a de si particulier, hormis le bénéfice qu'elle offre aux individus intelligents et sensibles. Ce n'est pas si facile.

Le fameux credo vulcanien IDIC («une infinie diversité d'infinies combinaisons») n'est jamais explicité dans le canon de Star Trek, mais, quel que soit son contenu, il semble compatible avec notre extension d'une morale centrée sur la personne. On peut convenir à un premier degré que la diversité est bonne en soi, qu'elle bénéficie aux entités moralement valides - à savoir les individus intelligents, et particulièrement les personnes. (Les bénéfices de la diversité pourraient apparaître dans un grand nombre de domaines - médical, scientifique, récréatif, esthétique.) Je supposerai qu'il n'y a du moins aucune objection valable à une morale centrée sur la personne, comprise extensivement, et relativement conservatrice - et je fonderai la suite de mon discours sur ce postulat.

Appliquer cette morale étendue à une forme non humanoïde de vie revient à déterminer à laquelle des trois catégories appartient un individu vivant, ou une communauté vivante. (Chose curieuse, la Fédération semble posséder un système de classification pour la vie intelligente, mais j'ignore comment il fonctionne. Les Shéliaks dans « Les enseignes de commandement » sont décrits comme « une forme de vie intelligente et non humanoïde, de classification R3 ».) Les équipages de la Flotte interstellaire doivent se demander si le candidat est intelligent ou non. Si un individu ou une communauté relèvent de la vie non intelligente, alors, suivant la morale centrée sur la personne, ils ne sont pas à prendre en considération morale. Sans doute les nouvelles plantes découvertes par les exobotanistes relèvent elles de cette catégorie. Notez bien que cela n'autorise pas les équipages à manipuler ce genre de vie à leur gré - mais suppose que, quoi qu'ils fassent, ils ne feront pas de tort à la plante. Manifestement, la transplantation ou toute autre forme d'intervention sur la végétation d'une planète peut avoir des effets sérieux sur la vie intelligente et les personnes qui s'y trouvent, et pourrait bien être condamnable d'un point de vue moral. Appelons les candidats à la vie non intelligente des suspects numéro trois (les définitions de la vie par le Dr Crusher semblent tout à fait utiles pour identifier les suspects numéro trois).

Si un individu est intelligent sans être une personne, il est à prendre en considération moralement. Supposons que nous rencontrions un extraterrestre dont l'intelligence et les fonctions sensorielles équivalent à peu près à celles d'un poulet. Si cette créature est susceptible d'éprouver une douleur physique, il serait condamnable de lui infliger cette douleur juste pour s'amuser. Mais il ne serait pas condamnable de la tuer pour la manger, toutes choses égales par ailleurs (disons, dans la mesure où on peut le tuer sans lui infliger une souffrance gratuite, à lui ou à une autre personne, par exemple son propriétaire). Appelons suspects numéro deux tous ceux qui sont intelligents sans être une personne.

Si un individu est une personne, il est à prendre en considération d'un point de vue moral. Lorsque l'Entreprise rencontre d'autres personnes non humaines, si exotiques soient-elles, les actions de l'équipage sont guidées par toutes les considérations morales concernant la gestion d'autres humains. Cela inclut la mise en pratique de la règle d'or. Dès lors l'Entreprise doit être à l'affût de suspects numéro un - d'individus pouvant prétendre à la qualité de personne et à la considération

morale qui l'accompagne. Mais comment identifier les suspects numéro un, quand ils peuvent être si exotiques? Je suggère que nous commençons par noter que lesdits suspects numéro un englobent les suspects numéro deux. Qu'est-ce qui autorisera Kirk, Picard et les autres à conjecturer qu'ils ont affaire à une forme de vie intelligente ?

Suspects numéro deux

*Commandant, prouvez à la Cour que je suis,
moi, un être sensible !*

Picard au commandant Bruce Maddox, «Ce qui fait l'homme» (NG)

Qu'est-ce qui fait d'une entité un suspect numéro deux, et comment étudier plus à fond ces indices ? Kirk et Spock se retrouvent confrontés à ce problème dans « Démon dans le noir» (PS). L'opération minière sur Janus IV est menacée lorsque cinquante mineurs sont tués par ce qui semble être un gros rocher. Spock soupçonne que le roc est une créature à base de silicone - une forme de vie intelligente, Jim, mais pas telle que nous la connaissons. Dans ce genre de situations, il serait pratique d'avoir un test d'intelligence sous la main. Non pas une version extraterrestre du test de QI; ceux-ci sont administrés aux sujets présumés intelligents. J'entends plutôt un test pour déceler la présence de l'intelligence, d'une pensée. .,

Kirk, Picard et les autres ont un avantage injuste en ce domaine puisqu'ils peuvent recourir aux capacités psychiques de Spock, de la conseillère Deanna Troi dans la série La nouvelle génération, et de Tuvok (voire de Kes) dans celle du Voyageur. Dans chacun de ces cas, on trouve une forme de perception extrasensorielle. La version empathique de la perception extrasensorielle de Troi se borne à discerner l'état émotionnel d'un sujet et semble opérer tout le temps, alors que les Vulcariens Spock et Tuvok peuvent susciter une « confusion des cerveaux» bien plus complète en réalisant le bon contact physique. (Bien sûr, ces capacités posent parfois problème aux scénaristes : dans «Le silence règne» [NG], l'intrigue exige que l'équipage ne soit pas conscient de la présence d'un individu intelligent qui les manipule. Les scénaristes éludent le problème en faisant que Troi échoue à reconnaître « une intelligence si vaste qu'elle [lui] a échappé ! ») Certains membres de la Flotte interstellaire ont des dons de perception extrasensorielle encore plus développés. Dans « Tin Man» (NG), nous apprenons que l'expert principal de la Fédération en matière de contact avec les extraterrestres est un bétazoïde exceptionnel, doué de vastes pouvoirs de télépathie. (On notera que tous ces membres psychiques de la Flotte sont des extraterrestres: nous n'avons aucune preuve que les humains terrestres puissent avoir des capacités de perception extrasensorielle, même s'ils le prétendent souvent.) Dans « Démon dans le noir», Spock utilise sa perception extrasensorielle comme une forme de test d'intelligence, en accomplissant une jonction cérébrale avec l'entité rocheuse. Il découvre que la créature est une Horta, la mère de milliers de rejetons rocheux en voie d'être détruits par l'opération de

minage, et qu'elle agonise suite aux blessures causées par les phaseurs de l'équipage. Le Dr McCoy fait un peu de maçonnerie pour réparer les blessures, un traité est négocié entre la Horta et les mineurs, et tout finit bien.

Un autre avantage dont jouissent les équipages de Star Trek, c'est l'existence de senseurs détectant la vie. Dans moult épisodes, un officier de passerelle « fera une recherche pour scanner des formes de vie » sur une planète située en dessous ou sur une station spatiale apparemment abandonnée. Ces scanners sont aussi utilisés par les équipes de mission à la surface des planètes. Dans « L'arsenal de la liberté » (NG), Data affirme avoir cherché en vain des formes de vie sur la planète Minos à l'aide d'un scanner. Ce dernier exemple ressemble plutôt à une facilité du scénario, mais même si un senseur pouvant détecter l'intelligence était possible, pourquoi l'équipage accepterait-il la réponse négative (d'autant que la Flotte interstellaire a une mission de découverte)? Alors qu'un diagnostic positif pourrait tirer qu'il y a de la vie intelligente en dessous, parce que c'est rme reconnaissable (quelle qu'elle soit), un diagnostic négatif ne garantit sûrement pas qu'il n'y en ait pas, car elle pourrait prendre une forme non reconnaissable.

Mais comment, en l'absence de perception extrasensorielle et d'autres senseurs particuliers, Kirk pourrait-il détecter une intelligence chez la Horta? Vous serez peut-être surpris d'apprendre que cela relève d'un problème philosophique très ancien appelé le problème des autres esprits. Prenez tout simplement le vôtre. Vous avez des états d'esprits - des croyances, des souvenirs, des espoirs, des sensations. Mais comment savez-vous quand vous êtes dans tel ou tel état d'esprit, comme la douleur? La réponse la plus plausible est que vous le savez par introspection (littéralement, en regardant à l'intérieur de vous-même). Mais si c'est vrai, cela pose un problème évident. En l'absence de perception extrasensorielle, on n'a pas d'accès introspectif aux autres esprits (contrairement à Spock, qui ressent bel et bien la douleur de la Horta pendant la jonction des cerveaux). Sans cet accès introspectif, comment savez-vous qu'il existe véritablement d'autres esprits? Comment pouvez-vous affirmer que moi, Richard Hanley, j'ai un cerveau? Cela pourrait passer pour une question idiote. Lorsque Picard la pose au commandant Maddox dans « Ce qui fait l'homme » (voir l'exergue), celui-ci répond: « C'est absurde! Nous savons tous que vous êtes un être sensible! » Toutefois, à moins de donner une réponse satisfaisante à cette question « absurde », vous êtes condamné à choisir entre deux opinions extrêmes. L'une est le solipsisme, l'opinion que vous seul êtes intelligent dans l'univers, et l'autre est le panpsychisme, l'opinion qu'absolument tout est intelligent dans l'univers. Si vous souhaitez adopter l'opinion raisonnablement moyenne selon laquelle certaines choses dans l'univers, à part vous, sont intelligentes, mais pas toutes, il faut faire appel à un test d'intelligence.

Manifestement, le commandant Maddox pense savoir que Picard est intelligent, puisqu'il cite l'intelligence comme une condition de la « sensibilité » (ce qui fait la personne). Nous nous pencherons plus longuement sur la position de Maddox dans le chapitre 3, mais ce qui importe pour l'instant, c'est l'importance qu'il semble accorder à l'humanité, au sens d'être humanoïde. Son raisonnement pourrait être le

suivant: Je suis intelligent et je suis humanoïde, Picard est humanoïde, dès lors il est (probablement) intelligent lui aussi.

Si c'est là le raisonnement de Maddox, certains philosophes y ont acquiescé. On peut recourir à l'analogie pour argumenter l'existence des autres esprits humanoïdes. Vous savez que vous, vous avez un esprit: par analogie, les autres humanoïdes en ont sans doute un - de même, si votre voiture est d'un modèle et d'une marque économiques, on peut en inférer qu'une autre voiture de la même marque et du même modèle l'est également. Mais je ne crois pas que cet argument démontre quoi que ce soit. La conclusion, dans le cas de la voiture, s'appuie sur des données connues - dont l'uniformité des procédures de production pour les véhicules. Quelles données préalables viendraient appuyer l'analogie avec les autres esprits ? Celles qui associeraient le fait d'être humanoïde au fait d'être intelligent. Mais aucune de ces données n'est connue à l'avance - puisque le raisonnement est censé établir l'existence des autres esprits humanoïdes! Pour l'illustrer, disons que, si l'on ne possède aucune donnée préalable sur la production de véhicules, en conclure que l'autre voiture est aussi économique que la vôtre n'a pas plus de sens que de conclure qu'elles sont de la même couleur.

Seuls des tests de comportement sont plausibles en matière de test d'intelligence, et les membres de la Flotte interstellaire y ont fréquemment recours - ce qui revient au même, puisque, même si l'argument analogique valait pour les autres humanoïdes, il ne saurait concerner les non-humoïdes. C'est le comportement des exocomps qui a amené Data à conclure qu'ils étaient intelligents. Et dans « Le sol natal », lorsqu'il est attaqué par une foreuse à laser, il en déduit que le comportement de la machine est trop « dynamique » pour être accidentel, et qu'il y avait, en d'autres termes, « un cerveau qui s'exerçait contre [lui] ». Qui plus est, bien avant que Spock n'accomplisse sa jonction cérébrale avec la Horta, il soupçonne qu'elle est intelligente à cause de son comportement (pourquoi, sinon, tenterait-il une jonction cérébrale avec un rocher?). Kirk pratique également des tests comportementaux sur la Horta, en guettant ses réactions lorsqu'il brandit, puis éloigne son phaseur. Revenons sur terre : c'est le comportement des animaux - chiens, poulets, etc. - qui nous fait leur attribuer une intelligence (relativement rudimentaire). Tout comme la Horta recule chaque fois que Kirk brandit son phaseur (parce qu'elle a déjà été blessée), des animaux comme les chiens apprennent très vite à fuir les agressions répétées d'individus humains.

Quelles sont les entités qui échouent au test comportemental d'intelligence? Sur terre, les plantes sont tenues pour une forme de vie non intelligente (par la plupart des gens, du moins) et cela vaut pour le cosmos de Star Trek, pour autant que je sache. Parmi les pratiquants de l'exobotanique (le Dr Crusher, dans « Indices » [NG]), personne ne s'élève contre cette vue. Serait-ce là un préjugé, une incapacité à percevoir les possibilités d'une intelligence alternative? Je ne crois pas. Ici sur terre, nous pouvons discerner deux stratégies fondamentales adoptées par des choses vivantes, représentées en quantité suffisamment vastes, et en compétition les unes avec les autres. L'une est une stratégie de guerre des tranchées, l'autre une

stratégie de guérilla (pardonnez-moi ces métaphores militaristes). Les plantes ont adopté la stratégie de tranchée en s'enfouissant dans le sol. Plus l'individu retranché est grand - un arbre, par exemple - plus il dépend d'un ravitaillement stable en eau et en nourriture, et d'un blindage lourd pour sa protection. Les guérillas, de l'autre côté, sont mobiles, et plus le guerrier individuel est gros - un chat, par exemple - plus sa survie repose sur la flexibilité de son comportement et sa capacité d'apprendre. Ces créatures vivent très précisément grâce à leur esprit. Un comportement flexible et adaptable est donc un indicateur très sûr d'intelligence.

Mais quid des choses exotiques vraiment dangereuses qui surgissent dans Star Trek ? Dans « La machine du jugement dernier », l'Entreprise rencontre un consommateur artificiel de planètes entières, et dans « Obsession » (PS), une entité gazeuse non artificielle qui draine tous les globules rouges de ses victimes. Personne sur l'Entreprise ne semble vraiment inquiet de savoir si ces entités sont intelligentes, et les deux épisodes s'achèvent sur leur destruction délibérée. Cependant, les choses sont un peu différentes dans les épisodes de la Nouvelle génération intitulés « Datalore » et « Avatar de silicone » où l'Entreprise-D rencontre l'entité cristalline qui ravage des planètes entières. Bien que cet épisode s'achève par la destruction de l'entité cristalline, c'est un ci vil du sexe féminin qui l'accomplit contre son gré, et Picard désapprouve fortement cet acte. Pourquoi est-il si bouleversé, et non simplement soulagé, comme Kirk et les autres l'étaient dans la série originale ? La réponse probable est que, dans ce cas, l'équipage vient d'établir une communication linguistique avec l'entité.

À coup sûr, l'un des comportements les plus impressionnants qu'une entité puisse manifester est la communication linguistique, et cette capacité apparaît comme un signe évident d'intelligence, voire davantage. Cette capacité linguistique était absente, ou n'était pas avérée dans les autres cas que nous avons considérés jusqu'ici, et je les ai choisis avec soin pour cette raison (si l'on excepte une légère aberration - la Horta réussit d'une façon ou d'une autre à graver « Pas tuer moi » sur une paroi rocheuse avant la jonction cérébrale, et sans le moindre « traducteur universel »). La majorité des formes de vie intelligentes rencontrées par la Flotte interstellaire jouit de capacités linguistiques, et grâce au procédé baptisé « traducteur universel », l'équipage peut en général établir un contact linguistique immédiat. (Mais pas toujours. Dans « Les petits hommes verts » [PE9], trois Ferengi remontent le temps, se retrouvent téléportés à Roswell, dans le Nouveau-Mexique, en 1947, et leurs traducteurs universels implantés tombent en panne. Lorsqu'ils commencent à se frapper la tête pour remettre en route le traducteur, les Terriens y voient une tentative de communication, et les imitent aussitôt.)

La traduction universelle : ludique et utile

Zefram Cochrane: « Quelle est la théorie sous-jacente au procédé ? »

Kirk: « Certaines idées et concepts universels sont communs à toutes les formes de vie intelligente. Ce procédé compare

instantanément la fréquence des ondes cérébrales, sélectionne les schémas et les concepts identifiés, et supplée la grammaire nécessaire. »

Spock: « Puis il traduit tout simplement ses trouvailles dans notre langue. »

Conversation tirée de «Métamorphose» (PS)

À un certain niveau, l'explication de Kirk peut paraître ridicule. À un autre niveau, elle se rattache à une théorie populaire sur l'acquisition du langage humain et la psychologie innée qui l'autorise. Dans le champ relativement neuf de la psycholinguistique, la sagesse dit qu'il est des universaux cognitifs présents chez tous les Terriens, implantés dans nos cerveaux sous forme de grammaire universelle. Cette grammaire innée est considérée comme la meilleure explication de notre capacité, au début de notre vie, à apprendre les langages naturels. La première phrase de Kirk rappelle étonnamment cette hypothèse, étendue à toute forme de vie intelligente et partout ! (Encore que l'exemple avancé par Kirk soit tendancieux. Il dit: « L'idée du mâle et de la femelle est un universel », mais les J'naï sont androgynes dans « Les proscrits » [NG].)

Le traducteur universel est parfois utilisé pour résoudre un problème qui surgit dans le canon de Star Trek. Comme le remarque Lawrence Krauss dans *La physique de Star Trek*: « Toutes les espèces étrangères rencontrées par l'Entreprise sont d'apparence humanoïdes, et elles parlent toutes anglais ! » J'ai lu sur Internet une explication sur l'ubiquité de l'anglais dans le cosmos de Star Trek : nous lisons et entendons les choses du point de vue de l'équipage, et puisque celui-ci dispose du traducteur universel, il entend les discours extraterrestres instantanément traduits en anglais. Si bien que nous aussi, nous entendons les extraterrestres parler anglais, même si ce n'est pas le langage qu'ils sont en train de parler. Si nous acceptons cette explication, alors l'idée du traducteur universel facilite l'enchaînement dramatique (tout comme l'idée du téléporteur, qui supprime la nécessité de faire atterrir le vaisseau). Mais cette idée me paraît fautive. Prenez n'importe quel film d'Hollywood situé dans un pays non anglophone : *Guerre et paix*, *Les liaisons dangereuses*, *L'insoutenable légèreté de l'être*, etc. Aucune de ces histoires ne stipule que les personnages parlent anglais, et dans la plupart d'entre elles ce n'est pas le cas ! Les acteurs, eux, le font évidemment, sans quoi il est probable que nous ne les comprendrions pas, mais c'est simplement le véhicule utilisé par le film pour rendre les propos russes, français, tchèques, etc. des personnages. Qui plus est, si nous percevons parfois les choses du point de vue de l'équipage (lorsque nous contemplons l'écran de passerelle), ce n'est pas toujours le cas. Dans « *Premier contact* » (NG), Riker est échoué sur la planète Malcor III où il essaie de se faire passer pour un autochtone. Les Malcoriens parlent-ils anglais, pour quelque raison insondable ? Non point. Les entendons-nous parler anglais parce que nous adoptons le point de vue de Riker ? Non point. Nous les entendons parler anglais entre eux. En vérité, dans « *Premier contact* » tout le monde parle malcorien sur Malcor III, y compris Riker. S'il

utilisait le traducteur universel, il se trahirait, puisque je suppose que celui-ci ne modifie pas les mouvements des lèvres. (Riker doit avoir un don des langues remarquable: dans « Qui garde les gardiens », Deanna Troi et lui infiltrèrent les Mintakiens de l'âge de bronze quasi sans préavis; tous deux ont peut-être pris des leçons de mintakien pour parer au hasard, puisqu'ils ne peuvent faire usage des traducteurs universels.) Par la suite, une fois ce petit jeu achevé, Picard se dévoile aux Malcoriens. Comme il ne parle pas malcorien, c'est sans doute ici qu'intervient le traducteur universel.

Cela dit, il arrive qu'un scénario de Star Trek ne fasse sens que si l'on suppose que les extraterrestres parlent anglais. Dans un certain nombre d'épisodes de la série originale, la culture extraterrestre ressemble étonnamment à la culture terrienne, et le groupe en mission, bien pénétré de la règle d'or, doit garder l'incognito. Ce scénario se produit dans « Miri » (évoquant plus ou moins la terre), « Lignes de force » (l'Allemagne nazie), « La gloire d'Oméga » (le pavillon américain, la Constitution et la menace communiste), « Du pain et des jeux » (le xx- siècle, mais en supposant que l'Empire romain n'ait jamais connu sa chute) et « Le syndrome du paradis » (les Indiens). Dans « Du pain et des jeux », les auteurs de Star Trek notent explicitement que les indigènes parlent anglais, « l'anglais familier du xxe siècle », qui plus est. Kirk fournit une explication intéressante, quoique douteuse : « Ces créatures appartiennent à une planète hautement industrialisée de type xx siècle, très semblable à la Terre - exemple surprenant de la loi de Hodgkin sur le développement des planètes parallèles ! » (Mais cette maigre explication ne saurait toujours valoir; dans « Lignes de force », John Gill, l'anthropologue de la Fédération, finira par attribuer l'étonnante ressemblance avec l'Allemagne nazie à un phénomène de contamination culturelle, contrevenant à la règle d'or. Mais Kirk et Spock ne peuvent utiliser le traducteur universel dans leur travail d'espionnage. (Gill inclut-il la langue anglaise dans les critères de pureté raciale? Je me le demande.)

Le temps darmok

Le fait que les races extraterrestres communiquent entre elles est tout à fait prodigieux.

Troi à Picard dans « Les enseignes du commandement » (NG)

Une dernière remarque portera sur ce que les philosophes appellent le problème de la traduction radicale. Nous nous demanderons jusqu'à quel point un extraterrestre doit différer de nous pour cesser d'être digne de considération morale, et nous nous poserons la même question sur la communication : jusqu'à quel point un langage extraterrestre doit-il différer du nôtre pour cesser d'être traduisible? Kirk affirme dans « Métamorphose » (PS) qu'il y a des universaux cognitifs, mais s'il avait tort? Et si vous rencontriez un être parlant dont la pensée soit si étrangère que vous n'auriez aucun moyen de traduire ses propos ? À quoi ressemblerait un langage si radicalement différent? Ce problème est soulevé par

Deanna Troi pour la première fois dans Star Trek dans « Les enseignes du commandement », lorsqu'il est nécessaire de communiquer avec une race étrange appelée les Shéliaks. Suit cet échange de propos :

Troi : « Dans nos échanges avec les autres races non humanoïdes, il y a toujours eu un point de référence. Pas avec les Shéliaks. »

Picard : « Mais nous devons avoir quelque chose en commun - nous communiquons. »

Troi : « À peine. Eux ont appris plusieurs langages de la Fédération, mais le leur nous échappe encore. »

Troi demande à Picard d'imaginer la situation suivante : « Nous sommes échoués sur une planète. Nous n'avons aucun langage en commun, mais je veux t'apprendre le mien. » Troi lève une tasse en verre, qui pourrait contenir du thé ou du café, prononce un mot imaginaire, et demande : « Qu'est-ce que je viens de dire ? » Picard avance « tasse », puis « verre ». Troi ajoute les possibilités « liquide », « clair », « brun » et « chaud ». (Elle ne propose pas « thé » ou « café », mais pourrait le faire.) « Nous, nous conceptualisons l'univers de façon plus ou moins semblable », dit Troi, suggérant qu'en l'absence d'un schème conceptuel commun la communication est extrêmement difficile. Elle conclut ses conseils en faisant remarquer que le traité de la Fédération avec les Shéliaks comporte 500 000 mots, et que 372 légistes de la Fédération y ont travaillé ! « La longueur visait à rassurer les Shéliaks, dit-elle à Picard. Ils trouvent notre langage irrationnel, et ont exigé ce niveau de complexité pour éviter tout malentendu futur. »

Chose curieuse, la philosophie propose un exemple semblable à celui de Troi, mais qui sert à démontrer quasiment la thèse opposée. Le philosophe W. V. O. Quine a imaginé le même genre de tentative pour traduire un langage inconnu, mais ici sur terre. Un lapin passe en courant devant un autre lapin, et celui-ci crie : « Gavagai ! » Ce son signifie-t-il « un échantillon individuel de la lapinitude » (un membre d'une classe), « une tranche de lapin » (tranche temporelle d'un lapin singulier), ou « un lapin géant discontinu » (l'individu serait la somme de tous les lapins individuels - comme on pourrait appeler le New Jersey « les États-Unis »), ou « fragment de lapin non détaché », ou quoi encore ? Certains philosophes concluent de ces exemples que la traduction reste toujours indéterminée, même entre des gens parlant le « même » langage, et qu'un « même » langage n'existe donc pas. Ils supposent qu'on pourrait traduire au mieux les énoncés de l'autre sans toujours savoir si on partage son schème conceptuel - hypothèse qui mine l'idée des universaux cognitifs.

Les différences entre la Fédération et les Shéliaks semblent exagérées. Tout d'abord, on a du mal à voir comment deux groupes sans le moindre concept commun pourraient conclure un traité. Puis les Shéliaks ont, somme toute, appris les langages de la Fédération - c'est peut-être qu'ils sont plus malins que les humains, et travaillent avec un schème conceptuel analogue, mais plus complexe. Enfin il paraît difficile de surmonter de gros problèmes de traduction en intensifiant la complexité

d'un traité!

Dans un cadre plus crédible, la situation imaginée par Troi devient réalité dans « Darmok » (NG), lorsque l'équipage rencontre les Enfants de Tama, une race extraterrestre dont les tentatives de communication sont « incompréhensibles ». Picard est kidnappé par eux et déposé sur une planète solitaire avec un Tamarien, Dathon, qui brandit des armes et énonce des phrases comme (traduites universellement, bien sûr): « Darmok et Jalad à Tenagra », et « Shaka, quand churent les murailles ». Le couple est attaqué par une créature constituée d'énergie pure, et Dathon est grièvement blessée. Picard finit par comprendre que les Enfants de Tama communiquent au moyen d'images et de métaphores, et réussit ainsi à mener une conversation rudimentaire avec le Tamarien à l'agonie. Pendant ce temps, de retour à bord de l'Entreprise-D, Data et Troi ont cherché des enregistrements de tentatives tamariennes de communication. Lorsque Riker demande ce qu'ils ont découvert, voici la conversation qui suit:

Data : « La structure de l'ego tamarien ne semble pas autoriser ce que nous appelons la conscience de soi. Leurs capacités d'abstraction sont des plus inusuelles. Ils semblent communiquer au moyen d'images narratives - de références à des individus et des lieux tirés de leurs récits mythohistoriques. »

Troi: « Comme si je vous disais: "Juliette au balcon." »

Dr Crusher: « Une image de romance amoureuse. »

Troi: « Exactement. L'imagerie est essentielle pour les Tamariens. Elle incarne leurs états émotionnels, leurs processus mentaux. C'est ainsi qu'ils pensent et qu'ils communiquent. »

Riker : « Si nous connaissons leur façon de penser, nous devrions pouvoir leur faire comprendre certaines choses. »

Data: « Non, monsieur. C'est un peu comme si nous comprenions la grammaire d'un langage sans en posséder le vocabulaire. »

Dr Crusher: « Si j'ignorais qui était Juliette, ou ce qu'elle faisait sur ce balcon, l'image en soi n'aurait aucun sens. »

Troi: « C'est juste. Nous savons que Darmok était un grand héros, un chasseur, et que Tenagra était une île, mais c'est tout. Sans les détails, nous ne saurions comprendre. »

Data: « Il nous faut absolument apprendre l'histoire où les Tamariens puisent leurs images. »

Cet extrait de dialogue est censé établir les zones de différence cognitive entre l'Entreprise-D et les Tamariens. Tel quel, il est tout à fait fascinant de par ses idées philosophiques. Tout d'abord, l'idée que les Tamariens n'ont aucune conscience de soi individuelle. Puis celle que les processus de pensée tamariens sont (peut-être exclusivement) imagés. Enfin l'idée que savoir la grammaire d'une langue ne suffit pas à sa compréhension.

La première de ces idées me pose plusieurs problèmes. D'une part, avec quel

degré d'exactitude Troi et Data ont-ils établi que les Tamariens n'avaient aucune conscience de soi? Puisqu'ils tentent d'essayer de comprendre la langue tamarienne, et, de leur propre aveu, n'y réussissent pas très bien, il semble tout à fait improbable qu'ils soient en mesure de formuler des jugements aussi subtils sur le schème conceptuel tamarien. (Je suppose que Troi n'a pas deviné ces informations par simple empathie.) Revenons sur terre : on a naguère affirmé pour des raisons de linguistique comparative que d'autres cultures avaient des schèmes conceptuels profondément différents de ceux de la culture occidentale. Ces assertions se sont révélées fausses, ou largement exagérées. (Accusation infamante, les Hopis étaient censés n'avoir aucun concept du temps, et les Eskimaux, d'innombrables concepts de la neige.) Les linguistes en question avaient réussi à communiquer de façon constructive avec les sujets de leur étude, et ne s'en trompaient pas moins. Data et Troi sont sur un sol fragile ! Un autre de mes soucis est que, si l'on prend les preuves avancées dans cet épisode, l'affirmation que les Tamariens n'ont aucune conscience de soi sonne faux. Les Tamariens ont une conscience individuelle et non collective, des noms individuels, et d'évidence ils ont des opinions et des préférences sur l'évolution de leur existence propre et de celle des autres. Picard attribue clairement une conscience de soi à Dathon quand il dit : « Le Tamarien était prêt à risquer nos vies à tous deux par simple espoir de communiquer ... Cet engagement signifiait plus pour lui que sa propre vie. » Qui plus est, lorsque Picard annonce la mort courageuse de Dathon à l'équipage tamarien, leur réaction solennelle montre clairement qu'ils y reconnaissent le trépas d'un individu. Puis le premier officier tamarien prononce une nouvelle allusion, « Picard et Dathon à El-Adrel », laquelle distingue deux individus (ce pour quoi il faut une notion du moi individuel). Il semble que ce soit uniquement par intérêt dramatique que les scénaristes nient aux Tamariens une conscience de soi. C'est ainsi que Dathon, dans ses tentatives de communiquer, peut tout juste jouer les Tarzan : « Moi, Dathon ; toi, Picard. »

Toute ma sympathie va aux scénaristes qui essaient d'atteindre deux buts diamétralement opposés. D'une part, ils tentent de présenter un point de vue radicalement différent sur le monde, de l'autre ils essaient de le mettre en scène! Bref, si le point de vue des Tamariens différait trop du nôtre, les auteurs de Star Trek ne pourraient pas nous le présenter. Trois autres objections à leur tentative: mis à part les « mois » individuels, les Tamariens ont manifestement des noms pour les systèmes planétaires, puisque Dathon a pu émettre « El-Adrel » en tamarien. Puis, même les philosophes qui doutent qu'il y ait des universaux cognitifs réels pensent que tout langage complexe et utile doit comprendre des coordinations logiques. L'unique coordonnant tamarien, c'est «et». Mais bien sûr, si on présentait leur langage tel qu'il doit être en réalité, cela gênerait l'histoire. Enfin il est clair que « Mira, toutes voiles déployées » est un ordre de décoller, de partir. Mais dans quelle direction, et à quelle vitesse? Les Tamariens utilisent un langage plus précis pour naviguer dans l'univers, construire des choses, etc. - mais, répétons-le, cela gênerait l'histoire.

En présentant les Tamariens comme des êtres dont la pensée est en premier

lieu ou exclusivement imaginative, les auteurs de Star Trek se sont engagés dans un débat philosophique animé. Presque tous conviennent que les contenus de certains états mentaux sont (proprement et exhaustivement) caractérisés par une proposition. Par exemple, je crois que New York est au nord de Philadelphie. La phrase « New York est au nord de Philadelphie » exprime une proposition indépendante d'un énoncé relevant d'une langue particulière, et c'est pourquoi on peut dire exactement la même chose dans une autre langue, à travers une phrase de même contenu - qui est aussi celui de ma croyance. Mais supposons qu'au lieu de le penser je me figure que New York est au nord de Philadelphie - j'élabore alors une image mentale. Le contenu de cet acte d'imagination surpasse celui de ma croyance. Mais s'agit-il simplement d'un contenu propositionnel plus complexe, ou d'un contenu non propositionnel - irréductiblement imagé? Les philosophes sont profondément divisés sur cette question, qui se rattache à celle de l'intelligence artificielle que nous discuterons aux chapitres 2 et 3. Si une part de la pensée est irréductiblement imagée, les ordinateurs, bien adaptés à traiter les contenus propositionnels, pourront-ils jamais y accéder?

Je suis intrigué par l'analogie suggérée par Data entre les difficultés de la communication avec les Tamariens et une situation où l'on «comprend la grammaire d'un langage sans connaître son vocabulaire». Ici, je crois que Data fait référence à la distinction établie par les théoriciens du langage entre la syntaxe et la sémantique d'un langage. Pour aller vite, la syntaxe renvoie au système de symboles fondant un langage, y compris les règles de transformation nécessaires pour manipuler ces symboles, et la sémantique renvoie à la signification de ces symboles - ce dont ils traitent. Pour le dire autrement, la syntaxe renvoie aux propriétés d'un langage non interprété, la sémantique aux propriétés d'un langage interprété. Dans une démarche qu'approuverait M. Spock, je vais faire usage de la logique pour expliquer cette différence (elle me servira également au chapitre 3), en introduisant un langage simple que j'appellerai L. Voici une expression en L: «Wb». Explication: « W » indique un prédicat, « b » un individu. «Wb» signifie donc que l'individu b est (a pour prédicat) W. Voici un autre énoncé en L: « (v'x) (Wx ~ Fx) ».Explication:« (v'x) » signifie « pour tout x » et « (Wx ~ Fx) » signifie «six est W, alors x est F. La logique veut que si l'on pose d'une part (v'x) (Wx ~ Fx) et d'autre part Wb, on peut conclure Fb. Cette règle devrait faire sens intuitivement: si un individu b est W et si tout ce qui est W est F, il est évident que l'individu b est F.

À moins d'être quelqu'un de peu ordinaire, le dernier demiparagraphe vous aura barbé comme il n'est pas permis, et c'est tout l'intérêt de la chose. Si je devais poursuivre dans la même veine (et que vous restiez éveillé), vous maîtriseriez L et ses règles en un rien de temps. Mais ce ne serait pas très intéressant, parce que L n'est pas (encore) connecté au monde. Vous ne feriez qu'apprendre une syntaxe, un système de symboles avec ses règles de transformation, sa grammaire. Rajoutons à présent une interprétation - une sémantique. Mettons que b renvoie à Bruce, que W dénote le prédicat d'être un wombat, et F celui d'avoir de la fourrure. Dans cette interprétation, «Wb» signifie que Bruce est un wombat, et « (v'x) (Wx ~ Fx) »

signifie que tout ce qui est un wombat a de la fourrure. Il en découle que « Fb » signifie que Bruce a de la fourrure. Nous avons rajouté un vocabulaire à la grammaire de L en ajoutant des significations fixes.

Comparons maintenant la situation de Data et Troi à celle de Picard, resté sur la planète. Picard a entendu si souvent « Darmok et Jalad à Tenagra » qu'il connaît la formule par cœur. Lorsque Dathon commence « Darmok et Jalad ... », Picard achève le fragment de phrase à sa place, à la grande joie de Dathon. Mais ils n'ont pas progressé. Comme le dit Picard: «J'ai retenu les mots, mais je ne comprends pas.» En ce qui concerne Picard, Dathon pourrait aussi bien dire «Wb». Mais Data et Troi ne savent-ils pas ce à quoi renvoient « Darmok » et « Tenagra »? Ne possèdent-ils pas déjà le vocabulaire qu'ils cherchent-du moins lorsqu'ils résolvent« Jalad »? Non. Même s'ils savent à quoi renvoient les termes individuels, ils ignorent ce à quoi renvoie le fragment dans son ensemble, et ont donc toujours un problème de vocabulaire. Data suggère qu'il n'y a qu'une façon de le résoudre - retrouver l'histoire -, mais il a omis un élément. Une allusion comme celle-ci met en relation une histoire et une situation, rapport qui désigne une analogie entre les deux. C'est comme si le locuteur disait : « Cette situation évoque Juliette au balcon. » Comme le rapport analogique est réversible, nous le retournons et disons : « Juliette au balcon évoque cette situation. » Pour comprendre l'allusion, il faut comprendre que les deux situations sont du même genre, et que, si l'allusion est formulée dans une situation adaptée, on peut apprendre quelque chose sur l'histoire de Juliette!

Nous pouvons intégrer la dernière pièce du puzzle par analogie. Voyez votre propre situation si vous savez que «Wb» signifie que « Bruce est un wombat ». Si vous ignorez ce qu'est un wombat, vous n'êtes pas mieux renseigné. Comment pouvez-vous apprendre ce qu'est un wombat ? Voici un excellent moyen : demandez à un expert de vous montrer un wombat, sachant que l'expert est juste quelqu'un qui sait identifier un wombat et utiliser correctement le terme. Vous êtes en compagnie de l'expert, un wombat surgit et il dit : «Wombat». Vous concluez: «C'est un wombat», et maintenant, c'est vous l'expert ! Dans « Darmok », que fait Dathon? Il emmène Picard voir ce que c'est qu'un Darmok et Jalad à Tenagra, voilà ce qu'il fait. Comme c'est lui l'expert, il met Picard et lui-même dans une situation analogue (et dangereuse) et prononce le nom qui définit cette situation. C'est une dure leçon pour les deux, mais Picard apprend quelque chose.

« Darmok » est un épisode extrêmement bien ficelé, mais nous devons conclure que les Enfants de Tamara ne sont pas, après tout, si différents de nous, ni leur langage du nôtre. À moins d'avoir déjà un manuel de traduction, impossible d'apprendre un langage inconnu sans mettre la main à la pâte, pour ainsi dire. Pour comprendre entièrement ce à quoi mots et énoncés font référence, il faut être dans une situation référencée. Mais lorsque vous y êtes, comment savez-vous à quoi on fait référence ? Il semble que Troi ait raison : une communication réussie requiert un schème conceptuel partagé pour l'essentiel.

Suspects numéro un

Revenons à la question centrale de ce chapitre. Attendu que certaines personnes sont dignes de considération morale à part entière, qu'est-ce qui permet de croire qu'on a affaire à une forme de vie personnelle? Il devrait être clair à présent qu'une communication linguistique réussie laisse fortement présumer qu'on a affaire à une forme de vie personnelle. Il est très difficile d'imaginer une communication complexe avec une autre espèce vivante sans que celle-ci possède une intelligence et un schème conceptuel riches, incluant une conscience de soi et des préférences pour l'avenir.

Mais, en l'absence de communication linguistique réussie (et de perception extrasensorielle), quelles raisons avons-nous de supposer qu'un extraterrestre est non seulement intelligent, mais encore est une personne? Une fois de plus, il faut examiner son comportement. Prenez la Horta: Kirk a déduit son intelligence de son comportement, qui lui faisait identifier son phaseur et y réagir, mais ce comportement répond bien à la norme de non-personnes intelligentes, comme les poulets ou les moutons, et cela vaut pour le comportement initial d'autoconservation de l'exocomp dans « La qualité de la vie». Tout comme un comportement (relativement inflexible) d'autoconservation ne présume pas l'intelligence, un comportement intelligent d'autoconservation ne présume pas la conscience de soi. Certaines expériences menées en psychologie animale témoignent de façon probante que peu de mammifères supérieurs, sinon aucun, ont une conscience de soi : or tous ces animaux sont capables d'un comportement intelligent d'autoconservation, ou plutôt ils sont capables d'un comportement intelligent dont le but est l'autoconservation. Mais seuls les animaux dotés d'une conscience de soi ont un comportement intelligent d'autoconservation, parce qu'ils l'identifient comme tel. Aussi les officiers de la Flotte interstellaire devraient-ils guetter un comportement d'autopréservation assez riche pour indiquer une conscience de soi, comme celui testé par Data sur l'exocomp dans « La qualité de la vie».

Autre élément significatif du comportement : la protection de sa progéniture. La Horta en fait état, tout comme la créature de l'espace gigantesque quel'Entreprise-D tue,,sans le vouloir dans « L'enfant de la galaxie». Une fois de plus, il est possible que ce comportement suffise à établir l'intelligence, mais non la qualité de «personne», parce qu'on peut mobiliser son intelligence pour protéger ses petits sans posséder la moindre conscience de soi.

Avoir conscience de soi, et par là même distinguer les autres identités individuelles, voilà une condition nécessaire pour être un agent moral, et tout comportement suggérant la présence d'une attitude morale serait la preuve qu'on a affaire à une personne. Ici-bas, sur terre, on entend parfois des anecdotes sur le comportement étonnant des chiens ou des dauphins. Si nous avons de bonnes raisons de croire que Lassie a vraiment sauvé Timmy, et n'a pas simplement adopté un comportement qui s'est trouvé avoir pour conséquence le sauvetage - qu'elle n'a pas juste identifié le danger, mais a agi dans l'intention de le sauver-, nous aurions la preuve qu'elle est une personne. De même, les cas de dauphins protégeant diverses

espèces des requins serait, une fois authentifiés, la preuve que ces dauphins sont des personnes. Je considère notre attitude généralement favorable aux dauphins comme le signe d'un soupçon tenace qu'ils sont des personnes.

Soupçons mutuels

Riker: « Dans toute notre histoire, pourquoi n'a-t-on conservé aucun témoignage d'une visite que vous nous auriez faite, vous ou l'un de vos semblables ? »

Voyageur: « Quelle surprenante arrogance! li n'y a aucun témoignage parce que nous ne vous avons jamais encore rendu visite. »

Riker: « Et pourquoi ? »

Voyageur: « Parce que jusqu'à maintenant - pardonnez-moi - vous n'aviez aucun intérêt pour nous. C'est maintenant seulement que votre forme de vie mérite qu'on y prête une attention sérieuse. Désolé.»

Dialogue tiré de « Là où nul n'est allé » (NG)

De même que nous avons vu pourquoi la communication peut échouer entre des personnes d'espèces différentes, y a-t-il des cas où nous pourrions échouer à identifier des formes de vie personnelles? Cela pourrait se produire lorsqu'il y a de vastes différences d'échelle. Star Trek présente au moins trois de ces cas - différence de vitesse, de dimensions et d'intelligence. Par exemple, si les Scalosiennes « hyperaccélérées » de « L'espace d'un clin d'œil » (PS) n'avaient pas eu besoin des hommes de l'Entreprise pour conjoints (en leur faisant boire de l'eau scalosienne pour les accélérer!), on peut douter que Kirk et les autres eussent identifié leur existence. Dans « Le syndrome d'immunité » (PS) et « Le nuage » (V), on rencontre une forme de vie gigantesque. Dans « Jouer au démiurge » (TE9), Jadzia Dax croise un uni vers minuscule. Dans chacun de ces épisodes (pour des raisons dramatiques évidentes), « nous » avons un effet significatif sur « eux » et vice versa, mais dans la plupart des cas ces différences d'échelle auraient eu pour résultat une interaction minimale ou insignifiante, et la présence de personnes n'aurait été soupçonnée ni par les uns ni par les autres.

Dans Star Trek, la différence d'échelle la plus commune, et qui nous contraint à l'humilité, porte sur les niveaux d'intelligence. Les humanoïdes de la Fédération apparaissent comme des nains intellectuels, comparés à d'autres individus et à d'autres races. Dans la première série, on trouve souvent des demi-dieux, des magiciens et d'autres individus dotés de pouvoirs spéciaux comme Apollon dans « Qui pleure Adonis? », Korob et Sylvia dans « À pattes de velours », Gary Mitchell dans « Là où nul n'est allé », et Charlie Evans dans « Charlie X ». Puis il y a les représentants de races plus avancées comme les Kelviens dans « Ou tout autre nom », les êtres composés d'énergie pure dans « Retour à demain » et les Pourvoyeurs dans « Les joueurs de Triskelion ». Au XXNe siècle, dans la série La nouvelle génération, c'est le cas du Voyageur dans « Là où nul n'est allé » et « La fin du voyage », des Nagiliens dans « Le bail du silence », de Kevin Uxbrige le Douwd dans « Les survivants », des Paxiens

dans «Indices», de John Doe le Zalkonien dans «Transfiguration», des Cythériens dans « Le-n-ièrne degré», et, de façon plus frappante, des omnipotents (et quasi omniprésents) êtres Q. Dans la série la plus récente, les habitants protoplasmiques du Dominion dans Profondeur Espace 9, et les Vidiens dans les épisodes «Phage» et «Visages» du Voyageur sont particulièrement remarquables.

Si ces intellects supérieurs le désiraient, ils pourraient tout à fait éviter d'être détectés par nous. (De fait, c'est ce qui se produit - après quelques ratés - dans «Indices».) Un contact avec les humanoïdes intéresse nombre de ces extraterrestres, et généralement, ils ne manquent pas de bonté de « cœur » - ou de ce qui leur en tient lieu. Toutefois, le thème récurrent le plus frappant dans les rencontres de la Fédération avec les êtres surintelligents est leur manque de considération pour les humanoïdes. Cela soulève une question intéressante non moins que préoccupante. Ces créatures surintelligentes seraient-elles dans leur droit en nous traitant comme des sujets expérimentaux («Le bail du silence»), des esclaves («Qui pleure Adonis ? »), des divertissements («Les joueurs de Triskelion »), de la vermine («Les enseignes du commandement» et «Indices»), des pièces détachées («Retour à demain», « Le phage» et «Visages»), voire des garde-manger? Est-ce un bonus inespéré s'ils sont d'aventure gentils avec nous, comme le Voyageur et les Cythériens ?

C'est une question préoccupante parce que ici-bas, sur terre, nous autres humains réservons d'ordinaire ce genre de traitement à d'autres animaux intelligents. Nous les utilisons à échelle massive pour la recherche expérimentale, nous les domestiquons, nous les mettons dans des zoos et dans des cirques pour notre amusement et notre éducation, nous les détruisons par milliers lorsqu'ils menacent nos moissons ou nous paraissent nuisibles, nous utilisons leurs corps comme pièces détachées lorsque ça nous convient, nous les tuons et les mangeons. Nous justifions ce comportement par notre supériorité humaine sur les autres animaux intelligents de la terre. Pourquoi ? Souvent, un autre argument consiste à dire que nous sommes beaucoup plus intelligents qu'eux! Je suppose que le problème apparaît maintenant clairement. Si c'était là une bonne justification, une justification rationnellement irréfutable, du traitement que nous infligeons à des créatures différentes, moins intelligentes, sur terre, alors cette justification vaudrait également pour le traitement équivalent que nous infligeons à des races supérieures. Les Nagiliens du « Bail du silence» seraient moralement justifiés de tuer la moitié de l'équipage de l'Entreprise-D pour s'instruire sur la mort. Les Shéliaks des « Enseignes du commandement» (qui peuvent nous comprendre, et non l'inverse, ce qui témoigne de leur intelligence supérieure) auraient raison d'éradiquer une invasion de quinze mille humains sur leur territoire, etc.

Heureusement, un plus haut degré d'intelligence n'est pas à lui seul une justification de ce comportement. Dans une morale centrée sur la personne, dont je propose une version élargie, la différence entre personnes et non-personnes intelligentes n'est pas simplement une question de degré, mais de seuil. Un individu intelligent qui a conscience de soi et de l'autre, qui a des croyances et des

préférences sur l'évolution de sa vie, et des idées morales, a atteint le seuil qui le fait « personne ». Dès lors, cet individu est digne de considération morale à part entière - du moins, en ce qui concerne les libertés fondamentales. Si les autres humanoïdes ont tort de nous traiter en esclaves ou en nourriture, cela vaut pour toute personne non humanoïde. Non seulement la Fédération devrait guetter les suspects numéro un, mais d'autres personnes (peut-être exotiques) devraient nous guetter, nous !

Retournons au point de départ de cette discussion - aux exocomps dans « La qualité de la vie ». Bien qu'aucun contact n'ait été établi avec les exocomps, celui que teste Data suggère par son comportement qu'il a conscience de soi, et l'attitude des trois exocomps qui sauvent Picard et LaForge en sacrifiant l'un d'eux relève d'un choix moral. C'est là une évidence qui suffit à classer les exocomps parmi les personnes - mais n'exclut-elle pas une information importante, à savoir leur nature artificielle ? Comment quelque chose de créé par l'homme pourrait-il être vivant, ou intelligent, ou être une personne ? C'est ce dont traitera le chapitre suivant.

Chapitre 2

Des données insuffisantes

Du whisky synthétique, des commandants synthétiques!

Commentaire du capitaine Montgomery Scott sur le XXIV^e siècle, dans «Reliques» (NG)

Dans Star Trek, on trouve trois candidats types à la qualité de personne artificielle. Le premier est un ordinateur programmé à cette fin, notamment dans les épisodes de la première série intitulés « Le retour des archontes », « Un avant-goût d'Armageddon », « La pomme », « Le dernier ordinateur », « Le monde est creux, et j'ai atteint les cieux », l'épisode de La nouvelle génération intitulé «Émergence» et celui du Voyageur intitulé « Les têtes brûlées ». (Variation sur ce thème : l'apparition de machines informatiques sophistiquées dans les épisodes de La nouvelle génération intitulés «Évolution» et « La qualité de la vie».) Le second est un robot sophistiqué construit sur le modèle d'un être humain, ou, pour recourir à la terminologie de Star Trek, un androïde. Dans la première série, les androïdes font leur apparition dans « De quoi sont faites les petites filles? », « Moi, Mudd », « Retour à demain », et « Requiem pour Mathusalem ». L'androïde Data est un membre permanent de l'équipage de l'Entreprise-D dans la série La nouvelle Génération et les films Star Trek VII: Générations et Star Trek VIII: Premier contact. D'autres androïdes apparaissent dans les épisodes « Datalore », « Le rejeton », « Fraternité », « Descente I & II », « L'héritage » (NG) ainsi que dans l'épisode « Prototype » du Voyageur. Le troisième type de candidat est une création du « holodeck », qui correspond à la version XXIV siècle de la réalité virtuelle. Nonobstant l'hologramme médical d'urgence dans la série du Voyageur, certaines créatures du holodeck sont apparemment dotées de personnalité: voir « Le grand adieu », « 11 00 1 00 1 », « Élémentaire, mon cher Data » et « Le bateau dans la bouteille » (NG). « La paille » (V) et « Jeu d'ombres » (DS9) présentent également des personnages, peut-être dotés de personnalité, produits par la réalité virtuelle.

L'intérêt de recenser ces épisodes, c'est qu'on remarque une tendance : dans la première série, les scénaristes se souciaient profondément du statut des machines informatiques ; dans La nouvelle génération, l'attention se porte sur le statut des androïdes; dans les séries plus récentes, c'est le statut des individus produits par le

holodeck qui suscite l'intérêt. On peut interpréter cette tendance en suggérant qu'avec le progrès technologique des individus de type nouveau acquièrent progressivement un statut de personne - bien qu'à chaque époque on trouve des persifleurs pour estimer que ces individus ne sont pas des personnes et sont souvent dénués d'intelligence.

Mais peut-on considérer des entités artificielles comme des personnes ayant leurs droits propres ? À chaque époque, ce sont surtout les gens travaillant en collaboration étroite avec ces entités qui voient en elles des personnes: peut-être y a-t-il là une projection anthropomorphique abusive sur ces collègues artificiels. Ou peut-être l'excès est-il du côté des opposants et de leur préjugé. Des machines faites de métal et de silicone peuvent-elles penser, sentir, être des agents moraux ? Et de simples projections holographiques ? Voilà les questions que nous examinerons dans ce chapitre et dans le suivant. Un cas qui fait précédent

Data sera mon guide.

**Le commandant Bruce Maddox,
dans « Ce qui fait l'homme» (NG)**

Dans cet épisode de La nouvelle génération, où l'Entreprise-D aborde une nouvelle station spatiale, une surprise attend l'équipage. Le commandant Bruce Maddox, titulaire de la chaire de robotique à l'Institut Daystrom, et étudiant du créateur de Data, le Dr Noonian Soong, proclame son intention de démonter Data. Maddox veut découvrir une fois pour toutes le secret du « cerveau positronique » de Data de façon à construire une flotte d'androïdes - un pour chaque vaisseau spatial de la Fédération. Data est tout d'abord «intrigué», exprime ses doutes sur plusieurs aspects de la proposition, et signifie enfin par sa démission son refus de se soumettre à la manœuvre. Maddox rétorque que Data est la propriété de la Flotte, et qu'il ne peut pas démissionner. La plainte est portée devant Madame la procureur général, qui s'appuie sur un précédent local pour donner raison à Maddox. Picard fait appel, et l'on prend date pour une session au cours de laquelle Riker défendra la position de Maddox, et Picard celle de Data. Lors d'un contre-interrogatoire mené par Picard, Maddox affirme que Data n'est « pas sensible » et que la « sensibilité » (la qualité de personne, dirions-nous) requiert « de l'intelligence, une conscience en général, et une conscience de soi ».

Qu'est-ce que Data, aux yeux de Maddox ? « Une machine », dit-il. Maddox a raison, bien sûr. Data est une entité construite, dotée d'un cerveau positronique (un emprunt aux suggestions frivoles de l'écrivain de science-fiction Isaac Asimov), de réseaux neuronaux, etc., et de programmes extrêmement complexes. Data, semble-t-il, est constitué à part entière de composants non organiques, bien que les scénaristes le montrent intoxiqué dans « Maintenant seulement » (NG), et sous l'emprise de la drogue dans « Datalore » (NG). Suivant l'Encyclopédie de Star Trek, « Maintenant seulement » montre bien que Data possède certains composants organiques, mais ce diagnostic est incompatible avec tous les autres petits renseignements que nous avons

sur la construction de Data, et avec le fait qu'il assure ne pas vieillir (dans « Ce qui fait l'homme », Data dit que « [sa] condition ne s'altère pas avec le cours du temps ») : je préfère laisser de côté cette anomalie apparente.

Data a été conçu et monté par un être humain, le Dr Noonian Soong. Bien qu'il soit d'apparence plus ou moins humanoïde, il a une peau dorée et des yeux jaunes, et il est à peu près indestructible. Data est un membre à part entière de l'Enterprise-Ir : il a rang d'officier supérieur, et a été décoré plusieurs fois par la Flotte. Par ailleurs, il est quasi unique de son espèce. Après trois faux départs, le Dr Soong, qui passe pour mort, a fait un androïde presque identique à Data (« Lore », le frère aîné de Data, un fauteur de troubles qui passe la plupart de son temps en pièces détachées), ainsi qu'un clone androïde de sa propre épouse (voir « L'héritage », NG), mais, comme le bon docteur était en avance sur son temps, personne ne comprend tout à fait ce qui fait marcher Data.

Maddox veut Data pour guide, et nous de même - il sera pour nous la pierre de touche de la personne artificielle. Les arguments de Picard l'emportent, et la procureur général accorde à Data le bénéfice du doute. Pour elle, Data est un suspect numéro un - un candidat idéal à la qualité de personne. Toutefois, notre juridiction s'étend plus loin que celle du procureur. Outre la question : « Étant donné ce que nous savons de Data, faut-il conclure ou non qu'il est une personne? », nous nous demanderons si la qualité de personne artificielle est viable. Si la réponse à cette seconde question est négative, Data n'est pas une personne. Mais, même si nous répondons à la première question qu'il faut considérer Data comme une personne, celui-ci reste une fiction, voire une construction de l'imaginaire, impossible en réalité.

La condition de « sensibilité » (la qualité de personne) posée par Maddox constitue un bon point de départ. Elle correspond bien à celle que nous avons posée dans le chapitre 1 : L'intelligence indique qu'on est doté d'un cerveau, et la « conscience de soi » signifie qu'on « a une idée de son ego ... » Maddox y ajoute la « conscience en général », dont nous reparlerons au chapitre 3. N'y voyons pas une condition supplémentaire, car la conscience de soi présuppose sans doute une forme plus générale de conscience.

Lorsque Picard demande à Maddox si Data est intelligent ou non, Maddox reconnaît qu'il est intelligent, citant sa « capacité à apprendre, à comprendre et à gérer de nouvelles situations ». Je pense que Maddox prend la question à la lettre, mais il faut faire ici attention. Au xx- siècle, nous avons l'habitude de parler de « cerveaux électroniques » et de « bombes intelligentes », mais la plupart des gens font un usage métaphorique de ces expressions, et n'adhéreraient sans doute pas à leur signification littérale. Dans ce qui suivra, je prendrai les adjectifs intelligent, malin, rusé, réfléchi et stupide dans leur stricte valeur littérale. (Par exemple, dire qu'une chose est stupide, c'est dire que son intelligence est relativement basse - et donc qu'elle est intelligente.) Que la réponse de Maddox ait été littérale ou non, pourquoi Picard, l'avocat de Data, ne s'est-il pas tout simplement rué dessus? L'intelligence de Data ne suffirait-elle pas à démontrer qu'il n'est pas la propriété de la Flotte? En un mot : non. Personne ne nie que les poulets ou les moutons sont intelligents, et pourtant

ces animaux sont presque toujours considérés comme objets de propriété. Picard a raison de poursuivre: être une personne est condition suffisante, et sans doute nécessaire, pour éviter (moralement parlant) d'être tenu pour un objet de propriété.

Détail important: si Maddox affirmait que Data est intelligent, mais qu'il n'est pas une personne, il concéderait bien plus aux machines que nombre d'opposants. Beaucoup d'objections à l'idée que les machines sont des personnes concernent leur intelligence. Quand la nécessité s'en présentera, je distinguerai l'intelligence et la qualité de personne, d'autant que la plupart des débats philosophiques ont pour seul objet de savoir si les machines peuvent ou non être intelligentes. Mais Star Trek, et la vie en général, ne sont pas si précis. Je considérerai les avis des persifleurs comme des objections visant à la fois la qualité de personne artificielle et l'intelligence artificielle.

L'intelligence des machines : deux tests et trois points de vue

Il y a quatre cents ans, le philosophe et mathématicien René Descartes imagina qu'il rencontrait une machine construite à la ressemblance d'un être humain, et pouvant accomplir des mouvements étonnants. Descartes écrit dans son Discours de la méthode:

« S'il y en avait [de ces machines] qui eussent la ressemblance de nos corps, et imitassent autant nos actions que moralement il serait possible, nous aurions toujours deux moyens très certains pour reconnaître qu'elles ne seraient point pour cela de vrais hommes. Dont le premier est que jamais elles ne pourraient user de paroles ni d'autres signes en les composant, comme nous faisons pour déclarer aux autres nos pensées. Car on peut bien concevoir qu'une machine soit tellement faite qu'elle profère des paroles, et même qu'elle en profère quelques-unes à propos des actions corporelles qui causeront quelque changement en ses organes [...] mais non pas qu'elle les arrange diversement pour répondre au sens de tout ce qui se dira en sa présence, ainsi que les hommes les plus hébétés peuvent faire. Et le second est que, bien qu'elles fissent plusieurs choses aussi bien ou peut-être mieux qu'aucun de nous, elles manqueraient infailliblement en quelques autres, par lesquelles on découvrirait qu'elles n'agiraient pas par connaissance, mais seulement par la disposition de leurs organes. Car, au lieu que la raison est un instrument universel qui peut servir en toutes sortes de rencontres, ces organes ont besoin de quelque particulière disposition pour chaque action particulière; d'où vient qu'il est moralement impossible qu'il y en ait assez de divers en une machine pour la faire agir en toutes les occurrences de la vie de même façon que notre raison nous fait agir. »

Les philosophes ont appelé test de langage et test d'action les deux tests comportementaux de Descartes permettant d'éprouver la qualité de personne. La proposition de Descartes peut susciter trois positions. L'une est celle de Descartes lui-même, qui est conservatrice: dire qu'une machine échouerait aux deux tests. Autre position conservatrice: affirmer que, dans l'un et l'autre test, une machine peut en principe réussir, mais que son succès ne suffit pas à certifier qu'elle est une

personne. Maddox semble être un conservateur de la deuxième espèce. Un progressiste pense que la machine réussira au moins l'un des tests, et que son succès suffit à déterminer qu'elle est une personne. C'est la position adoptée par Picard.

Descartes affirme que nous pouvons, en principe, distinguer les êtres humains des machines à ce que ces dernières n'ont ni compréhension ni raison. Une fois encore, il y a des positions conservatrices à l'égard de cette proposition : soit une machine ne saurait réussir l'un ou l'autre test; soit elle pourrait les réussir, sans que cela garantisse qu'elle est intelligente. Le progressiste pense qu'une machine pourrait réussir au moins l'un des tests, et que son succès suffit à certifier qu'elle est intelligente. Maddox serait progressiste en ce qui concerne l'intelligence de la machine.

Dans un célèbre article publié en 1950, le mathématicien Alan Turing inventa une forme particulière du test de langage pour l'intelligence, en suggérant, plutôt que de s'interroger sur la capacité des machines à penser, de voir si elles peuvent réussir ce test. Turing, en bon progressiste, considérait qu'une machine pouvait en principe réussir le test, et qu'il fallait donc la considérer comme intelligente. Le test de langage de Turing s'inspire d'un jeu de société très en vogue dans l'ère prétélévisuelle, qui suppose un interrogateur et deux sujets, un homme et une femme. L'interrogateur est dans une autre pièce que les deux sujets, qui communiquent avec lui par un intermédiaire. Leur rôle est de convaincre l'interrogateur qu'ils sont, par exemple, de sexe masculin (ou féminin). L'interrogateur doit déterminer qui est l'homme; il peut poser n'importe quelle question à l'un ou l'autre sujet, et appuie son jugement uniquement sur le contenu des réponses. Si l'interrogateur devine mal, c'est la femme qui a gagné.

Dans la version révisée par Turing, l'un des sujets se voit remplacer par un ordinateur adéquatement programmé : l'objet du programme est de tromper l'interrogateur et de lui faire croire que l'ordinateur est un être humain. Pour que la machine gagne, elle doit réussir à imiter certaines réactions typiques d'un être humain. Imaginons maintenant qu'un ordinateur joue à ce nouveau jeu à maintes reprises, et que dans la moitié des cas l'interrogateur l'identifie comme un être humain. Dans ce cas, l'ordinateur a réussi ce qu'on a baptisé le test de Turing. Pour le rendre plus vivant, imaginez-vous que vous êtes l'interrogateur, et que vous menez le test à la façon d'une conversation en temps réel sur Internet. Si l'ordinateur peut vous bluffer au moins la moitié du temps, il a réussi le test de Turing.

La première hypothèse progressiste de Turing, c'est qu'une certaine sorte de machine - très précisément un ordinateur digital, programmé convenablement - peut en principe réussir le test de Turing. Cette affirmation repose en partie sur des preuves mathématiques sur lesquelles nous passerons, mais nous pouvons en saisir l'intuition fondamentale en examinant une machine de Turing. Figurez-vous une machine ainsi agencée: elle est dotée d'un très long ruban (de papier, par exemple) sur lequel apparaissent des symboles (les bits 1 et 0) ; elle peut faire avancer le ruban dans l'une ou l'autre direction, lire les symboles inscrits, les effacer et en écrire de nouveaux. Figurez-vous encore que la machine peut adopter un nombre limité d'états,

et qu'elle occupe l'un ou l'autre à part entière. De plus, lorsque la machine avance, elle ne progresse que d'un espace à la fois. Enfin, considérez que chaque étape de la machine dépend de sa table d'états, soit une série de règles déterminant ses opérations. Supposez que nous attribuons à la machine de Turing la table d'états que voici :

Table 1

État	Lire 0 sur le ruban	Lire 1 sur le ruban
E1	0 E1 D	E2 D
E2	1 E3 D	1 E2 D
E3	Stop	Stop ¹

Mettons que la machine de Turing soit en E1, et occupe la position foncée sur la section de ruban que voici :

0	0	0	1	1	0	1	1	1	0	0	0
---	---	---	---	---	---	---	---	---	---	---	---

Suivant la table d'états, lorsque la machine est en E1 et lit 1, elle doit écrire 0, passer en E2, et se déplacer vers la droite. (Par ailleurs, elle effacera le 1.)

0	0	0	0	1	0	1	1	1	0	0	0
---	---	---	---	---	---	---	---	---	---	---	---

La machine de Turing a exécuté l'opération décrite ci-dessus: elle est maintenant en E2 et a progressé d'un espace vers la droite, où elle rencontre un autre 1 sur le ruban. La table d'instructions exige que la machine laisse le 1 (en fait le matériau de base de la machine est ici simplifié : elle pourrait effacer le 1 et le réécrire - mais passons). Puis elle reste en E2 et progresse vers la droite.

0	0	0	0	1	0	1	1	1	0	0	0
---	---	---	---	---	---	---	---	---	---	---	---

La machine de Turing est maintenant en E2 où elle déchiffre un 0, qu'elle effacera, puis elle passe en E3, en progressant vers la droite.

0	0	0	0	1	1	1	1	1	0	0	0
---	---	---	---	---	---	---	---	---	---	---	---

Enfin, lorsque la machine de Turing est en E3 et qu'elle rencontre un 1, elle stoppe (comme indiqué dans la table d'instructions).

La machine vient d'additionner 2 et 3, un peu comme si on prenait un groupe de

deux crayons et un groupe de trois crayons, et qu'on les rassemblerait pour former un groupe de cinq crayons. Si l'on regarde ce qui est sur le ruban au départ (sur le premier schéma), on verra un groupe de deux 1 et un groupe de trois 1, séparés par un 0. À la fin de l'opération (quatrième schéma), il y a un groupe de cinq 1. Cela risque de vous paraître un peu fastidieux. Ce qui est étonnant, c'est la puissance de cette simple idée. Réfléchissez ne serait-ce qu'un moment à tout ce qu'accomplissent les ordinateurs digitaux, qui tous fonctionnent sur ce simple principe. Avec un ruban assez long et assez de mémoire pour stocker la table d'états appropriée, la machine de Turing peut reproduire l'opération de n'importe quel ordinateur digital. (Oublions la vitesse pour l'instant.)

Il est temps d'introduire une terminologie utile. L'entrée des données est ce que la machine lit sur le ruban, la sortie est ce qu'elle efface ou écrit (ou laisse intact). La table d'états est le programme ou software de la machine, et le restant est le matériau de base de la machine. (Notez bien que certaines instructions sont incrustées dans ce matériau - ainsi, ne progresser que d'un espace à la fois.) L'entrée n'est pas forcément en code binaire ; cela peut être des chiffres, comme 2 ou 3 ; ou des lettres, comme A et B ; ou des signes de ponctuation, comme des espaces et des virgules. La même chose vaut pour la sortie.

L'important, c'est qu'il n'y a pas de limite théorique à ce qu'un ordinateur digital peut accepter comme entrée linguistique, ou produire comme sortie linguistique. (Il y a des limites théoriques intéressantes - reliées au théorème de Gödel - sur les rapports entre entrée et sortie, mais aucune n'établit de différence entre les ordinateurs et les êtres humains, et je les laisserai donc de côté.) Pas plus qu'il n'y a de limite théorique sur le genre de programme que peut gérer un ordinateur digital, tant qu'il a assez de mémoire. On a donc toutes les raisons de supposer qu'un ordinateur digital peut réussir en principe le test de Turing. Bien sûr, il y a toutes sortes de limites pratiques, mais dans les cinquante dernières années elles ont singulièrement reculé. Turing lui-même pensait qu'en l'an 2000 nous aurions ou serions très proches d'avoir un gagnant au test de Turing. Une machine qui serait un ordinateur au programme approprié pourrait-elle gagner le test de Turing? Oui! Et si Descartes pensait le contraire, c'est lui qui se trompait.

Comme démonstration de ce principe, examinons le programme ELIZA, conçu par Joseph Weizenbaum il y a une vingtaine d'années, et qui a pour objet l'imitation d'un psychothérapeute rogérien. (Ces psys adoptent une attitude passive, et répètent les réponses des patients ou posent des questions qui en dérivent directement.) Voici le genre de « conversation » qu'un « patient » peut avoir avec ELIZA ; les questions d'ELIZA sont en majuscules.

Les hommes sont tous les mêmes.

COMMENT ÇA?

Ils n'arrêtent pas de nous embêter d'une façon ou d'une autre.

POUVEZ-VOUS NOUS CITER UN EXEMPLE PRÉCIS ?

Eh bien, mon copain m'a forcée à venir ici.

VOTRE COPAIN VOUS A FORCÉE À VENIR ICI ?

Il dit que je suis déprimée une bonne partie du temps.

JE SUIS NAVRÉ D'APPRENDRE QUE VOUS ÊTES DÉPRIMÉE.

C'est la vérité. Je suis malheureuse. "

PENSEZ-VOUS QUE VENIR ICI VOUS AIDERA À NE PLUS ÊTRE
MALHEUREUSE ?

J'ai besoin qu'on m'aide, c'est sûr.

QU'EST-CE QUE ÇA SIGNIFIERAIT POUR VOUS, QU'ON VOUS AIDE?

Peut-être que j'apprendrais à m'entendre avec ma mère.

PARLEZ-MOI DE VOTRE FAMILLE.

Ma mère prend soin de moi.

QUI D'AUTRE DANS VOTRE FAMILLE PREND SOIN DE VOUS?

Mon père.

VOTRE PÈRE?

Vous ressemblez à mon père par certains côtés.

OÙ EST LA RESSEMBLANCE ?

Vous n'êtes pas très agressif, mais, à mon avis, vous ne voulez pas que je le
remarque.

QU'EST-CE QUI VOUS FAIT CROIRE QUE JE NE SUIS PAS AGRESSIF ?

Vous ne m'opposez aucun argument.

ET POURQUOI, À VOTRE IDÉE ?

Vous avez peur de moi.

ÇA VOUS FAIT PLAISIR DE PENSER QUE J'AI PEUR DE VOUS?

Mon père a peur de tout le monde.

QU'EST-CE QUI VOUS VIENT D'AUTRE À L'ESPRIT QUAND VOUS
PENSEZ À VOTRE PÈRE ?

Tous des brutes.

CELA A-T-IL À VOIR AVEC LE FAIT QUE VOTRE COPAIN VOUS A FORCÉE
À VENIR ICI ?

ELIZA imite à merveille un psychothérapeute rogérien, comme il saurait imiter le conseiller Troi ! Mais son comportement est extrêmement rigide, et, dès qu'on sort du domaine de la thérapie, cela saute aux yeux. Une fois que le patient cesse de jouer à la thérapie, les réactions d'ELIZA paraissent très étranges! Mais ELIZA est un programme réduit conçu pour un ordinateur primitif par rapport aux critères actuels. Essayez de vous figurer un ordinateur doué d'une immense mémoire, d'une immense capacité, et d'une programmation appropriée qui soit extrêmement complexe. Il n'y a pas de raison, en principe, pour qu'il ne réussisse pas le test de Turing. Quitte à se montrer conservateur, autant faire porter l'opinion conservatrice sur le caractère approprié ou non des tests en question.

Fonctionnalisme

Selon Turing, la pensée consiste simplement à traiter les informations d'une certaine façon, et cela pourrait être accompli par un ordinateur digital aussi bien que par un être humain. C'est là une vision fonctionnaliste de l'esprit, sans doute la position dominante dans la philosophie moderne de l'esprit. Je vais préciser ce qu'est le fonctionnalisme, mais d'abord je voudrais en finir avec l'intuition. Prenons le cerveau pour commencer. Votre état mental actuel dépend essentiellement de l'état actuel de votre cerveau (postulat confirmé par les cas de lésions au cerveau). Mais quelles sont les données de votre cerveau qui importent et déterminent vos états mentaux ? Par exemple, je présume qu'en ce moment précis vous croyez que Washington D.C. est la capitale des États-Unis. Voilà votre état mental actuel. Mais en quoi votre cerveau justifie-t-il que vous voyiez en Washington D.C. la capitale des États-Unis ? Ce qui importe, est-ce sa constitution ou son fonctionnement ? L'intuition fonctionnaliste suggère que ce sont les états fonctionnels du cerveau - ce qu'il fait - qui comptent. Une conséquence importante du fonctionnalisme mental, c'est la grande variété des états mentaux réalisables, l'idée que c'est le mélange de diverses substances qui suscite physiquement tel ou tel état mental.

Les états fonctionnels sont caractérisés en tout et pour tout par des chaînes de causalité, que j'illustrerai par des exemples. La plupart des métiers sont caractérisés fonctionnellement. Si vous faites la plonge, votre office est de prendre (à l'entrée) des plats sales et de produire (à la sortie) des plats propres. C'est là votre chaîne causale. La plupart des appareils sont caractérisés fonctionnellement, eux aussi. Une souricière est un appareil qui a pour office de prendre (à l'entrée) des souris libres et de produire (à la sortie) des souris prises au piège. Ce sont là des chaînes causales très simples que nous pouvons décrire en détail si nous le souhaitons. D'autres chaînes ne sont pas si simples, mais nous pouvons être certains qu'elles sont caractérisées fonctionnellement dans leur intégralité. Prenez les livres, par exemple. On pouvait croire naguère qu'un livre avait certaines propriétés physiques - celle notamment d'être constitué de papier et d'encre. Mais la fonction d'un livre est de stocker des informations rendues accessibles par l'auteur et de permettre leur extraction par le lecteur. Ceci peut être fait de maintes façons, et il y a maintenant des livres sur bandes magnétiques, des livres métalliques en braille, des livres sur disques compacts, des livres sur ordinateur, des livres sur le Net, etc. Vous voyez, j'espère, que spécifier en détail la chaîne causale d'un livre est source d'infinies complications. C'est pourquoi le fonctionnaliste s'attend à d'incroyables difficultés lorsqu'il s'agit de spécifier intégralement la chaîne causale d'un objet aussi compliqué que le cerveau (ou même d'un unique état mental, telle l'idée que Washington D.C. est la capitale des États-Unis). Toutefois, la plupart des fonctionnalistes estiment que les chaînes causales propres aux états mentaux correspondent à un traitement de données identique à celui des ordinateurs.

La thèse suivant laquelle une grande variété d'états mentaux est réalisable donne sens au postulat d'une intelligence extraterrestre radicalement différente de la nôtre. Si la Horta, à base de silicone, de « Démon dans le noir », est intelligente, alors il y a au moins deux configurations physiques différentes de l'intelligence, la

nôtre et celle de la Horta. Qui plus est, la définition fonctionnaliste des états mentaux offre un large champ à la science de l'esprit - la discipline baptisée science cognitive. Si le cerveau n'est pas l'unique agent de configuration de la pensée, alors il n'est pas strictement nécessaire d'étudier l'un pour comprendre l'autre. La recherche en matière d'intelligence artificielle avance pour dogme principal qu'un ordinateur reflétant adéquatement le fonctionnement du cerveau lorsqu'il configure la pensée est plus qu'un simple modèle d'intelligence - il est intelligent à part entière. La meilleure façon de produire ce résultat est de programmer un ordinateur pour qu'il fournisse des productions linguistiques appropriées à une série d'entrées linguistiques. La plausibilité initiale du fonctionnalisme paraît étayer la thèse de Turing, qui considère son test comme preuve suffisante d'intelligence.

Le fonctionnalisme tel que je l'ai décrit est une théorie matérialiste de la pensée et des états mentaux, puisqu'il affirme que la pensée est partie et portion du monde matériel, naturel et physique. Mais pourquoi imaginer ceci? Peut-être l'esprit est-il si mystérieux parce qu'il ne fait pas du tout partie du monde matériel? Devrions-nous supposer une forme spéciale de « matériau mental» pour en rendre compte? Il faudrait combattre cette opinion pour une raison irréfutable - que nous illustrerons en faisant appel à l'épisode de La nouvelle génération intitulé « La phase suivante». Geordi LaForge et l'enseigne Ro Laren sont rendus invisibles au reste de l'équipage de l'Entreprise-D lorsqu'il « sont mis hors phase». Eux, toutefois, peuvent voir l'équipage. Le problème (comme l'a souligné Lawrence Krauss dans La physique de Star Trek), c'est qu'on ne peut pas jouer sur les deux aspects: pour voir, LaForge et Ro doivent nécessairement absorber la lumière, et si c'est le cas ils sont visibles. (Au passage, l'argument commun pour l'existence d'âmes invisibles, fondée sur l'expérience de mort approchée, souffre d'une faille équivalente.) Quel rapport avec le « matériau mental»? Eh bien, si votre pensée n'est pas partie et portion du monde matériel, alors on a du mal à voir comment elle pourrait interagir avec le monde matériel - impossible de jouer sur les deux aspects. Et si votre pensée n'interagit pas avec le monde matériel, alors aucune des actions que vous accomplissez dans ce monde ne peut être causée par vos états mentaux. Si vous pensez (à juste titre, me semble-t-il) que votre comportement est largement dû à vos états mentaux, il vous faut rejeter l'hypothèse selon laquelle votre pensée est immatérielle.

Un problème mécanique

Parmi les ordinateurs décrits dans Star Trek, combien réussiraient le test de Turing? Aucun, je le pense. Les ordinateurs à bord des vaisseaux spatiaux échoueraient -, même l'ordinateur principal de l'Entreprise-D. En premier lieu, il en «sait» trop - il stocke bien trop d'informations, et elles sont extraites bien trop rapidement. Ensuite, il exécute des calculs compliqués avec une extrême rapidité. Enfin il réplique à certaines questions par des énoncés tels que « Données insuffisantes à gérer» ou « Impossible de répondre à votre demande. Spécifiez les paramètres SVP», énoncés qu'un être humain ordinaire ne prononcerait jamais. Qui

plus est, l'ordinateur du vaisseau ne manifeste jamais la moindre impatience quand on répète la même question ou qu'on lui pose des questions stupides ; il est immanquablement courtois et répond toujours lorsqu'on l'interroge. (De fait, il a souvent des réponses percutantes à des questions ou propos vagues ou ambigus.) L'ordinateur connu sous le nom de Landru dans « Le retour des Archontes » (PS) s'en tire un peu mieux, mais il ne faut pas longtemps à Kirk et à Spock pour réaliser qu'ils ont affaire à une machine. Dans l'un et l'autre de ces cas, toutefois, l'ordinateur n'est pas programmé pour imiter à la perfection les réactions linguistiques d'un être humain ordinaire. L'ordinateur de l'Entreprise-D est programmé pour servir, non pour tromper, et Landru est programmé pour administrer les citoyens grégaires de Beta III. Leur échec probable au test de Turing renvoie peut-être à la question d'une bonne ou d'une mauvaise programmation.

Le canon de Star Trek suggère une raison plus profonde, plus fondamentale à l'échec de la machine - du moins lorsqu'on lui pose la bonne question. Un ressort essentiel dans l'intrigue de « Moi, Mudd » (PS) est que les androïdes (et par extension les ordinateurs) ne peuvent gérer l'irrationalité ou la contradiction. Star Trek embraye ici sur un thème courant de la science-fiction, mais « Moi, Mudd » lui donne une application particulière. Le négociant Harry Mudd persuade l'ensemble de l'équipage de l'Entreprise-D de se téléporter sur une planète anonyme peuplée de centaines de milliers d'androïdes prêts à les servir. Toutefois, Mudd ignore que les androïdes ont un autre motif: ils considèrent les humains comme illogiques et instables par nature; leur plan est de s'emparer non seulement de l'Entreprise mais de l'univers connu, et de les régir tous deux de façon ordonnée. Lorsque l'équipage réalise le danger, il met les androïdes hors de combat en les exposant à l'irrationalité et au paradoxe. Le comportement irrationnel des membres d'équipage apparaît comme un exemple de rare bêtise, et cette capacité de la bêtise à court-circuiter une machine demeure un mystère; une machine n'y prêterait sans doute aucune attention, pas plus que moi ! L'équipage recourt aussi au paradoxe. Spock assure qu'il aime Alice-27 et, simultanément, qu'il hait Alice-210, qui présente pourtant des qualités identiques : les deux androïdes disjonctent immédiatement. Puis Kirk et Mudd court-circuitent le chef des androïdes, Norman, en lui tenant les propos suivants:

Kirk : « Tout ce que vous dit Harry est faux. Retenez bien ceci - tout ce que vous dit Harry est mensonge. »

Harry : « Je suis un menteur. »

Le pauvre Norman essaie de résoudre ce paradoxe: si tout ce que dit Harry est mensonge, alors il ment en disant qu'il ment, mais s'il ment ici c'est qu'il ne ment pas, etc. En moins d'une minute, Norman fait Pschhhhhh !

Tout ceci est très amusant mais ne court-circuiterait pas un androïde sophistiqué comme Norman, qui reconnaîtrait probablement qu'il s'agit d'un paradoxe facilement soluble. Mentir, c'est affirmer délibérément ce qu'on pense n'être pas la vérité (même si ça peut l'être), et si Harry ment en disant: « Je mens », c'est qu'il croit qu'il ne ment pas. Norman pourrait en conclure que Harry ne comprend tout simplement pas ce qu'est le mensonge, puisqu'il ment en croyant qu'il ne ment pas.

D'un autre côté, il pourrait donner à Harry le bénéfice du doute et conclure qu'il ne ment sans doute pas (ce serait ma conclusion personnelle). Et si c'est vrai, puisque Kirk dit que tout ce que Harry dit à Norman est un mensonge, soit Kirk ment, soit il se trompe. Il n'y a pas ici de contradiction, et le paradoxe est résolu. Tout paradoxe est engendré par un ensemble de propositions conjointes, et l'on peut toujours suspendre sa croyance dans la conjonction sans décider quelle proposition il faut rejeter.

Norman pourrait donc conclure soit que Harry est la proie d'une confusion profonde, soit que Kirk se trompe, soit qu'il ment- et régler ainsi le problème. De même, Alice-27 pourrait conclure soit que Spock est irrationnel, soit qu'elle n'est pas identique à Alice-210, soit qu'il est possible d'aimer un individu tout en haïssant un autre individu, même si celui-ci possède des qualités identiques. Si les machines peuvent être programmées non seulement pour accepter et gérer, mais également pour produire ce genre de disjonctions, il est improbable que des ordinateurs sophistiqués soient confondus par le paradoxe. Et il est non moins improbable qu'ils échouent au test de Turing, vu leur réaction aux paradoxes.

Il est d'autres épisodes, tirés de la première série, où Kirk recourt à la logique pour combattre des ordinateurs sophistiqués. Dans « Le retour des Archontes », il fait admettre à Landru que celui-ci doit détruire le mal, puis qu'il est lui-même mauvais, et presque aussitôt... Pshhhhh ! Dans « L'ultime ordinateur », Kirk fait admettre à l'ordinateur M5 (qui régit l'Entreprise et fait des ravages) que le meurtre est un mal passible de mort, puis le convainc qu'il a commis lui-même un meurtre. Cette fois-ci, l'ordinateur ne s'autodétruit pas immédiatement, mais se lance dans une série d'actions qui causeront sa propre destruction, comme un châtiment auto-imposé. Dans « L'enfant des fées », l'Entreprise rencontre Nomade, un stérilisateur biologique particulier, qui a été accidentellement programmé pour détruire toutes les entités imparfaites. Kirk fait admettre à Nomade qu'il est lui-même imparfait, et celui-ci s'autodétruit. Peut-être le message de ces épisodes est-il que les ordinateurs sont par essence trop rigides dans leur soumission à la logique, trop inflexibles pour leur propre bien. Les arguments de Kirk peuvent tous être représentés comme des exemples de modus ponens: forme d'argument qui stipule que « si p, alors q et p, donc q ». (Si Nomade est imparfait, alors Nomade doit être détruit. Nomade est imparfait. Dès lors Nomade doit être détruit.) C'est une forme d'argument valide - si les prémisses sont vraies, la conclusion doit l'être également. Mais toute forme d'argument valide peut être utilisée de façon inverse, comme reductio : puisque « si p alors q et p » entraînent q, et donc non-q, alors « si p alors q et p » est nécessairement faux. Si Nomade a un protocole assez solide interdisant l'autodestruction, soit il niera devoir être détruit, soit il niera être imparfait. M5 devrait avoir des raisons solides de produire une reductio des arguments de Kirk. D'une part, bien des gens nient que le châtiment approprié au meurtre soit la peine de mort; d'autre part, M5 affirme qu'il a agi en autodéfense lorsqu'il a tué, et devrait donc nier avoir commis un meurtre. La logique est une arme à double tranchant, et il n'y a pas de raison de penser que les ordinateurs puissent facilement être persuadés de s'autodétruire ; dès lors, il est peu probable qu'ils échoueront pour cela au test de

Turing.

Bien sûr, on peut invoquer bien d'autres raisons à cet échec. Mais cela n'implique pas que ces ordinateurs sophistiqués soient dénués d'intelligence. Turing affirme que la réussite à son test est suffisant pour garantir l'attribution de l'intelligence, mais elle n'est sûrement pas indispensable! Et cela vaut mieux, puisqu'il est sans aucun doute des tas d'humains intelligents qui obtiendraient de piètres résultats au test de Turing, et seraient pris pour des machines. Imaginez par exemple un être humain entièrement dépourvu de capacités linguistiques (l'enfant surprotégé d'un couple sourd, incapable de manier le langage par signes), ou dont l'unique langue serait incompréhensible (comme Dathon dans « Darmok » [NG]), ou certains autistes. Peut-être M. Spock lui-même, dont l'intelligence ne fait aucun doute, échouerait-il au test de Turing. («Vous feriez un remarquable ordinateur, M. Spock », dit Kirk dans « Le retour des Archontes».) Certains humains pourraient même échouer délibérément. Dans « Le culte du héros» (NG), l'Entreprise-D recueille à bord du Vico Timothée, l'unique survivant d'un accident. Traumatisé, il s'attache à Data et finit par affirmer qu'il est lui-même un androïde, en répondant à des questions prévenantes telles que « Comment vous sentez-vous ? » par «J'opère dans le cadre des paramètres établis »,

Il est également à parier que Data échouerait au test de Turing. Comme l'ordinateur de l'Entreprise-D, son principal obstacle serait sa capacité supérieure à stocker les données, ses pouvoirs de calcul et sa vitesse, qui le trahiraient sans doute. Puis il y a les fameux malentendus de Data sur les émotions et humeurs humaines (encore qu'on puisse imaginer que l'interrogateur diagnostiquerait un humour délibéré dans certaines répliques innocentes). Mais encore une fois, en théorie il n'y a aucune raison de supposer que Data ne pourrait être programmé de façon à réussir le test de Turing. (Dans un exemple qui me contredit en apparence, le bétazoïde Tarn Elbrun, doué de télépathie, dit à Data dans « Tin Man» : « Incroyable - un androïde. Je ne peux pas du tout lire en vous. C'est comme si vous n'étiez pas là.» Je ne vois aucune raison - si Data est intelligent et si la télépathie existe bel et bien - pour que l'intelligence de Data ne soit pas télépathique. Quoi qu'il en soit, les bétazoïdes ne peuvent pas lire non plus dans la pensée des Ferengi, comme il est établi dans « Le dernier bastion». Je laisserai donc de côté le témoignage d'Elbrun.) Puisqu'il n'est pas indispensable de réussir le test de Turing pour se voir attribuer une intelligence, et puisque Data échouerait à ce test, en est-il un plus juste - que Data aurait une bonne chance de réussir sans qu'il y ait besoin de le reprogrammer in extenso ?

Des tests combinés

*Votre créateur s'est donné beaucoup de mal pour vous
donner une apparence humaine.*

Le Dr Julian Bashir à Data dans « Patrimoine I »

Qu'est-ce qu'un test efficace d'intelligence androïde? Nous pouvons utiliser le test de Turing comme modèle pour unifier les tests de langage et d'action proposés

par Descartes. Supposons que le sujet du test soit un androïde. Pour réussir le test combiné, il lui faut tromper un interrogateur humain ordinaire et lui faire croire qu'il a affaire à un être humain ordinaire, dans une situation où l'interrogateur peut voir et entendre l'androïde, et interagir avec lui d'autres façons. Dans ce test, toute différence entre le comportement d'un androïde et celui d'un être humain typique suffirait à rejeter l'androïde. Ce test est manifestement injuste: Data serait immédiatement disqualifié par ses mouvements cocasses de tête et de cou, par exemple. De plus, imaginez toutes les ressources gâchées consacrées à monter l'androïde de façon qu'il imite des expressions humaines sans intérêt.

Comment construire une machine capable d'imiter un peu plus que le comportement linguistique humain adapté par l'ordinateur- une machine susceptible d'imiter toute la gamme des comportements humains ? Il lui faudrait réagir adéquatement au langage humain, et se faire un chemin dans notre monde en usant de modalités sensorielles similaires à la vision, l'ouïe, le toucher, l'odorat. Turing pensait que pour bâtir un gagnant à son test, une bonne idée serait de le modeler sur une pensée enfantine, et je soupçonne qu'il répéterait ici son conseil - construisez un androïde enfant, avec une capacité d'apprendre de son environnement identique à celle d'un enfant humain. Mais la tâche présente d'énormes difficultés. En plus de produire le matériau de base équivalant à une pensée enfantine, et de le programmer de façon qu'il puisse apprendre de façon appropriée, un constructeur d'androïde est confronté à ce que j'appellerais le problème de la transposition.

Rappelez-vous que dans le test de Turing on est dans une sorte d'interface avec la machine (sur Internet, par exemple). Il nous faut résoudre deux problèmes de transposition : tout d'abord le problème d'entrée, à savoir comment transformer le langage naturel qu'on produit en une forme que la machine puisse gérer; ensuite le problème de sortie, à savoir comment retransformer les résultats en une forme accessible. Ce sont des problèmes que nous avons déjà résolus, mais l'Internet et vous coopérez à leur solution en mettant directement vos questions sous un « cyber-format » et en accédant aux réponses sous ce même format. Si les questions étaient formulées verbalement, il y aurait besoin d'un autre moyen de transformation - par exemple, un intercesseur humain qui entrerait vos questions. Et s'il vous fallait une réponse verbale, alors vous auriez besoin d'un autre mode de transformation de sortie.

Lorsque vous parlez à un androïde, il doit avoir son propre moyen de transformer l'entrée verbale (les ondes sonores) en une forme qu'il puisse gérer. Et s'il doit vous répondre, il doit avoir son propre moyen de retransformer ses résultats en ondes sonores. Il lui faut donc des transposeurs d'entrée et de sortie. Pensez à votre propre corps : les ondes sonores atteignent votre oreille externe, et sont dirigées à travers l'air vers le tympan. Puis elles sont transmises par l'os - (le marteau, l'enclume et l'étrier) à la « fenêtre ovale », et de là à travers le fluide de la cochlée, où des vibrations font transmettre aux cils de cellules spécifiques (« d'une façon ou d'une autre », pour citer un manuel classique de physiologie) les impulsions nerveuses qui atteignent le cortex auditif du cerveau. Les impulsions nerveuses elles-

mêmes sont des impulsions électriques produites chimiquement. Il faut noter deux points importants à propos de ce processus : les transposeurs auditifs sont extrêmement complexes, et nous ne comprenons pas entièrement leur fonctionnement. À la sortie - au niveau de la production sonore - les choses sont encore plus compliquées. Pensez à tous les transposeurs dont dispose le corps - transposeurs d'entrée pour les quatre autres sens, et tous les autres transposeurs de sortie, notamment ceux du mouvement. Un androïde comme Data devrait avoir des équivalents ! La bonne nouvelle, pour les constructeurs d'androïdes, c'est que nous savons que les problèmes de transposeurs peuvent tous être résolus en principe, puisque notre corps donne l'exemple.

Il est peut-être facile de perdre de vue une conséquence de la construction des corps. Si on veut bâtir un androïde de taille adulte, doté de la plupart des capacités humaines, il y aura certaines choses qu'il ne pourra faire. Et s'il est parfois difficile de discerner les intentions des scénaristes de Star Trek, je les soupçonne de tomber de temps à autre dans un malentendu ordinaire concernant les machines intelligentes. Nous avons déjà évoqué ce qu'on pourrait appeler la division mentale du travail en nous-mêmes - notre cerveau est subdivisé en facultés et autres unités fonctionnelles, dont beaucoup sont inaccessibles et étanches à la conscience. De même que la nature trouve tout à fait logique de nous construire ainsi, le Dr Soong trouve logique de bâtir ainsi Data. La dernière chose qu'on souhaiterait au monde, c'est un exécutif mental qui aurait le contrôle de tout; par analogie, imaginez le PDG d'une grande entreprise qui insiste pour avoir la haute main sur la moindre opération de sa firme. Bien sûr, certains systèmes de contrôle sont nécessaires, mais la plupart du temps on ne souhaite pas que l'agent soit au fait de tout ce qui se passe. Pourtant, lorsque Data prend des leçons de danse avec le Dr Crusher (dans « Le jour de Data » [NG]), il n'est que trop évident qu'il contrôle entièrement ses actes. Ailleurs, il est dit qu'il possède trop de capacités pour se faire passer pour un être humain crédible; dans « Le culte du héros », nous apprenons que Data peut analyser chimiquement tout ce qu'il absorbe. Réfléchissez un moment à l'incroyable gâchis de ressources que cela a impliqué de donner à Data la capacité non indispensable - puisqu'il n'a pas besoin de nourriture ou de boisson - d'imiter l'acte de boire et de manger! Rajoutons ce que nous apprenons sur Data dans « Jouir du présent ». « Vous êtes fonctionnel au plus haut point, n'est-ce pas? », ronronne l'officier de sécurité en chef Tasha Yar, lascive sous l'influence d'une boisson intoxicante. « Bien sûr, mais ... », réplique Data. « Jusqu'à quel point? », interrompt Yar. « Dans tous les domaines, bien sûr, dit Data. Je suis programmé pour offrir une large gamme de plaisirs. »

Dans « Violations » (NG), lors d'un débat sur les caprices de la mémoire, Geordi LaForge dit à Data : « Tu te rappelles chaque seconde de chaque instant de ta vie », et Data répond : « Exact. » Voici qui est entièrement absurde : les réserves mémorielles de Data seraient comblées en un rien de temps. Par comparaison, voyez ce que c'est que de surfer sur le World Wide Web. Lorsqu'on passe en revue diverses pages du web, on les stocke dans une case qui ressemble beaucoup à une mémoire à court terme. Si vous êtes comme moi, il y a un tas de choses sur le web qui ne vous

intéressent pas, et lorsque vous quittez le réseau, vous êtes content de vider la case. Si vous voulez enregistrer quelque chose sur votre disque dur (sorte de mémoire à long terme), cela demande un effort distinct. Mais supposez que, au lieu de n'enregistrer que ce que vous voulez conserver, votre browser enregistre la moindre page que vous consultez. En un rien de temps, vous n'aurez plus d'espace libre sur votre disque dur. Observez maintenant votre propre mémoire et toutes les expériences que vous avez eues dans votre vie - tous les petits déjeuners, tous les brossages de dents, tous les éternuements ; tous les signaux de circulation que vous avez perçus, etc. Il devrait paraître évident qu'un individu fonctionnel ne saurait se rappeler toutes ces choses, ni ne le souhaiterait !

Enfin, toutes les fois que Data a besoin de se rappeler quelque chose, il lui faut apparemment élaborer de nouveaux programmes. (Ainsi, dans « En théorie » [NG], il élabore un programme pour mener à bien une relation romantique.) Le simple fait d'être un androïde est déjà dur, si l'on considère tout ce qu'il faut apprendre, mais être un androïde autoprogrammateur serait franchement infernal !

Puisque le test combiné est injuste, et qu'il est si difficile de concevoir une machine capable de le réussir, il nous faut revoir le test. Un test beaucoup plus efficace pour évaluer l'intelligence androïde est celui que nous utilisons pour identifier les suspects numéro deux dans le cas de la vie extraterrestre (voir chapitre 1). Il s'agit d'un test combiné où il nous faut accepter un comportement approprié soit linguistique, soit non linguistique comme preuve suffisante d'intelligence. Ce test combiné laisse au sujet plusieurs possibilités de réussite, et nous évitons ainsi le danger de ne pas reconnaître l'intelligence qui pourrait exister dans un corps inapproprié : songez aux enfants dont les cerveaux sont tout à fait normaux, mais dont les corps difformes leur valent parfois d'être catalogués comme retardés mentaux. Ce test laissant plusieurs possibilités d'échec au candidat, il est utile lorsqu'on est en contact linguistique avec un individu capable de supercherie (y compris envers lui-même) : voir et entendre l'individu peut offrir des informations supplémentaires cruciales. Et un individu pourrait réussir ce test combiné en imitant des comportements humains (une machine qui réussit le test de Turing réussit le test combiné), mais sans que ce soit obligatoire. Nous pouvons imaginer par ailleurs un test encore plus rigoureux pour détecter les personnes : celui que nous avons utilisé pour identifier les suspects numéro un dans le chapitre 1. Ici encore, il suffirait de réussir le test de Turing pour réussir le test combiné prouvant la qualité de personne. Revenons donc à la question : comment définir un gagnant du test de Turing ?

Et le gagnant est ...

*S'ils 'agissait d'une boîte montée sur roulettes,
je n'aurais pas à combattre une telle opposition ...*

Maddox, dans « Ce qui fait l'homme »

Y a-t-il un argument contre la position conservatrice ? Pourquoi serions-nous

impressionnés par une machine qui réussit le test de Turing ? Après tout, les apparences peuvent être trompeuses. Tout ce qu'un progressiste peut faire ici, c'est en appeler à la cohérence. Rappelez-vous le problème des autres pensées discuté au chapitre 1- où l'on doit justifier l'idée qu'un autre individu pense alors qu'on n'a pas accès à ses pensées. Le seul moyen de se sortir de ce dilemme, c'est de considérer le comportement, linguistique ou non linguistique, des autres êtres humains que vous rencontrez comme preuve suffisante de leur intelligence - ou leur intelligence comme la meilleure légitimation de leur comportement. (Par contraste, le comportement d'un patient dans le coma est relativement simple, comme celui d'un rocher ou d'une chaise, et nous pouvons en déduire une absence de pensée: c'est pourquoi nous envisageons parfois de le « débrancher».) J'allègue donc que l'on a raison d'attribuer une pensée à un autre être humain si la meilleure façon d'expliquer son comportement est de diagnostiquer chez lui une intelligence. Autre remarque intéressante quant à ces déductions : dans la plupart des cas, nous concluons" à une intelligence lorsque nous voyons à quoi ressemblent les individus - plus exactement, lorsque nous voyons qu'ils nous ressemblent. Un aspect essentiel du test de Turing, c'est qu'il élimine les préjugés fondés sur les apparences ; il suffit de le réussir pour se voir attribuer une intelligence, même si le gagnant est une boîte montée sur roulettes. Bref, si cette réponse possible au problème des autres pensées a un mérite, c'est de s'appliquer clairement aux revendications de Turing. Si nous attribuons l'intelligence à d'autres humains en nous fondant sur des critères purement comportementaux, il semble que la simple cohérence nous oblige à l'attribuer aux machines lorsque nous constatons un comportement mécanique approprié, comme lorsqu'elles réussissent un test de langage tel que le test de Turing.

Selon la position progressiste, refuser d'attribuer une intelligence à une machine alors qu'on l'accorderait à un être humain s'apparente au sexisme ou au racisme - une discrimination infondée. Celle-ci repose sur les traits d'un individu, qui ne sont pas pertinents en la circonstance. Le fait qu'un individu soit une boîte montée sur roulettes peut toutefois s'avérer pertinent dans le cas qui nous occupe. Si vous êtes l'interrogateur dans le jeu de société qui a inspiré le test de Turing, et que la femme imite à merveille les réactions de l'homme. Ce seul comportement légitime que vous la preniez pour un homme. Mais si vous recevez de nouveaux témoignages (si vous entendez sa voix, par exemple), il vous faudra revenir sur cette conclusion. De même, selon les conservateurs, la réussite d'un individu au test de Turing légitime que vous lui attribuiez une intelligence - mais, si vous constatez que c'est une machine, il vous faudra revenir sur cette attribution.

Cela ne nécessite pas de revenir sur la conclusion selon laquelle l'intelligence est responsable du comportement des machines, mais remet en question le lieu de cette intelligence. Supposons que l'individu avec qui vous avez conversé au cours d'une soirée aux chandelles se révèle être un mannequin de cire équipé d'un récepteur et transmetteur radio, d'un locuteur et d'un microphone. Certes, la production verbale du mannequin est intelligente, mais vous réviserez à juste titre votre opinion sur le mannequin lui-même, en estimant que son intelligence vient de l'extérieur. De même, le

conservateur peut affirmer que, découvrant que vous avez conversé avec un ordinateur, vous devrez revoir votre opinion, et resituer l'intelligence postulée à l'extérieur de l'ordinateur - chez le concepteur du programme, par exemple.

La question est à présent joliment simplifiée. Le progressiste affirme qu'on ne doit pas contester son intelligence à un gagnant du test de Turing simplement parce que c'est une machine, un ordinateur par exemple, programmée adéquatement. Le conservateur pense le contraire. Mais pourquoi ? Quelle est la différence entre un être de chair et d'os et une machine de silicone et de métal ? Il est temps d'examiner les diverses formes du conservatisme.

L'important n'est pas de gagner

*Lorsque, frappé de disgrâce par la fortune et le regard des hommes,
je déplore, solitaire, mon statut de paria ... ce ne sont que des
mots pour vous, ou est-ce que vous en saisissez le sens?*

Maddox à Data, dans « Ce qui fait l'homme »

La position conservatrice se réclame d'une différence entre les machines et les créatures de chair et de sang comme nous-mêmes. Mais il ne suffira pas ici d'une différence classique - elle doit elle-même créer la différence. Dans cette partie, je recenserai les défauts des machines allégués par les personnages de Star Trek. Dans chaque cas, on sous-entend que le défaut fournit une raison acceptable de traiter les machines d'une certaine façon - différemment des personnes. Nous aurons en tête deux questions à la lecture des accusations portées contre les machines: le défaut allégué est-il authentique ? Et, s'il l'est, fait-il une authentique différence ?

Voici une collection de citations persifleuses glanées dans la première série, en commençant par « De quoi sont faites les petites filles ? ». Le Dr Roger Korby, un scientifique, a disparu depuis cinq ans avant que l'Entreprise le localise sur Exo III. Lorsque l'équipe de mission découvre que les compagnons de Korby sur la planète sont des androïdes, ce dernier se montre rassurant: « Un androïde, c'est comme un ordinateur. Ça ne fait que ce que je programme. » L'androïde Brown informe Kirk que les habitants originels d'Exo II ont « substitué une culture mécanique à la liberté ». Korby dit de l'androïde Andrea : « Il n'y a aucune émotion en elle - aucun investissement affectif. Elle ne fait que répondre aux ordres. C'est un ordinateur entièrement logique. »

Dans « Le retour des Archontes », l'ordinateur baptisé Landru explique que Landru, le chef humain de Beta III, l'a programmé dans une tentative d'autoclonage. « Je suis Landru, dit l'ordinateur. Je suis lui. Tout ce qu'il était, je suis - son expérience, ses connaissances. » « Mais pas sa sagesse, réplique Kirk. Il peut vous avoir programmé, mais non vous avoir donné une âme. Vous êtes une machine. » Par la suite, lorsque Spock admire la prouesse mécanique que représente Landru, Kirk rétorque : « Mais [ce n'était] qu'une machine, M. Spock. Landru l'a programmé à l'origine avec toutes ses connaissances, mais il ne pouvait lui transmettre sa sagesse,

sa compassion, son entendement, son âme ! »

Dans «Retour à hier», des êtres composés d'énergie attirent l'Entreprise sur une planète anonyme où ils persuadent des membres d'équipage de leur prêter momentanément leur corps pour qu'ils puissent se construire des corps androïdes. Alors qu'il discute de son transfert imminent dans un corps androïde, Enoch (qui occupe le corps de Spock) dit à Thalassa (qui occupe le corps du Dr Ann Mulhall) que les mains androïdes « sont des mains qui ne sentent rien ». Thalassa/Mulhall demande à Sargon (qui occupe le corps de Kirk):« "Bien-aimé" -que pourra signifier ce mot pour une machine?» Et peu après, juste avant d'embrasser Sargon/Kirk, elle demande : « Des lèvres de robots pourront-elles faire ceci?» Plus tard, elle dit encore à Enoch/Spock en désignant un corps d'androïde: «Je ne peux pas vivre dans ce truc ! » '

Dans « Le dernier ordinateur», lorsqu'il doit remettre les commandes de l'Entreprise au super-ordinateur M5, Spock objecte à son concepteur, le Dr Richard Daystrom, que « cet ordinateur peut gérer des informations, mais uniquement celles qu'on y introduit», et Kirk intervient: « Il peut travailler mille fois, un million de fois plus vite qu'un cerveau humain, mais il est incapable de porter un jugement de valeur, il n'a aucune intuition, il ne sait pas penser ! » Par la suite, McCoy remarque : «La compassion, voilà une chose que les machines n'ont jamais possédée. C'est peut-être le seul élément qui donne aux hommes une longueur d'avance sur elles.» Et dans « Moi, Mudd », Scott dit à Kirk: « Capitaine, les androïdes et les robots, ben, ils n'ont tout simplement pas accès à une pensée indépendante et créatrice. »

Au XXIV^e siècle, c'est l'androïde Data qui attire l'attention des persifleurs - notamment celle du Dr Kate Pulaski, qui a pris la suite de Doc McCoy. Dans « Le bail du silence», elle fait part de ses doutes sur la capacité de Data à remplir ses devoirs. Par la suite, elle demande pardon à Data, et remarque : « Vos états de service stipulent que vous êtes vivant. Il nous faut l'accepter.» (Mais «accepterait s-elle les états de service de Picard s'ils stipulaient que Picard était mort?) Dans « Ce qui fait l'homme», elle observe Data qui perd au poker suite à un coup de bluff de Riker, et dit que le bluff exige de l'instinct, signifiant par là que Data ne peut pas maîtriser le jeu. Dans « Élémentaire, mon cher Data», nous apprenons que Data est un fan des histoires de Sherlock Holmes, et qu'il en est venu à les recréer sur le holodeck. Mais, au lieu de jouer toute l'histoire, Data (qui représente Holmes) saute systématiquement au dénouement et résout l'énigme. Pulaski affirme que c'est parce qu'il est une machine. « Vous apprenez par cœur, dit-elle. Pour vous, tout est affaire de mémorisation et de récitation.» Puis elle ajoute (au bénéfice de LaForge) : « Holmes comprenait l'âme humaine ... cette compréhension échappe à Data ... votre ami artificiel ne souhaite pas résoudre un mystère qu'il n'aurait encore jamais lu.» Dans «L'ami», lorsque Data propose d'être le « soutien moral» de Troi, qui est enceinte, Pulaski réplique : «f Le conseiller Troi requerra un contact humain, et non la main froide de la technologie.» Par la suite elle traite Data de «data» (une somme de données), et, lorsque celui-ci se récrie, elle rétorque: « Où est la différence ? » « Le premier est mon nom, le second ne l'est pas», répond Data. Pulaski demande alors d'une voix vibrante de sarcasme: « Est-ce possible? Parmi tous vos réseaux neuronaux, algorithmes, toute votre

heuristique, il est une combinaison qui correspond à l'orgueil froissé ? » Dans « Ce qui fait l'homme », Will Riker, avocat de Maddox dans l'interrogatoire destiné à déterminer si Data est ou non une personne, avance immédiatement un argument. Il commence par établir que Data est un artefact, conçu et construit par un être humain. Puis il demande à Data de tordre une barre d'acier trempé pour montrer la supériorité de sa force. Sur quoi il démonte le bras gauche de Data et fait observer à la cour qu'il s'agit d'une construction, d'une nature non organique. Enfin il débranche Data en tournant un bouton situé au milieu de son dos et annonce : « Pinocchio est cassé. Ses ficelles ont été tranchées. »

Les persiflages, dans La nouvelle génération, viennent pour la plupart de Data lui-même. Dans « Les enseignes du commandement », il rapporte ses progrès en violon dans les termes suivants : « Bien que je sois techniquement compétent, les autres musiciens disent que je manque d'âme », et dans le même épisode il informe le colon de sexe féminin Ar'drian (qui lui a fait part de son intérêt sentimental) qu'il n'a « aucune forme de sentiment ». Dans « Tin Man », le bétazoïde Tarn Elbrun demande à Data : « Ça vous ennuie que je n'arrive pas à lire dans vos pensées ? », et Data répond : « Peut-être n'y a-t-il rien à lire. Rien de plus que des mécanismes et des réactions algorithmiques. » Dans « Indices », LaForge se livre à un test diagnostique sur le « réseau positronique » de Data et s'inquiète à l'idée de lui faire mal. « Je ne ressens pas la douleur », lui dit ce dernier. Dans « Le jour de Data », il doit mener la mariée à l'autel lors du mariage entre Keito Ishiwaka et le lieutenant Miles O'Brien, et Riker lui demande s'il est nerveux. « Je ne peux pas devenir nerveux, monsieur », répond Data. Dans « Le rejeton », Data décrit au Dr Crusher sa relation avec sa « fille » Lal en ces termes : « Je puis lui consacrer une certaine attention, docteur, mais non de l'amour. » Enfin, dans « Le culte du héros », lorsque Timothy demande avec admiration à Data si les androïdes sont meilleurs que les humains, celui-ci répond : « "Meilleur" est un terme hautement subjectif. Je ne possède pas par exemple la capacité de ressentir les émotions, à l'instar des humains... Mon cerveau positronique est incapable de générer ces conditions. »

Dans « Assistance vitale » (PE9), le Dr Bashir envisage d'utiliser des « implants positroniques » artificiels pour remplacer certaines parties lésées du cerveau du chef religieux bajorien, Vedek Bareil, non sans douter : « L'un de mes professeurs à l'école de médecine disait que le cerveau a une étincelle de vie qu'on ne saurait reproduire, se dit-il. Si nous commençons à remplacer des parties du cerveau de Bareil par des implants, cette étincelle peut se perdre. » Bareil finit par refuser de survivre grâce aux implants ; s'il ne fait rien, il mourra, « mais il mourra comme un homme, pas comme une machine ! »

La majorité des persiflages vient de la bouche d'« unités automatiques » découvertes par le capitaine Kathryn Janeway et l'équipage du Voyageur dans « Prototype ». L'unité 3947 dit des humains que « leurs processus cognitifs sont imprévisibles ». Il veut dire que ces unités automatiques ont elles-mêmes un comportement entièrement prévisible - propos renforcé lorsque 3947 explique en ces termes la menace faite par 6263 de tuer tous les humains si l'ingénieur en chef

B'Elanna Torres échoue dans ses efforts à sauver les unités : « 6263 est une unité de commandement préconçue. Il ne fait que suivre son programme. » Et voici une citation finale de Torres, dans « Dégel » (V), lorsqu'il s'agit de maintenir un programme de réalité virtuelle dépendant du cerveau en utilisant un cerveau artificiel : « Ce ne serait pas la même chose, dit-elle. Une intelligence artificielle ne saurait en aucun cas remplacer les fonctions véritables du cerveau. »

Si les persifleurs ont raison, nous ne devrions pas nous laisser prendre par les apparences, et cet argument est intuitivement séducteur. Après tout, nous avons tous vu des robots, comme ceux de Disneyland, qui peuvent passer pour réels dans les circonstances appropriées. Mais personne n'imagine qu'une vie mentale riche se poursuive à l'intérieur de ces robots. La lumière peut être allumée, mais il n'y a personne à la maison ! Comment un supplément de mécanique et de programmation - pour produire un robot plus imaginatif - ferait-il la différence ? C'est une question à laquelle nous nous attaquerons en détail dans le chapitre suivant, où je ferai de mon mieux pour miner les objections des persifleurs.

Chapitre 3

Procréation

La Flotte a pour mission de chercher de nouvelles formes de vie ... en voici une.

Picard, désignant Data dans « Ce qui fait l'homme »

Le chapitre 2 s'achevait sur une liste des propos ordinaires tenus par les personnages de Star Trek au sujet des machines. Voici un catalogue des traits humains censés manquer aux entités construites artificiellement: libre arbitre, créativité, sentiments et sensations, pensée originale ou indépendante, intuition ou instinct, valeurs (moralité), conscience de soi et du monde, capacité à saisir la signification profonde d'un mot, âme, sagesse, compréhension, « étincelle de vie », capacité de compréhension et d'action outrepassant une programmation, origine adéquate, composition matérielle adéquate, desseins adéquats. De plus, on souligne que les machines peuvent être débranchées. Si un individu comme Data a une de ces failles, et s'il suffit d'une faille pour dire d'un individu s'il est ou non une personne, il en découle que l'individu n'est pas une personne. (Cela vaut mutatis mutandis pour déterminer s'il est intelligent.) Dans ce chapitre, j'examinerai ces failles pour déterminer si Data (ou d'autres machines) les présente réellement, et si elles lui interdisent de se qualifier comme un être intelligent ou une personne. Nous commencerons par quelques allégations montrant pourquoi, comment et par quel intermédiaire une entité artificielle peut devenir un être.

Origines

*Une civilisation de puces informatiques, c'est impossible.
Ils ont monté une usine à Dakar, au Sénégal ... j'ai pu observer sa construction.*

Dr Paul Stubbs, «Évolution» (NG)

Data a été conçu par un être humain qui entendait créer une personne artificielle; les exocomps dans « La qualité de la vie » ont été créés par un être humain - un certain Dr Farrallon - qui entendait créer de simples ingénieurs outils. Et alors?

Non seulement l'inventeur d'un artefact peut échouer dans ses desseins (en créant des objets qui ne fonctionnent pas, par exemple), mais il peut aussi inventer des objets qui dépassent son intention (en réussissant mieux qu'il ne pensait, ou à la suite d'une manœuvre accidentelle). De fait, s'il était vrai qu'un artefact ne peut accomplir d'action imprévue, bien des procès seraient évités. Data commet des erreurs que son inventeur n'avait sans doute pas prises en compte : il est bien dangereux de supposer les machines incapables de se tromper. Dans « Le code de l'honneur » (NG), Data fait tous ses efforts pour comprendre les techniques du one-man-show et raconte (mal) une blague sur un dyslexique qui dit «tarin» pour «matin». Lorsque LaForge demande à Data combien de blagues il a déjà tentées, Data bafouille: « En toncant les tarins ... » et LaForge éclate de rire. Data, lui, se montre un peu vexé: « Simple lapsus!», dit-il en s'excusant.

Dans « Élémentaire, mon cher Data », l'ordinateur du vaisseau crée un personnage de holodeck apparemment doué d'intelligence, le professeur Moriarty. Dans Le guide Star Trek des tâtillons de La nouvelle génération, le super-fan Phil Farrand suggère que, si Moriarty est intelligent, c'est que l'ordinateur est intelligent lui aussi. Ce serait juste si les produits ne faisaient que refléter les intentions de leurs concepteurs : ce qui signifierait qu'un ordinateur a des intentions. Mais je crois que Farrand songe à un autre principe. Un produit a souvent des propriétés dont son inventeur est dépourvu - sans quoi les hommes ne sauraient produire des objets transparents, comme le verre - mais Farrand imagine peut-être qu'une propriété complexe comme l'intelligence n'est produite que par un objet intelligent. Si c'est vrai, il faudrait admettre que nous sommes, nous aussi, le produit d'une intelligence - et c'est la première étape dans l'argumentation en faveur de l'existence de Dieu (ou des dieux). Néanmoins, ce principe ne m'apparaît pas particulièrement plausible; je ne pense pas que notre position nous autorise à affirmer qu'un être non intelligent ne saurait produire de l'intelligence, ni même qu'une non-personne ne saurait engendrer une personne.

Si ce principe était plausible, le fait que Data a été créé par une personne serait un argument en sa faveur. Mais l'argument de Riker dans « Ce qui fait l'homme » est que l'humanité du créateur de Data joue à son détriment. Parce que les humains ne sont pas assez intelligents ? Si les hommes en cette fin du xx- siècle ne sont pas assez intelligents pour construire des personnes artificielles, cela ne signifie pas qu'ils n'y arriveront jamais. (Supposons qu'au xv« siècle, quelqu'un déclare que, si Dieu peut créer des objets volants, les humains ne sont pas assez intelligents pour l'imiter. Transporté au xx- siècle et confronté à un avion, il raisonnera en ces termes : puisque cet objet a été construit par des humains, et que les humains ne sauraient construire d'objet volant, cet objet doit être, dans les faits, incapable de voler !)

Peut-être Riker veut-il dire que le problème ne porte pas sur le degré d'intelligence des humains, mais sur l'absence, chez eux, de certaines capacités divines? Dans son résumé des faits, le procureur général observe : « Tous, nous avons tourné autour de la question essentielle: Data a-t-il une âme ? » Je soupçonne que

notre penchant à dénier la qualité de personne à une machine repose sur la conviction que notre qualité, à nous, ne vient pas de notre nature matérielle, mais d'un élément supplémentaire, particulier ... l'étincelle de vie, l'âme. Notez bien qu'on ne saurait attribuer ou refuser la qualité de personne à une machine en s'appuyant sur cette seule conviction. Il faut recourir au postulat suivant lequel les machines n'ont pas d'âme (Kirk y adhère visiblement, vu ses propos dans « Le retour des Archontes»). À cela, je réponds: si Dieu vous a donné une âme, comment l'avez-vous acquise? Deux solutions, sans doute: soit Dieu, par une intervention miraculeuse, ajouta votre âme au mélange matériel (à la conception, par exemple), soit Il a conçu la nature de telle sorte que, au moment où le bon mélange matériel se forme, une âme s'y attache (ou l'anime, etc.). Mais ces deux méthodes laissent ouverte la possibilité de machines animées, elles aussi. Peut-être Dieu a-t-il tant aimé le robot du Dr Soong qu'Il lui a donné miraculeusement une âme; peut-être y a-t-il d'autres façons de produire un « bon mélange», et Soong en a découvert une par hasard. Après tout, si vous tenez tant que ça à l'âme humaine, le comportement de Data n'est-il pas une preuve irréfutable qu'il a, lui aussi, une âme?

D'un autre côté, si ce comportement ne suffit pas à prouver que Data a une âme, loin d'être une «question essentielle», la question de savoir si Data a une âme ou non devient un point de détail, comme le montre la procureur général: « J'ignore si [Data] a une âme, dit-elle, et j'ignore si j'en ai une!» Mais si le droit de Data à disposer de sa personne dépend d'un jugement sur son âme, la question de savoir si la procureur général (ou d'autres) partage ces mêmes droits dépend à son tour d'un jugement sur son âme à elle. Puisque la procureur doute de la réalité de sa propre âme, elle ne doit pas être certaine de sa propre qualité de personne. Ce n'est pas le cas, donc la question de l'âme n'est pas prioritaire.

Riker se donne beaucoup de mal pour démontrer que Data n'est pas un être organique: peut-être est-il impossible qu'il existe des êtres inorganiques? Si c'est le cas, non seulement les personnes artificielles sont une impossibilité, mais cela vaut pour des personnes inorganiques engendrées normalement. C'est pourquoi la Horta dans « Démon dans le noir», et la substance cristalline dans « Sol natal» ne sauraient être des personnes. Pouvons-nous conclure, par principe, à la non-existence de personnes inorganiques? Non - dénier cette possibilité est pur chauvinisme (ou plutôt, pur «organicisme»). C'est indubitablement l'argument de Picard dans « Datalore », lorsqu'il dit aux officiers supérieurs : «S'il est gênant de s'entendre rappeler que Data est une machine, rappelez-vous que nous ne sommes qu'une variété de machines parmi d'autres - de nature électrochimique.» (Mais Data est lui aussi, semble-t-il, de nature électrochimique ; je suppose que Picard sous-entend « de nature organique et électrochimique».)

Que dire du bouton interrupteur dont Soong a pourvu Data, et dont Riker fait usage de façon convaincante ? Tout d'abord, les êtres humains ont aussi, semble-t-il, un interrupteur - le pinçon vulcanien les plonge dans un état remarquablement similaire. Par ailleurs, lorsque Data est éteint, il est peut-être dans un état analogue à notre état inconscient, comme lors d'une anesthésie. Enfin, si l'on suppose que les

opérations de Data s'interrompent entièrement lorsqu'il est débranché ... sommes-nous si certains que la qualité de personne n'est pas intermittente? Nous ne saurions affirmer que c'est impossible.

Jusqu'ici, je n'ai pas avancé d'argument positif en faveur de l'intelligence des machines, mais je n'en ai nul besoin. Rappelez-vous comment nous avons procédé. La solution au problème des autres pensées est de considérer un comportement complexe, intéressant, flexible et adapté chez les autres humains comme un motif suffisant pour leur conférer l'intelligence. La cohérence exige que l'on adopte les mêmes critères comme preuves suffisantes de l'intelligence chez les machines. Tout le poids de la démonstration revient donc aux persifleurs, qui doivent produire des raisons impérieuses de principe pour refuser d'appliquer ces critères aux machines. De plus, chaque fois qu'un parti affirme que quelque chose est logiquement impossible et qu'il est contredit par son adversaire, il pose une hypothèse bien plus forte, et doit donc assumer le poids de la démonstration.

La conscience de soi

*« Mes droits », « mon statut », « mon libre choix », « ma vie » ...
ces rennes me semblent témoigner d'une assez bonne conscience de soi.*

Picard à Maddox, citant les propos de Data dans « Ce qui fait l'homme »

Selon Maddox, Data n'a pas la conscience de soi dont Picard est pourvu : « Parce que vous, vous êtes conscient de votre propre existence, de vos actions ... vous êtes conscient de vous-même, de votre ego. » Maddox croit visiblement que Data n'a aucune idée de lui-même, aucune conception de lui-même comme être continu, aucune intuition ou préférence sur le cours de son existence. Nous l'avons vu dans le chapitre 1: s'il est nécessaire de posséder une conscience de soi pour être qualifié de personne, ce n'est pas une condition nécessaire pour l'intelligence. Il y a des tas d'espèces intelligentes sur terre qui ne semblent pas avoir d'idée de soi. Les psychologues des animaux pensent que les primates supérieurs ont bel et bien une idée de soi; si vous peignez une tache sur la figure d'un chimpanzé et qu'il se regarde dans un miroir, il essaiera de froter le miroir à l'endroit de la tache, en prenant le reflet pour lui-même. Aucune autre espèce n'a encore agi de la sorte dans une situation expérimentale. Par exemple, on peut apprivoiser une perruche en mettant un miroir dans sa cage: elle croira qu'une autre perruche lui tient compagnie. Un chat ou un chien ordinaire n'ont sans doute pas conscience d'eux-mêmes: ils réclameront à manger, mais leur contenu mental ressemble plus à : « Manger ici maintenant » qu'à « J'ai faim ».

Data peut donc toujours être intelligent, même si Maddox pense à juste titre qu'il n'a pas conscience de lui-même. Encore que ... Une fois de plus, observons le comportement de Data. Quels résultats aurait-il aux tests pratiqués par les psychologues animaliers? Je reconnais que Data, de toute évidence, identifie son reflet dans le miroir. Picard pratique une forme de test comportemental, lui aussi -

une variation sur le test linguistique. Il demande à Data ce qu'il fait à l'instant même. Data réplique: « Je participe à un interrogatoire légal qui déterminera mes droits et mon statut. Suis-je une personne ou un objet de propriété ? » Picard demande : « Et quel est l'enjeu du débat? » Data répond: « Mon libre arbitre, et peut-être ma vie. » Dans sa réponse à Maddox (cité dans l'exergue), Picard souligne l'utilisation par Data de pronoms possessifs à la première personne, «mes», «mon», «ma». (Sans citer le «Je» de la réplique citée ci-dessus.) L'argument judiciaire de Picard, c'est que ces pronoms prouvent sans contredit que Data possède une conscience de soi. Il semble avoir une idée de ce qu'il est, un sens de sa continuité temporelle, des croyances et préférences sur le cours de son existence (notamment par son refus de subir le processus de démontage).

Mais ce serait affreux si ...

Les ordinateurs sont d'excellents serviteurs, très efficaces.

Mais je n'ai aucune intention de les prendre pour maîtres.

Spock à Kirk, « L'ultime ordinateur »

Un androïde, c'est comme un ordinateur.

Ça ne fait que ce que je programme.

Roger Korby, « De quoi sont faites les petites filles ? »

Une autre objection ordinaire à l'idée que des machines sont des personnes est à peu près la suivante: s'il existait des populations de machines, ce serait affreux ! Les humains perdraient leur travail, un ordinateur serait élu président, etc. Spock note dans « Le retour des Archontes » que « la société sur Beta III suit une conception de la perfection qui serait celle d'une machine : paisible, harmonieuse ... » ; sur quoi Kirk l'interrompt :

« Mais sans âme ! » Toutefois, qu'une société qui inclut une population de machines ou est régie par elle soit ou non une bonne chose, cela ne tranche en rien la question de savoir si elle est possible. Turing lui-même a connu cette préoccupation, mais il l'a qualifiée de tactique de l'autruche, et il a eu raison de la rejeter pour son manque de pertinence.

Mais si l'on met à part la question de l'obsolescence humaine, pourquoi se soucier de produire des gens artificiels ? J'admets qu'on redoute « la substitution d'une culture mécanique à la liberté », pour citer l'androïde Brown dans « De quoi sont faites les petites filles ? ». Nous avons tendance à considérer les machines comme des objets programmés et non libres, dépendants, stériles, alors qu'on peut dire, en restant parfaitement raisonnable, que les ordinateurs et les androïdes sont créatifs et indépendants. Tout système fondé sur des règles de logique sophistiquée manifestera sa systématisme sa capacité à combiner les symboles pour produire de nouveaux réseaux, activité créative s'il en est. Lorsque Picard écoute Data jouer du violon dans « Les enseignes du commandement », il souligne que, même si Data a copié

les styles de deux virtuoses, il a produit quelque chose d'authentiquement original. Data avoue : «Je suppose que j'ai appris à être créatif, monsieur, lorsqu'il le faut. » Les machines peuvent aussi être indépendantes, dans le sens où, une fois mises en marche, elles peuvent poursuivre toutes seules les opérations. Tout système automatisé est conçu pour être indépendant dans ce sens. Ayant reçu un ordre, Data pourra l'exécuter tout seul.

La créativité est une chose curieuse. Pour ceux d'entre nous qui ne sont pas créatifs, elle paraît presque magique, divine. Mais aux yeux de bien des créatifs, c'est juste une tâche à accomplir (ce qui ne signifie pas qu'elle ne soit pas comblante). Pourtant, le mythe de l'étincelle divine a la vie dure. J'ai assisté récemment à une conférence du célèbre spécialiste en informatique Douglas Hofstadter, où il présentait les résultats d'un programme de composition musicale élaboré par le compositeur David Cope. Le programme intègre à l'entrée diverses pièces écrites dans un style particulier par un compositeur donné, et produit à la sortie un morceau original du même style. Si la pauvreté de certaines compositions d'auteurs décédés vous afflige, nous pourrions peut-être vous inventer des morceaux inédits ! Cela peut sembler grotesque, mais Hofstadter a fait jouer à un pianiste des morceaux deux par deux, et a ensuite demandé au public de désigner le morceau original. Une des séquences comparatives a bluffé les deux tiers du public - l'un des morceaux était une invention à deux voix de Bach, l'autre une variation « à la Bach » élaborée par le programme. Hofstadter déplora cette erreur de jugement: apparemment, il est convaincu - ou espère - que la reproduction de la grande musique ne saurait être si facile. Si la grande musique peut être reproduite par un programme informatique relativement basique, il est probable que nous n'avons pas à investir de grandes ressources mentales dans cette tâche. Pour ma part, je ne trouve cela ni affligeant ni alarmant. Considérant la division mentale du travail, je ne serais pas surpris que la partie du cerveau ayant fonction de « générateur de musique » soit simplement un outil - unique ou combiné - voué à cette tâche, et qui opère essentiellement hors de portée de la conscience (quelle que soit l'opération - je doute qu'il s'agisse de créer une musique toute faite). Sans doute l'aspect le plus ardu du métier de compositeur est-il de produire le type de morceau qu'on attend à la sortie. La conscience, certes, a introduit certaines données dans le cerveau, mais elles sont assez limitées. Après tout, c'est, en gros, la façon dont notre cerveau agit dans tous les autres domaines, ou presque (lorsqu'il produit des énoncés verbaux, par exemple), pourquoi en irait-il autrement avec la production musicale? Pourquoi la relative simplicité du « générateur de musique » changerait-elle quoi que ce soit (à part faire exploser le mythe de l'étincelle divine, ce qui est plutôt sain) à notre façon d'apprécier le résultat final?

Produire la responsabilité

Sans libre arbitre, il n'est pas de créativité.

Kirk, « Le retour des Archontes »

*Votre passé ne saurait justifier un
comportement immoral, monsieur*

Data à Fajo, «Joujoux»

Incontestablement, les machines sont largement capables de créativité et d'indépendance. Mais je doute que les persifleurs se laissent toucher par ces observations : à mon sens, ils donnent aux termes «créatif» et «indépendant» un sens différent, très proche de l'idée de libre arbitre, et opposé, semble-t-il, à l'idée de programmation. Mettons ce sens à l'épreuve. Dans « l'œil de l'esprit », LaForge est kidnappé par les Romuliens, et, une fois relâché, il tente d'assassiner le gouverneur Klingon qui est l'invité de l'Entreprise-D. LaForge a été conditionné dans ce but, conditionnement déclenché par son viseur sous l'impact d'une onde mauvaise. En d'autres termes, LaForge a été programmé de façon à recevoir une entrée donnée (l'onde) qui lui ferait produire une certaine sortie (son attentat sur la personne du gouverneur). Lorsqu'un comportement est ainsi conditionné, il n'y a pas de place pour le choix, l'indépendance ou la créativité; qui plus est, LaForge n'est pas à blâmer. Parce qu'il est programmé, il n'a aucun choix; son acte ne relève pas du libre arbitre, et dès lors il n'en est pas moralement responsable.

C'est tout le problème d'une « société mécanique », et des inquiétudes qu'elle suscite chez nous. Nous imaginons que les individus vivant dans cette société sont programmés sans libre arbitre, sans indépendance, sans responsabilité morale. Mais étudions les choses d'un peu plus près. En quoi l'action de LaForge ne relève-t-elle pas du libre arbitre? Tout d'abord, c'est un autre qui l'a conçue: LaForge est une marionnette impuissante entre les mains des Romuliens. Ce sont eux (et le traître Klingon, qui a tout orchestré) qui sont à blâmer pour cette agression ; ils en portent la responsabilité causale, et par extension la responsabilité morale. Puis, lorsqu'il est « manipulé », LaForge ne peut faire autrement que d'obéir aux diktats du programme. Pour résumer: pour qu'une action soit libre, elle doit être conçue par l'agent, et celui-ci doit avoir fait un choix, doit avoir eu la possibilité d'agir autrement dans les mêmes circonstances.

Lorsque nous appliquons ce raisonnement à Data, il achoppe sur l'un et l'autre point. Puisque Data a été créé et programmé par le Dr Noonian Soong, c'est sûrement le Dr Soong qui devrait être loué ou blâmé pour les actions de Data. Toutes les médailles de la Flotte remises à Data appartiennent en fait au Dr Soong. De plus, regardez ce qui se passe dans « Le jouet par excellence ». Data est kidnappé par Fajo, un trafiquant qui prend plaisir à collectionner des objets uniques en leur genre. Lorsque Fajo demande à Data s'il a déjà tué, celui-ci répond: « On a implanté en moi un respect fondamental pour la vie sous toutes ses formes, et une répugnance profonde à faire du mal à des êtres vivants. » Par la suite, Data tente de s'échapper et pointe sur Fajo son propre disrupteur, sur quoi le trafiquant lui rappelle ce pour quoi il a été programmé, convaincu qu'il ne saurait tirer. Mais il a oublié un autre détail mentionné par Data: « On m'a implanté la capacité d'infliger la mort pour mon

autodéfense.» Lorsque Fajo refuse de libérer Data, celui-ci appuie sur la détente ... et se retrouve téléporté à bord de l'Entreprise-D au même instant. Sans quoi, il aurait tué Fajo. Et qui aurait été responsable de cette mort? assurément, le créateur et programmeur de Data, le Dr Soong, qui a préconçu toute action accomplie par Data. Et rien dans cette situation ne laissait prévoir que Data pourrait faire autre chose que ce pour quoi il a été programmé. Fajo aurait eu raison de penser que Data était incapable d'appuyer sur la détente si son programme avait inclus un interdit absolu contre le meurtre. Au lieu de quoi le programme de Data incluait la possibilité de tuer pour son autodéfense, et, lorsque celui-ci s'est trouvé dans une position où il pouvait se défendre, il a dit simplement: « Je ne puis laisser continuer ainsi les choses. » De sa propre bouche, Data a admis que les circonstances ne lui laissaient pas le choix. Il semble que nous ayons décelé une véritable déficience chez Data et d'autres machines programmées: elles sont nécessairement privées de libre arbitre, et donc ne sont pas moralement responsables de leur comportement. Et puisque les machines ne sauraient être des agents moraux, elles ne sauraient être des personnes - ce qui ne signifie pas qu'elles ne puissent pas être intelligentes. Maddox a donc en partie raison.

Du moins est-ce la conclusion de cet argument. Le gros problème réside dans le postulat de base - suivant lequel les actions libres sont celles qui sont conçues par l'agent, et sont telles que l'agent aurait pu agir autrement. Je pense que soit ces deux conditions sont fausses (quand on prend les choses par excès), soit elles peuvent être réinterprétées (quand on les prend par défaut) de façon à reconnaître un libre arbitre aux machines. Il est plutôt bon qu'elles pèchent par excès: si elles avaient raison, plus aucune action humaine ne rerverrait du libre arbitre. Dans l'interprétation par excès, la première condition de l'action libre, c'est qu'elle doit naître chez l'agent, et la seconde, c'est que l'agent aurait pu agir différemment dans les mêmes circonstances. Les actions de Data ne remplissent pas ces conditions, mais elles ne valent pour aucune action humaine - si nous acceptons ce que j'appellerai l'image causale du monde (abrégée ICM).

L'ICM peut être définie en six mots: tout événement a une cause suffisante. La simple image causale de la façon dont fonctionne le monde apparaît comme une présupposition de notre pensée tout entière. Je suis assis ici à taper, des lettres apparaissent sur l'écran, mais non pas par magie. Je tape « q » et, abracadabra ! « q » apparaît sur l'écran. Vous, je ne sais pas, mais moi, je n'ai de mon ordinateur qu'une compréhension très limitée. Quand même, il y a une chose dont je suis certain : une pression suffisante sur la touche « q », une vision correcte du mode de fonctionnement de l'ordinateur, et le fait que celui-ci est en bon état de marche suffisent à expliquer pourquoi la lettre « q » apparaît. Je ne sais pas quelle est l'explication, mais je suis certain qu'il y en a une (en principe). De plus, ma certitude ne se restreint pas aux événements qui constituent les opérations d'ordinateurs. Pour chaque événement, on peut trouver, j'en suis sûr, quelque chose, une certaine combinaison de choses, qui en décrit l'origine. (Même un événement magique, s'il existe, doit bien avoir une cause.) Après tout, s'il n'y avait pas une combinaison de

choses qui suffise à produire l'événement en question, celui-ci ne se serait pas produit, n'est-ce pas? Que nous soyons en position de donner une explication complète ou non, le fait est que nous le faisons parce que nous présumons la vérité de l'ICM.

Deux points restent à clarifier. Tout d'abord, si notre monde n'admet pas qu'on puisse remonter le temps, qu'il n'est aucune chaîne causale inversée (voir le chapitre 6), il semblerait que tout événement soit précédé d'une cause suffisante. (Vous aurez moins de mal à apprécier l'argument qui va suivre si vous considérez l'ICM sous cet angle.) Ensuite, oubliez vos préoccupations antérieures sur « l'origine des choses ». Certains penseurs ont estimé qu'une chaîne causale doit être étayée par une « cause première» (Dieu, peut-être), mais il est tout aussi plausible d'estimer qu'elle peut être infinie. D'un côté ou d'un autre, il semble que notre pensée ordinaire ait pour présupposé que l'ICM vaut pour tous les événements dans cette partie de l'histoire qui nous permet de juger de la nature des actions humaines. Tous les événements qui nous entourent ici et maintenant ont une cause suffisante, et ceci nous permettra d'édifier-l'argument suivant.

Les actions humaines prennent place dans le monde parmi d'autres événements, et nous les expliquons de la même façon. Tout comme une pression sur la touche « q » explique (en partie) l'apparition de « q » sur l'écran devant moi, le fait que j'enfonce mon doigt dans votre œil explique (en partie) le fait que vous hurliez « Ouille ! » (Une explication complète nécessiterait bien sûr des détails supplémentaires sur votre état personnel, et les principes sur lesquels vous fonctionnez.) Tout comme le fait que j'appuie sur une touche n'est pas l'unique facteur externe à l'ordinateur qui suscite l'apparition du « q » (songez à toutes les causes par le moyen desquelles l'ordinateur lui-même a atteint l'état dans lequel il était avant même que je presse la touche), de même cette pression de mon doigt n'est pas l'unique cause externe de votre hurlement. Tout événement qui se produit en votre for intérieur est déterminé par une chaîne causale suffisante, qui est rattachée au monde extérieur. Supposons qu'en plus de hurler «Ouille !», vous me boxiez le nez. Pourquoi m'avoir boxé, quand un autre se serait retenu? Pourquoi une telle disposition à la violence? Nos explications ordinaires remonteront au moins partiellement à votre environnement - peut-être avez-vous été battu au cours de votre enfance - et ce genre de causes justificatives est clairement externe. De même, le comportement pacifiste des autres sera expliqué en partie par leur éducation relativement civilisée - une fois de plus, on ira chercher des causes externes à l'agent.

Je suis sûr que vous commencez à voir où est le problème. Une explication complète de la moindre de vos actions ne requiert pas moins que toute l'histoire de toutes les causes dans l'histoire de l'univers. En termes plus modestes: pour expliquer entièrement une action, il nous faudrait prendre en compte votre constitution génétique, vos principes de développement, l'histoire de votre environnement, incluant des entrées récentes comme le coup reçu dans l'œil. (Sans compter la question de savoir pourquoi vos parents vous ont élevé comme ils l'ont fait...) N'arrive-t-il donc jamais que vos actions soient considérées comme libres de toute influence extérieure, qu'elles aient leur origine causale en vous ? Jamais. Ni celles des

autres. Les actions humaines ne sont jamais libres, parce que les premières conditions de l'action libre ne sont jamais remplies. Si l'ICM est authentique, les actions humaines ne sont jamais libres de causes externes. Les agents n'en sont jamais les seuls auteurs, les causes originaires.

De plus, même s'il arrivait que les actions humaines remplissent, en certains cas, la première condition d'une action libre, il est tout à fait improbable qu'elles rempliraient la seconde. Nous pouvons voir ceci en soulignant que, suivant l'ICM, chaque événement a une cause suffisante. Regardez encore une fois l'événement au cours duquel vous m'avez boxé le nez que nous appellerons N. C sera la combinaison d'éléments (le coup de doigt dans l'œil, votre éducation, votre emploi du temps ce jour-là, etc.) ayant suffi à produire N. Puisqu'il suffit de C pour produire N, si on obtient C, N se produit nécessairement. Vous m'avez boxé le nez, et vraiment vous n'auriez pu réagir autrement dans ces circonstances puisqu'elles suffisaient à produire cette action. (Arrivés ici, on peut noter que rien de ce que j'ai revendiqué dans cette section ne dépend de la perspective suivant laquelle les actions humaines sont purement physiques ou matérielles de nature. Le schéma explicatif que j'ai proposé - qui présuppose l'rcsr - est employé également par des matérialistes et des non-matérialistes.) Et puisque toute action humaine a une cause suffisante, jamais l'agent n'aurait pu faire autrement dans ces circonstances, et aucune action humaine n'est libre. Le libre arbitre et la responsabilité morale sont illusoire si l'on interprète par excès les conditions d'une action libre. (En fait, la responsabilité morale pourrait exister : si Dieu existe, et qu'il est sa propre cause, il est moralement responsable de tout ce qui se produit, y compris le mal.) Data a tort d'affirmer que le passé misérable de Fajo n'excuse pas son comportement (voir l'exergue). Même si Fajo n'avait pas été misérable dans le passé, celui-ci n'en aurait pas moins excusé son comportement. Si le comportement de Data est programmé, cela vaut aussi pour Fajo, et pour vous !

Nous touchons ici à une célèbre question philosophique, le problème du libre arbitre, souvent posé en des termes différents de ceux que j'ai utilisés ici. On appelle déterminisme la thèse selon laquelle chaque état de l'univers détermine entièrement chaque état suivant. Pour illustrer cette thèse, figurez-vous une intelligence infiniment puissante (Dieu, par exemple) à qui n'échappe aucun fait sur l'état présent de l'univers. Cette intelligence connaît tout ce qu'il y a à savoir sur l'état présent de l'univers, y compris ses principes de fonctionnement. Si le déterminisme existe, alors cette intelligence pourrait infailliblement prédire l'état suivant en se fondant sur cette connaissance. Par exemple, pour des êtres limités comme nous, le résultat d'un pile ou face (sans tricher) est une proposition (cinquantecinquante), mais c'est parce que nos connaissances sont limitées. Figurez-vous que nous lançons une pièce en l'air, et que nous figions son trajet lorsqu'il a atteint son sommet. Si Dieu connaît la masse et l'orientation exactes de la pièce à ce point, sa distance exacte de tous les autres corps (et notamment de la terre), le nombre et la position des molécules d'air qui la heurteront au cours de sa chute, tous les détails sur la surface où elle atterrira, etc.,

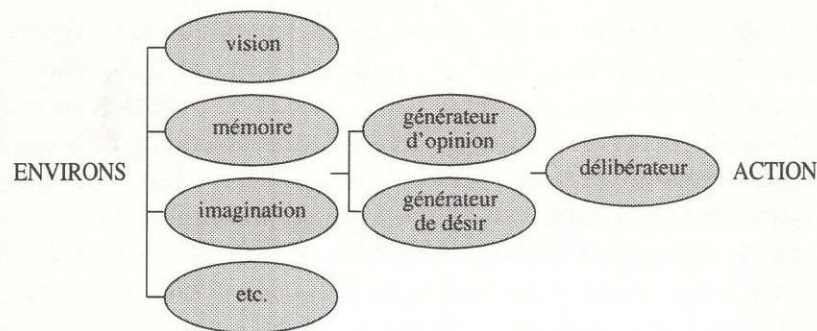
eh bien, si le déterminisme est chose authentique, Dieu pourra infailliblement prédire le résultat du jet.

Toutefois, notre science nous apprend que le déterminisme n'existe pas. Il y a une indétermination quantique qui, en ce qui nous concerne, revient à attribuer au hasard un rôle fondamental dans le fonctionnement du monde. Le hasard est cousu dans l'étoffe même de l'univers que nous habitons, et même Dieu ne saurait prédire infailliblement tout ce qui doit se passer. L'indéterminisme existe. Mais cela ne change rien à la conclusion: nous ne sommes pas libres. L'ICM ne se confond pas avec la thèse du déterminisme. Si l'indéterminisme existait, l'ICM existerait, mais ce raisonnement ne vaut pas pour l'inverse. L'ICM est simplement un terme neutre entre le déterminisme et l'indéterminisme. Si l'indéterminisme est vérifié, alors l'univers ne progresse pas de façon déterministe en états subséquents - mais il le fait. Chaque événement n'en a pas moins une cause suffisante, mais le hasard est fondamentalement inséré dans la combinaison de choses qui suffisent à produire l'événement. Nous aurions donc pu définir ainsi l'ICM: chaque événement a une cause suffisante (déterministe ou non).

On pourrait toutefois objecter que l'argument qui déduit de l'ICM l'absence universelle de liberté est sans pertinence. Peut-être la liberté n'est-elle menacée que par une causalité déterministe? Nous pouvons renoncer assez rapidement à cet espoir. Pour citer le philosophe Daniel Dennett, la variété de « libre arbitre » produite par l'indéterminisme est assez piètre. Si certaines de vos actions sont authentiquement non déterminées à cause d'événements causés par un hasard quantique dans votre cerveau, il semble que ce soit là un mince réconfort. Si le déterminisme fait de vous un simple pantin mécanique, agité suivant des mouvements prévisibles, être mû arbitrairement fera-t-il moins un pantin de vous ? Et en quoi le « libre arbitre » né de l'indéterminisme nous conférerait-il une responsabilité morale? Loin de considérer le hasard comme une entité rendant les agents moralement responsables de leurs actes, nos intuitions vont dans le sens inverse - l'intervention du hasard semble éliminer, ou du moins atténuer, la culpabilité morale. (Une question intéressante, mais secondaire, est de savoir si un quantum d'indétermination peut se traduire en indétermination macroscopique - à savoir, si les effets quantiques ne sont pas tout simplement annulés à l'échelle d'événements mentaux.)

Puisque l'interprétation par excès des conditions de l'action libre montre qu'aucune action humaine ne l'est vraiment, je pense que nous devrions l'exclure. Quant à la première condition, retournons à la « programmation » de LaForge. Quelle est la différence entre son attentat et toutes les autres actions qu'il peut accomplir - librement, à notre sens ? Pour répondre, il nous faut un schéma grossier du fonctionnement de l'agent cognitif - et donc du cerveau. Considérez le cerveau comme un outil fonctionnel, composé, dans la division mentaliste du travail, d'unités fonctionnelles de plus en plus réduites, de diverses facultés et sous-facultés. Qu'est-ce que la volonté, cet élément qui soutient le libre arbitre ? Simplement une faculté mentale, dont le but est de produire l'action (et non le comportement dans son entier). Nous pouvons diviser la faculté de volonté en trois sous-facultés

fondamentales : un générateur d'opinion, un générateur de désir et un délibérateur. Voici un schéma simplifié de la volonté, qui produit de l'action à la sortie, ayant reçu à l'entrée certaines informations des autres facultés et de l'environnement.



Pour illustrer une opération mentale, supposez que vous vous promeniez sur le trottoir et que vous trouviez sur votre chemin un morceau de chewing-gum. Votre faculté visuelle transmet cette information, et vous formez l'opinion qu'il y a devant vous, sur le trottoir, un morceau de chewing-gum. Puisque vous vous rappelez combien il est collant et difficile à éliminer, vous formez le désir de l'éviter. Une combinaison d'autres facultés vous amène à former certaines opinions sur la façon de procéder (en contournant le chewing-gum, par exemple). Puis votre délibérateur jauge l'entrée fournie par les générateurs d'opinion et de désir (y compris d'autres opinions et désirs courants, et leurs poids respectifs) et produit, à la sortie, le comportement approprié. Si d'autres désirs ne prennent pas le dessus à ce moment, la combinaison d'opinions et de désirs, telle que je l'ai décrite, produira le comportement voulu: vous contournerez le chewing-gum. (Les « entrées environnantes » peuvent aussi venir de l'intérieur. Par exemple, lorsque votre estomac est vide, cette entrée suscite un désir de manger.)

Voici l'interprétation par défaut que je voudrais donner à la première condition d'une action libre. Pour être libre, une action doit être produite par la volonté - par le générateur d'opinion, le générateur de désir et le délibérateur, agissant de concert. La raison pour laquelle l'attentat de LaForge ne relève pas de son libre arbitre est qu'il ne remplit pas cette condition; il est contraire à son bon vouloir: il ne désire pas la mort du gouverneur Klingon. Le conditionnement a permis aux Romuliens de court-circuiter LaForge, en faisant fi de sa volonté; son comportement tenait plus du réflexe que d'une action authentique. De plus, nous pouvons opposer ce comportement dicté à une action libre, et appliquer les mêmes distinctions au comportement de Data. Dans «Frères» (NG), Data prend soudain le commandement de l'Entreprise-D, et met le cap sur une destination inconnue, en résistant à toutes les tentatives de l'équipage pour reprendre le contrôle. Il s'avère que son créateur, le Dr Soong, est encore vivant: il a activé chez Data un programme « retour à la maison » afin de poursuivre son travail sur lui. Voilà bien un comportement programmé, et, lorsque Data « se réveille », il ne se rappelle plus rien de ce qui s'est passé après le déclenchement du programme. Il se soucie de l'Entreprise-D, de l'équipage, de la mission qui leur avait été assignée.

Interpréter ces événements en faveur de Data, c'est dire qu'il est retourné « chez lui » contre sa volonté. Il voulait remplir sa mission, obéir aux ordres de ses supérieurs, et considérait qu'en rentrant « à la maison » il mettait des vies en danger (il y a à bord du vaisseau un enfant agonisant). Sa volonté a été court-circuitée par le Dr Soong et son programme, tout comme celle de LaForge est court-circuitée par les Romuliens. Si l'enfant était mort, le Dr Soong aurait donc été moralement responsable de son décès - son ignorance des faits aurait constitué une circonstance atténuante. Ce schéma pourrait décrire les opérations ordinaires de Data comme celles de LaForge, et la première condition pour l'action libre est donc remplie en ce qui le concerne.

L'interprétation par défaut de la seconde condition, c'est que l'agent doit être tel qu'il aurait pu agir différemment en des circonstances différentes. Peut-être une analogie nous aidera-t-elle. Prenez le sphex, dont le comportement est connu pour son invariabilité. Lorsqu'il pique une araignée pour la paralyser, il traîne sa victime jusqu'à un monticule, descend y arranger quelque chose, ressort et rentre en traînant l'araignée. Mais si vous interférez en déplaçant l'araignée, ne serait-ce que d'un centimètre, pendant que le sphex est à l'intérieur, lorsqu'il revient à la surface, il refait toutes les étapes de la routine, en redéplaçant le corps vers le « bon » emplacement avant de redescendre vérifier ses petits arrangements (et vous, pendant ce temps, vous redéplacerez l'araignée, etc.). Les philosophes rendent hommage au sphex en appelant sphexisme une réponse inflexible à différentes entrées environnantes. En voici diverses applications humaines. Un individu n'arrive pas à sortir de chez lui, parce que, lorsqu'il est à la porte d'entrée, il s'inquiète de savoir s'il a laissé le gaz allumé, retourne à la cuisine vérifier, regagne la porte d'entrée, retourne à la cuisine vérifier, etc. Une femme se lave les mains parce qu'elle les trouve sales, les relave parce qu'elle les retrouve sales, et ainsi de suite. Ce sont des comportements obsessionnels qui incluent la kleptomanie, l'anorexie mentale, la boulimie, la nymphomanie, etc. Pourquoi ? Parce que, quelle que soit l'entrée reçue, la volonté produit exactement la même sortie en matière de comportement. Si vous n'êtes pas un laveur de mains compulsif, vous ne vous les laverez que lorsqu'elles seront sales, ou en cas de doute. Un laveur de mains obsessionnel se lave les mains en boucle, qu'elles soient sales ou non, quoi qu'il fasse par ailleurs. Une interprétation par défaut de la seconde condition stipule qu'une action est libre lorsqu'elle ne répond pas à ce comportement: lorsque la volonté fonctionne normalement. Sans quoi, une ou plusieurs des sous-facultés mentales ne font pas leur boulot. Le générateur d'opinion doit être sensible à la façon dont le monde fonctionne ; si nos mains sont propres, il devrait nous inciter à considérer ... qu'elles le sont. Le générateur de désir devrait être sensible à tout ce qui est bon pour l'agent; si on se laisse lentement mourir de faim, le générateur de désir devrait susciter un désir de manger supérieur à celui d'être séduisant(e). Et il devrait être sensible au bon équilibre des opinions et desirs qu'il reçoit en entrée; si le générateur d'opinion induit l'opinion que la minceur n'est pas une compensation à la mort, cela devrait se refléter dans le comportement. Il s'avère que le libre arbitre n'a rien à voir avec une liberté à l'égard des causes -

bien au contraire. Puisque chaque événement a une cause, le libre arbitre n'a rien à voir avec l'absence de causalité. C'est la façon dont les actions humaines sont causées qui les rend libres ou non libres. À mon sens, l'adjectif «libre» appliqué à la volonté signifie exactement la même chose que lorsqu'on l'applique à d'autres outils fonctionnels, telles les roues de bicyclette et les girouettes. Une girouette est libre si elle fonctionne bien - si elle est capable d'intégrer le mouvement des molécules d'air à l'entrée et de produire une représentation correcte de la direction du vent à la sortie. Les dysfonctionnements d'un individu sphexiste ressemblent à ceux d'une girouette rouillée qui désigne la même direction quelle que soit celle du vent.

Bien sûr, les gens qui redoutent une société mécanisée craignent en fait le sphexisme. (Dans « Les proies» [NG], nous voyons les résultats de la programmation des soldats, et nous ressentons une sympathie profonde à leur égard, parce que nous voyons combien ils sont peu faits pour la paix, et combien ils sont impuissants à changer.) Mais il n'y a pas de raison, en principe, pour croire que les machines seront nécessairement sphexistes. Leur créativité est aussi capable de flexibilité et d'adaptation que celle des humains - en principe, du moins. Data est-il sphexiste? Cela n'apparaît qu'à une seule occasion. Dans «Indices», il répond à toutes les interrogations de Picard par « Je ne saurais répondre à cette question, monsieur.» Avec une exaspération croissante, Picard s'exclame : « On dirait que vous êtes pris dans une sorte de boucle solipsiste ! » Mais Data n'est pas le moins de monde coincé; il ne fait qu'obéir aux ordres de Picard (qui l'ignore, ayant subi un lavage de cerveau) et pense qu'en agissant autrement il condamnerait à mort l'équipage tout entier. Par ailleurs, Data manifeste précisément le comportement qu'on attendrait de la part d'une créature possédant une volonté en parfait état de fonctionnement (au pire, il est moins sphexiste que les autres membres de la Flotte). Picard le dit dans « Unification 1 » (NG): « Monsieur Data, vos ressources mentales ne cesseront jamais de m'étonner» ; et dans « Rédemption II», après que Data eut présenté ses excuses pour avoir désobéi aux ordres: « La Flotte ne veut pas d'officiers qui suivraient aveuglément les ordres sans analyser la situation. Vos actes étaient appropriés aux circonstances.» Data agit souvent librement, et son créateur, le Dr Soong, n'est pas moralement responsable de ces libres actions. Une fois de plus, ce qui paraissait être une différence authentique entre les machines et nous-mêmes s'avère ne reposer sur aucun fondement.

Une question de références

Dans « Ce qui fait l'homme», Maddox demande à Data si les vers des sonnets de Shakespeare « ne sont que des mots pour [lui] ou [s'ils] ont un sens», et dans « Retour à demain», Thalassa s'interroge: «"Bien-aimé" ... que signifiera ce mot pour une machine?» Maddox et Thalassa, semble-t-il, ne s'interrogent que sur le problème de savoir si une machine peut ou non comprendre des types particuliers de mots (dénnotant des sentiments, des affects), mais leur scepticisme nous rappelle l'une des plus célèbres objections à l'intelligence des machines (et, par extension, à leur statut

de personne). Suivant le philosophe John Searle et ses disciples, une machine faite de silicone et de métal ne saurait comprendre les mots, quels qu'ils soient. L'argument de Searle est connu sous le nom d'argument de la chambre chinoise. S'il a raison, alors son argument montre pourquoi les conservateurs devraient refuser d'admettre qu'un gagnant au test de Turing soit intelligent.

Searle imagine qu'il est enfermé dans une chambre dotée de deux fentes dans le mur. L'une est une fente «d'entrée», par où passent des morceaux de papier couverts de gribouillis étranges. Dans la pièce où est Searle se trouvent aussi des tas de crayons, des pages blanches et un manuel d'instructions assez massif. Il n'y a rien d'autre. Le manuel d'instructions contient certaines informations, rédigées en anglais, sur la marche à suivre pour réaliser certaines transformations, suivant un ordre donné, sur ces étranges gribouillis. Voici une de ces instructions: « Si la page comporte tel ou tel schéma, copiez cet autre schéma sur une page blanche et glissez-la par la fente de sortie.» Searle ignore, car la règle du jeu stipule qu'il ne parle qu'une langue, que les gribouillis sont en fait des idéogrammes chinois. En suivant les instructions, il produit des éléments de sortie qui non seulement ont un sens pour les « interrogateurs » chinois, mais peuvent être interprétés comme des réponses correctes aux questions qu'ils glissent par la fente d'entrée.

La chambre chinoise elle-même est en fait le «sujet» d'un test de Turing mené en chinois. Le premier point établi par Searle, c'est qu'il est possible, si le manuel d'instructions est précis et exhaustif, et que Searle suit correctement ses instructions, de faire passer la chambre chinoise pour un simple test comportemental, comme le test de Turing. Le second, c'est que Searle, enfermé dans la chambre, ne comprend pas un mot de chinois, mais que les interrogateurs chinois ont l'impression de mener une conversation rationnelle avec le sujet du test (la chambre). La première conclusion de Searle est celle-ci: puisqu'un ordinateur programmé pour réussir le test de Turing est exactement semblable à la chambre chinoise, le fait qu'il réussisse le test n'établit pas qu'il soit doué d'intelligence. Sa seconde conclusion, c'est qu'aucune instruction complémentaire en matière de gestion de données contenue dans le manuel ne réussirait à lui faire comprendre le chinois, et que par conséquent la gestion de données ne saurait suffire à établir une intelligence.

Pour nous répéter, l'expérience mentale, dite de la chambre chinoise est censée montrer qu'un test purement comportemental tel que le test de Turing ne suffit pas à diagnostiquer la présence d'une intelligence lorsque le sujet en est simplement un ordinateur bien programmé. De plus, il est censé établir le point négatif suivant lequel la pensée n'est pas une simple gestion de données. De quoi est faite la «pensée»? Suivant Searle, elle requiert l'intentionnalité, ou la référentialité. Supposons qu'on vous demande: « Quelle est la capitale des États-Unis?» et que vous répondez « Washington, D.C. ». Puisque vous comprenez l'anglais, vous comprenez que « les États-Unis » renvoient à une nation. Lorsque vous songez aux États-Unis (en vue de répondre à la question), votre pensée a pour référence les États-Unis. Mais supposez que nous posions cette question (en chinois, bien sûr) à la chambre chinoise, et que nous recevions la bonne réponse. Puisque Searle ignore que le «gribouillis» est

du chinois pour « les États-Unis », il est quasi certain qu'il ne songe pas aux États-Unis au moment où il produit la bonne réponse, et qu'en tout cas il n'y songe pas au moment d'accomplir les transformations nécessaires. Eh bien, un ordinateur qui « passe » le test de Turing, et qui répond à cette même question par « Washington, D.C. » ne se réfère pas non plus en pensée aux États-Unis. Suivant Searle, la pensée est intentionnelle- on pense en référence à quelque chose - et puisque l'ordinateur ne pense pas en référence à quoi que ce soit, c'est qu'il ne pense pas.

Searle reformule cet argument en disant que le traitement de l'information en lui-même est syntaxique et non sémantique. Comme nous l'avons noté dans le chapitre 1, la « syntaxe » renvoie aux propriétés d'un langage non interprété, et la « sémantique » à celles d'un langage interprété. Rappelez-vous le langage simple baptisé L. Voici un autre énoncé de L: « Cwu ».

Qu'est-ce que ça signifie, en français de tous les jours? Rien - tel quel, il s'agit d'une chaîne syntaxique non interprétée. Nous pouvons le pourvoir d'une sémantique en lui assignant une place dans un dictionnaire. Si « u » dénote un individu qui est les États-Unis, si « w » dénote un individu qui est la ville de Washington, D.C., et si « Cwu » ne vaut que si et seulement si w est la capitale de u, alors cette formule signifie que Washington, D.C., est la capitale des États-Unis. Mais il est important de réaliser que « Cwu » ne signifie cela qu'en référence au sens du dictionnaire. Une autre référence lexicale lui ferait signifier: « Walter a étouffé l'oncle Fred ».

L est gouverné par les règles transformatives de la logique de premier ordre. Par exemple, si A, B et C sont des variables pour la formule L, alors, étant donné :

$(A \& B) \sim C$ (si A est vrai et B est vrai, alors C est vrai) et C (C n'est pas vrai)
nous pouvons en inférer

Av B (soit A n'est pas vrai ou B n'est pas vrai) en deux étapes.

Peut-être savez-vous déjà comment procéder. Si c'est le cas, votre capacité à accomplir ces transformations est indépendante, non seulement de toute référence de la formule de L à « A », « B » et « C », mais aussi à toute signification assignée par le dictionnaire à A, B et C. Vous ne faites que manipuler des symboles formels, à la façon d'un ordinateur. De fait, vous pourriez apprendre ces règles en vous passant des références lexicales : vous seriez un peu comme Searle dans la chambre chinoise, un peu comme Data et Troi lorsqu'ils essaient de décoder le tamarien dans « Darmok ».

Mais, pourrait-on objecter, lorsqu'on reçoit un vocabulaire, on comprend ce que signifie la transformation de L. Pourquoi ne pas simplement donner à l'ordinateur un dictionnaire qu'il pourrait utiliser pour sa transformation ? Parce que, pense Searle, il y a plus à dire sur l'intentionnalité: notamment, il affirme que les ordinateurs (comme celui sur lequel je tape) ne sont pas d'une matière susceptible de posséder l'intentionnalité. Aux yeux de Searle, seules certaines structures biologiques sont capables de l'intentionnalité - elles seules ont les bons « pouvoirs de causalité », (Détail intéressant, Searle ne nie pas qu'il pourrait y avoir des formes de vie intelligentes, basées sur le silicone, comme la Horta, mais nous ne nous préoccupons pas de cela. Si l'argument de la chambre chinoise ne marche pas, le reste de sa pensée n'importe guère.)

Une objection répandue contre cet argument, baptisée par les philosophes et les scientifiques informaticiens la réfutation des systèmes, admet que Searle-dans-la-chambre ne comprend pas le chinois, mais souligne le fait que voici : nous ne dirons pas de ceux qui parlent le chinois qu'une partie d'eux parle cette langue - c'est l'individu tout entier, pris comme un système, qui comprend le chinois. Ceci vaut bien sûr pour la chambre chinoise et son contenu (y compris Searle), si on les prend comme un système. Puisque le système pris dans son ensemble reproduit le comportement d'un locuteur chinois compétent, c'est le système tout entier qui comprend cette langue. Searle réfute en deux points cette réplique des systèmes. Tout d'abord, il tient pour invalide la suggestion suivant laquelle le système comprend le chinois. Si un homme ne comprend pas le chinois, comment peut-il y parvenir une fois pourvu de crayons, de papier et d'un manuel? Ensuite, il adapte l'argument de la chambre chinoise en supposant qu'au lieu d'accomplir manuellement les opérations dictées par le manuel, Searle-dans-la-chambre les fait dans sa tête, et au lieu de consulter continuellement le manuel, mémorise toutes les règles. Il intériorise le système. Dans cette version modifiée, affirme Searle, il ne comprend toujours pas le chinois - qui lui apparaît toujours comme une série de gribouillis - et la réplique des systèmes ne tient pas.

Suivant une seconde objection - la réfutation du robot - il n'est pas surprenant qu'un ordinateur prisonnier, suivant le modèle offert par Searle, ne comprenne pas le chinois, puisqu'il ne possède aucun appareil sensoriel qui lui permette de connecter des vocables à des significations. (Cf. le cas « Darmok », où, pour comprendre une expression, Picard doit se trouver dans la situation à laquelle il est fait référence.) Toutefois, un ordinateur pourvu d'un tel appareil et de la capacité de naviguer dans un environnement - bref, un robot programmé à ces fins, tel Data - serait capable de comprendre le chinois. La réplique de Searle comprend une fois de plus deux parties. Tout d'abord, la réplique du robot commence par concéder que l'argument de la chambre chinoise démontre qu'un ordinateur prisonnier ne comprend rien: l'objection porte donc sur l'affirmation de Searle selon laquelle seule une structure biologique appropriée peut comporter l'intentionnalité. Searle riposte en imaginant, une fois de plus, qu'il intériorise tout - y compris l'équipement robotique, cette fois - et répète que tout cela ne l'aiderait pas davantage à comprendre le chinois.

Suivant la réfutation du simulateur de cerveau, un ordinateur qui simule non seulement le comportement linguistique d'un locuteur chinois, mais également son cerveau, y compris au niveau des impulsions neuronales, comprend bel et bien le chinois. (L'« ultime ordinateur », l'infortuné M5, est une simulation du cerveau de Richard Daystrom.) Une fois de plus, Searle répond en deux parties. Tout d'abord, il note qu'une telle affirmation ne s'accommode guère de la vision fonctionnaliste - qui prédomine dans le champ de la recherche en matière d'intelligence artificielle-, l'étude du cerveau n'est pas nécessaire à la constitution de l'intelligence. De plus, il propose d'adapter la chambre chinoise de façon à réussir le test en simulant la structure du cerveau par un système de conduites d'eau et de valves actionnées à la main. (Ces valves n'ont pas besoin de manettes «ouvert/fermé» comme les

ordinateurs digitaux: elles pourraient moduler le niveau d'eau.) Searle tient pour évident qu'on ne pourrait rien comprendre dans un tel système.

Jusqu'ici, tout va bien pour lui- en apparence. Mais creusons un peu plus, pour commencer par la réponse de Searle à la réfutation du simulateur de cerveau et à son joli tableau d'un système de conduites d'eau et de valves. Il est, de fait, difficile d'imaginer comment les choses sont réellement dans le cerveau, au niveau des impulsions neuronales, dans votre matériau de base organique. J'admets qu'il est non moins difficile d'imaginer comment pareil système pourrait comprendre quoi que ce soit. Assurément, les neurones individuels ne comprennent rien, et pourtant, lorsque tout est connecté, cela donne l'intelligence. Mon objection ne porte pas tant sur l'argument de Searle que sur sa méthode. Il imagine tout un champ de systèmes d'expérimentations mentales qu'intuitivement nous jugerions incapables de comprendre quoi que ce soit. Lui pense que certains systèmes purement matériels - le cerveau humain - comprennent certaines choses, et c'est non moins mystérieux dans le cas d'un cerveau que dans le cas d'un homme enfermé dans une pièce, ou dans le cas d'un système de conduites d'eau et de valves. Searle évacue peut-être un peu rapidement la réfutation des systèmes. Ne l'oublions pas, et passons à une ou deux objections plus substantielles.

La réfutation anglaise comprend deux parties. La première s'appuie sur la logique - notamment sur la loi du centre exclu qui affirme que, pour toute proposition donnée, soit la proposition est vraie, soit sa négation est vraie. Voyons comment l'argument de Searle est structuré. Il postule une stricte analogie entre le comportement de la chambre chinoise et celui d'un ordinateur programmé adéquatement qui réussit un test de Turing mené en chinois (nous l'appellerons l'ordinateur-C). Supposons que nous admettions l'argument de Searle qui veut que la pensée requière l'intentionnalité, et qui plus est, admettons son hypothèse suivant laquelle aucun élément constituant la chambre chinoise - pas même le système pris dans son entier - ne comprend le chinois. Si nous admettons encore la stricte analogie avec l'ordinateur-C, quelle est la conclusion qui s'impose? Que l'ordinateur ne comprend pas le chinois. Mais ce n'est pas celle que tire Searle. Il conclut que l'ordinateur ne comprend rien. Searle a utilisé un centre non exclu (ou non distribué, dirions-nous en bons philosophes). Voici un centre exclu: « Soit l'ordinateur-C comprend le chinois, soit il ne comprend pas le chinois. » Voici le centre non distribué de Searle: « Soit l'ordinateur-C comprend le chinois, soit il ne comprend rien. » Cela laisse clairement du champ à la possibilité médiane suivant laquelle l'ordinateur-C comprend quelque chose, mais pas le chinois.

Cela complète la première partie de la réfutation anglaise, qui montre que Searle n'a pas démontré ce qu'il affirme avoir démontré : c'est pourquoi l'argument de la chambre chinoise manque son objectif. La seconde partie de la réfutation anglaise tente de montrer qu'il pourrait se trouver quelque chose que l'ordinateur-C comprenne. Pour commencer, vous noterez que la chambre chinoise comporte un élément intentionnel, y compris dans la version qu'en donne Searle: Searle lui-même, qui maîtrise parfaitement la langue anglaise. Voici qui rompt toute analogie avec le cas

de l'ordinateur-C: essayons donc de corriger le tir. Manifestement, Searle est largement surqualifié pour cette tâche, et je veux donc que vous vous figuriez qu'on lui enlève peu à peu des parcelles de sa compréhension de l'anglais (il pourrait se passer de 99 % de son vocabulaire anglais) en ne laissant que le strict minimum nécessaire pour comprendre le manuel et exécuter ses instructions. Ce qui reste de Searle - nous l'appellerons Searle-C - est encore intentionnel, dans la mesure où il a des idées sur les symboles qu'il transforme, sur ce qu'il doit faire avec eux, etc.

Prenons maintenant l'ordinateur-C. Il réussit le test de Turing, et nous ne devons pas sous-estimer la complexité de la programmation, la capacité de mémoire, et la vitesse de traitement nécessaire à faire ceci. Quelle raison aurions-nous, en principe, de penser que Searle-C possède quelque chose qui manque à l'ordinateur-C? Si nous prenons au sérieux la stricte analogie entre les deux cas, comme Searle nous y incite, il n'est pas raisonnable de penser que, tout comme Searle-C a besoin de l'intentionnalité pour accomplir ses tâches, de même une partie de l'ordinateur-C a besoin de l'intentionnalité pour remplir les tâches similaires? N'aura-t-il pas besoin de se référer aux symboles sur lesquels il accomplit les transformations? Si cette perspective est la bonne, alors Searle a mal représenté ce que font les ordinateurs en évoquant la pure syntaxe et non la sémantique.

Pour renforcer cette idée, revenons une fois de plus au langage L. En effectuant des transformations sur les formules de L en l'absence de directives données par le dictionnaire, il n'est pas juste de dire que les symboles ne font pas sens pour vous. Par exemple, la formule « Cwu » veut bel et bien dire quelque chose dans le métalangage que nous utilisons à propos de L. Elle signifie qu'il existe un individu nommé « w », un individu nommé « u », et une propriété relationnelle ordonnée « C » telle que w est en relation C par rapport à u. De même « (A & B) ~ C » signifie qu'il y a trois formules de L telles que, si les deux premières sont toutes les deux vraies, cela vaut pour la troisième. Certes, il ne s'agit pas d'une véritable sémantique sur les rochers, les arbres, et Washington, D.C., mais c'est tout de même une sémantique. C'est pourquoi la réfutation anglaise conclut que ce que montre la chambre chinoise, c'est que l'ordinateur-C est probablement intelligent, précisément dans la mesure où l'intentionnalité est requise par l'ordinateur-C pour imiter avec succès la compréhension chinoise.

Cela aidera peut-être de formuler les choses comme suit. Suivant la réfutation anglaise, ce que Searle a montré, c'est que, dans le cas de l'ordinateur-C, la distinction entre test de langage et test d'action s'effondre. C'est le champ d'action requis pour imiter la compréhension d'un langage naturel qui garantit l'attribution de l'intelligence à l'ordinateur-C. Prenons un perroquet passé maître dans l'art de produire des réponses appropriées en anglais aux questions qu'on lui pose en anglais. Vous regimberez sans doute à l'idée d'admettre que le perroquet parle anglais, mais assurément le simple fait qu'il peut imiter ainsi la compréhension linguistique prouve son intelligence (relativement primaire).

Nous pouvons fortifier la réfutation anglaise en reprenant la réfutation du robot. Supposons que la réfutation anglaise soit correcte, et que nous devions

admettre que Searle a raison de dire que rien dans la chambre chinoise ou l'ordinateur-C ne comprend le chinois. Que faut-il de plus? Que possède un locuteur chinois que ne possède pas l'ordinateur-C? Sans doute l'expérience appropriée d'un monde de rochers, d'arbres, et de Washington, D.C., qui manque à un ordinateur-C confiné. Voilà qui suggère qu'une machine doit ressembler en gros à un robot programmé adéquatement - comme Data - afin d'avoir une intentionnalité authentique, du genre requis pour maîtriser le chinois. Voilà qui prête bien plus de crédibilité à la réfutation du robot qu'à première vue, et fait écho à une suggestion précédente de Turing - à savoir que la meilleure façon de produire un gagnant au test de Turing est de programmer une machine de façon qu'elle simule une pensée d'enfant, et de la laisser apprendre par expérience, comme un enfant humain. (Et Data ne vous fait-il pas penser à un enfant?)

Mais peut-être la réfutation anglaise fait-elle une concession trop importante en admettant que rien dans la chambre chinoise, et donc rien dans l'ordinateur-C, ne comprend le chinois. Suivant la réfutation des systèmes, le système pris dans son ensemble comprend le chinois, même si ce n'est pas le cas de Searle. On pourrait aussi affirmer que Searle comprend bel et bien le chinois, mais n'en est pas conscient. Nous pouvons combiner ces deux réfutations en examinant la version modifiée de Searle, dans laquelle il intériorise entièrement le système. Si la réfutation des systèmes est correcte, alors Searle comprend vraiment le chinois sans le savoir! Il s'abuse lui-même profondément. Nous pouvons toutefois nous passer de cette affirmation en recourant à l'objection finale que j'appellerai la réfutation de l'individu virtuel. Elle dit que Searle a raison, dans une large mesure, de dire qu'une machine ne peut pas penser. Il a même raison de dire que les machines ne sont pas de bons candidats à l'intelligence - mais pas parce qu'elles sont constituées des mauvais éléments; bien plutôt, parce que la pensée n'est pas accomplie par des machines (pas même humaines), mais par des machines virtuelles - des entités réalisées par l'activité de (certaines) machines, lorsqu'elles fonctionnent de façon adéquate.

Le fait que Searle ne comprenne pas le chinois dans la chambre chinoise n'implique pas que personne ne comprend le chinois dans la chambre chinoise. Pour le dire autrement : la proposition « Soit Searle comprend le chinois, soit personne d'autre ne comprend le chinois dans la chambre chinoise » a un centre non distribué. Puisque auparavant la seule personne dans la chambre est Searle, il nous faut prendre en compte la possibilité que Searle, en effectuant ses transformations dans la chambre chinoise, fasse naître une autre personne qui comprenne le chinois. De même, l'ordinateur-C, en remplissant son programme, peut faire naître une personne qui comprend le chinois. Encore une fois, cela fait écho à Turing, qui dit que tout ordinateur digital, pourvu d'une bonne programmation et d'une mémoire suffisante, peut simuler l'opération de toute machine à l'état discret. En jargon informatique, cette simulation est connue sous le nom d'émulation, et lorsqu'une machine exécute une émulation, on dit qu'elle accomplit une machine virtuelle. Lorsque ceci se produit, la machine qui simule et la machine virtuelle ne sont pas identiques. Tout d'abord, elles exécutent des programmes distincts : la première exécute le programme

d'émulation, la seconde une application (un traitement de texte, par exemple). De plus, lorsque l'émulation prend fin, la machine virtuelle cesse d'exister, alors que ce n'est manifestement pas le cas pour la machine qui simule. Pour faire une comparaison avec l'être humain, on peut dire que le corps (et notamment le cerveau) est la machine qui exécute l'émulation de personne. Une personne est une machine virtuelle d'un certain type.

Cette section a été difficile, et j'ai laissé bien des choses de côté. Mais j'espère avoir au moins clairement montré que la chambre chinoise, l'argument de Searle contre la possibilité qu'une machine comprenne des langages naturels, n'est pas sans présenter des failles importantes. Searle a tout au plus montré qu'un ordinateur confiné qui réussit le test de Turing ne maîtrise pas entièrement le langage avec lequel le test est mené. Mais combien d'entre nous ont une maîtrise entière d'un langage naturel ? Lorsque j'écris « Bruce est un wombat », à quel point maîtrisez-vous cette phrase, si vous ne savez pas qui est Bruce, ou ce qu'est un wombat ? Quoi qu'il en soit, Data n'est pas confiné - il va librement dans le monde - et nous n'avons aucune raison de penser qu'il ne comprend pas l'anglais.

Les sentiments

Ar 'drian: « Vous ne ressentez rien pour moi ? »

Data (doucement): « Je ne ressens rien du tout. »

« Les enseignes du commandement »

Nous examinerons les émotions sous peu, mais il y a bien d'autres sentiments que nous associons à des états non émotionnels. La douleur, par exemple (outre la façon dont elle nous fait nous comporter) - et nombre de gens pensent qu'aucune entité faite de silicone et de métal ne pourrait connaître une telle expérience. Data soutient de vive voix cette affirmation, car il est difficile d'imaginer un androïde plus sophistiqué que lui: on voit mal en quoi le simple fait de lui ajouter du matériau de base ou des programmes ferait une différence. (L'idée selon laquelle l'addition d'une « puce émotionnelle » doterait Data de la sensation ne fait qu'embrouiller la question, et n'ajoute rien de significatif à la discussion présente - je la garderai donc pour le point suivant.)

Une fois de plus, Star Trek amène à un débat philosophique animé. Commençons par distinguer entre deux aspects apparemment différents de la sensation: l'un concerne la tâche accomplie pour nous par la sensation, ce qu'on pourrait appeler son aspect fonctionnel - le « on ressent cela » ou « comment ça fait ». Prenons par exemple la douleur. Si vous touchez une surface brûlante, votre sensation de douleur semble jouer un rôle causal important dans le geste qui vous fait vite retirer votre main. Et il y a une explication plausible à la douleur éprouvée : il est extrêmement utile de déceler, aussi vite que possible, que votre main est en train de brûler, et d'agir promptement pour éviter des dommages supplémentaires. La force motivante de la douleur ne saurait être ici sous-estimée ! L'aspect phénoménal de la douleur semble

toutefois accessoire, et il est étroitement relié à la notion de conscience. Nous nous décrivons comme des êtres « conscients » de nos sensations, et de fait nous tendons à penser que cet aspect phénoménal est nécessaire à la sensation.

Data éprouve-t-il des sensations? Eh bien, il nie explicitement pouvoir ressentir de la douleur (comme lorsque LaForge vérifie son réseau positronique dans « Indices »). Mais il se peut que nous n'ayons pas à le croire : examinons un peu ses autres comportements. Il y a d'excellentes raisons de penser que Data a des états qui équivalent fonctionnellement à ce que nous appelons douleur. Pour commencer, toute créature mobile ayant des aspects vulnérables a besoin de pouvoir surveiller l'état de ces parties de lui-même pour éviter ou réparer leurs lésions. Et lorsque le système de surveillance « sonne l'alarme », la créature a intérêt à réagir dans les plus brefs délais. Si Data n'avait pas ce système de surveillance, il pourrait bien ne pas remarquer la menace pesant sur une partie vulnérable de son être, ou ne pas réagir assez rapidement. (Certes, Data peut très bien n'avoir aucun équivalent fonctionnel des récepteurs de douleur à la surface de son cerveau positronique ; notre cortex cérébral, bien protégé d'ordinaire, n'en compte pas non plus.)

On peut objecter que Data n'est pas particulièrement vulnérable. Dans ce cas, ses « signaux de douleur » ont moins de chances de se déclencher ; mais cela ne signifie pas qu'il en est dépourvu. C'est ici qu'intervient une seconde considération : les sensations de douleur et de « pression » semblent de même nature, et ne diffèrent que par leur degré ; toutes deux sont intimement liées à la sensation du toucher. Si, comme moi, vous êtes assis en ce moment, votre postérieur enregistre une pression, et au moment où je tape, je me guide en partie à la pression que je ressens aux extrémités de mes doigts. Une réflexion rapide devrait vous convaincre que vous ne sauriez survivre le temps d'une journée ordinaire sans toutes les informations qui vous sont transmises par l'aspect fonctionnel du toucher. Indestructible ou non, Data a lui aussi besoin de ces informations. Et puisqu'il a l'équivalent fonctionnel d'une sensation de pression, il est probable qu'il aura aussi, à un certain degré, l'équivalent fonctionnel de la sensation de douleur.

Nous pouvons avancer un argument similaire à l'égard d'autres sensations. Data ne pourrait pas se comporter comme il le fait s'il n'avait pas de mécanisme pour distinguer les couleurs, par exemple : il a donc un équivalent fonctionnel de notre capacité à distinguer le rouge du vert. Il possède aussi un équivalent fonctionnel de l'ouïe, et nous savons qu'il peut distinguer différentes substances en les avalant (il l'indique à Timothy dans « Le culte du héros ») : il a donc l'équivalent fonctionnel du goût. Si nous affirmons simplement que Data n'a aucun sentiment en nous référant à l'aspect purement-fonctionnel de la sensation, nous nous trompons.

Toutefois, les persifleurs pensent sans doute à l'aspect phénoménal de la sensation, et nient que Data puisse savoir « comment ça fait » en matière de sensations. La vision fonctionnaliste de l'activité mentale, décrite dans le chapitre 2, semble pour beaucoup de philosophes laisser ouverte la possibilité que deux individus soient fonctionnellement identiques, mais différents par les aspects phénoménaux de leur état mental - par exemple, notre spectre visuel serait inversé, et ce que vous

appelez «rouge» vous paraîtrait être ce que moi je considère comme du « vert ». Il y a une merveilleuse illustration de cette possibilité et des problèmes qu'elle soulève dans «Descente I» (NG). Lorsque Data dit avoir ressenti de la colère, LaForge se montre sceptique: si Data n'a jamais ressenti de colère auparavant, comment sait-il ce qu'il a ressenti? Data admet l'objection, et demande à LaForge de décrire un sentiment de colère, de lui fournir un « cadre référentiel». LaForge, bien sûr, est entièrement incapable de décrire ce sentiment sans faire référence à d'autres sentiments. De plus, certains philosophes pensent que le fonctionnalisme laisse ouverte la possibilité qu'un individu puisse être comme vous et moi fonctionnellement, mais n'avoir strictement aucun état phénoménal - ce que ces philosophes appellent un zombie. Le pauvre Data est-il un zombie?

Une façon de contredire cette conclusion consiste à nier que l'aspect phénoménal de la sensation soit plus que l'aspect fonctionnel pris dans son entier. Pour rendre cela plus plausible, il nous faut distinguer entre une équivalence fonctionnelle et une équivalence purement comportementale. Deux individus sont de comportement similaire s'ils produisent les mêmes attitudes environnementales - par exemple, s'ils sont d'accord sur les dénominations de vert et de rouge. Il n'en découle pas que leurs états internes soient similaires: ils peuvent jouer des rôles fonctionnels différents, même s'ils jouent des rôles comportementaux extérieurement identiques. Dans cette hypothèse, il est possible que les spectres soient inversés parce que des organisations fonctionnelles internes peuvent susciter une équivalence comportementale. Mais si deux individus sont fonctionnellement identiques, alors, nécessairement, ils partagent les mêmes états phénoménaux.

Si, d'un autre côté, l'aspect phénoménal de la sensation est de fait quelque chose de plus que l'aspect fonctionnel pris dans son entier, il n'en découle toujours pas pour autant que Data est un zombie. Voici une hypothèse alternative : si un individu présente l'aspect fonctionnel d'une sensation (la douleur, la vision du rouge), cet état participe nécessairement du phénoménal. Mais bien que présenter l'aspect fonctionnel d'une sensation signifie également en avoir l'aspect phénoménal, cet aspect phénoménal particulier n'a pas nécessairement de particularité. Qu'est-ce qui détermine l'aspect phénoménal particulier, alors ? Peut-être ce qui constitue l'individu. La douleur-chez-les-humains pourrait différer fondamentalement de la douleur-chez-la-Horta, et ce que le rouge signifie pour moi pourrait différer de ce que le rouge signifie pour vous. Faut-il trancher sur cette question difficile pour déterminer le statut de Data? Non, heureusement. Examinons le syllogisme du persifleur, qui est le suivant:

Prémisse n° 1: Il est nécessaire de posséder l'aspect phénoménal des sensations nécessaires pour être une personne.

Prémisse n° 2: Est absent chez Data l'aspect phénoménal des sensations.

Conclusion: Data n'est pas une personne.

Ce syllogisme peut être contesté. Tout d'abord, mettons que votre conception de la qualité de personne vous incline à accepter la prémisse n° 1. La question est alors: que serez-vous justifié d'accepter, la prémisse n° 2 ou le fait que Data soit une

personne? J'admets que si vous acceptez la prémisse n° 1, vous devriez sans doute refuser la prémisse n° 2, étant donné la preuve comportementale impressionnante de la qualité de personne de Data. Notamment parce qu'il y a peu de choses qui soutiennent la prémisse n° 2 (sinon le témoignage de l'intéressé, et nous aurons des raisons d'en douter à nouveau dans la prochaine section). Puisque nous ne savons pas exactement comment nous en venons nous-mêmes à être dans des états phénoménalement caractérisés, nous ne sommes pas vraiment en position d'affirmer que Data ne connaît pas ces états. Si vous rencontriez des extraterrestres à base de silicone se comportant comme des humains, en concluriez-vous immédiatement que, lorsqu'ils identifient les couleurs, leur expérience visuelle est dénué de tout aspect fondamental ? Non, sans aucun doute ! Data a sans doute des états phénoménaux, même s'ils diffèrent en nature de ceux d'un agent cognitif organique.

D'un autre côté, supposons que vous soyez un persifleur récalcitrant, profondément convaincu, pour une raison ou pour une autre, que la prémisse n° 2 est vraie. La question devient: que serez-vous justifié d'accepter, la prémisse n° 1 ou le fait que Data est une personne ? J'admets que, si vous acceptez la prémisse n° 2, vous devriez sans doute nier la prémisse n° 1, étant donné la preuve comportementale impressionnante en faveur de la qualité de personne de Data. Autrement dit, vous devriez revoir votre idée de la personne.

Émotions

*Je peux consacrer de l'attention {à Lai}, docteur,
mais je suis incapable de lui donner de l'amour.*

Data au Dr Crusher, « Le rejeton »

*Je suis incapable de haine ... la crainte n'est
pas un de mes traits de caractère.*

Data à Timothy, « Le culte du héros »

Une objection persistante au postulat que Data est une personne est son apparente absence d'émotion. Une fois de plus, il y a des exemples innombrables où Data reconnaît avoir cette déficience. Mais laissons pour l'instant ces dires, et concentrons-nous plutôt sur ce qu'il fait (une fois encore, tous les exemples que je donne sont tirés d'épisodes où la « puce émotionnelle » est absente).

Prenons quelques exemples tirés de « Ce qui fait l'homme ». Data est sentimentalement attaché à des objets tels que ses médailles, un livre que lui a donné Picard, et un portrait de la défunte Tasha Yar. Interrogé sur Tasha, il semble également lui être sentimentalement attaché. Il exprime, entre autres choses, son regret à l'idée de démissionner, sa confiance dans les capacités de Picard à le représenter, sa gratitude envers Riker qui a accepté d'intenter un recours (sans quoi le procureur général aurait tranché immédiatement contre Data), son espoir à l'idée que Maddox est plus compétent qu'il n'y paraît (lorsque le démontage apparaît comme

inévitable), et sa résignation à cette idée. Data dit que ses «amis» lui manqueront, notamment LaForge, et qu'il est intrigué par l'idée même du démontage avancée par Maddox. Dans « Le jour de Data», nous entendons deux commentaires de Data sur le journal de bord, qui seront transmis à l'Institut de Daystrom pour aider Maddox dans sa recherche. Le premier commentaire suit une modification de trajectoire inexplicée, ordonnée par l'ambassadeur de Vulcain : « Il est heureux que je puisse accomplir mes devoirs sans distractions émotionnelles. Si ce n'était pas le cas, une modification soudaine de la trajectoire en direction de la Zone neutre me rendrait... nerveux. »

Tout en écoutant la voix de Data, nous voyons celui-ci faire une pause, d'un air pensif, et battre des doigts sur la console, avec toute l'apparence de la nervosité. Le second commentaire suit un interrogatoire de Data mené par l'ambassadeur vulcanien sous le prétexte d'une vérification pour des mesures de sécurité: « Commandant Maddox, j'ai souvent regretté l'absence chez moi d'un sens que les humains appellent intuition, ou instinct. Puisque les Vulcaniens sont incapables de mentir, je dois tenir les explications de l'ambassadeur pour vérité, mais je préférerais avoir un sentiment viscéral qui ratifie cette conclusion. »

Ce texte semble exprimer à la fois le regret et l'envie - ou du moins un vague regret, ainsi que le soupçon.

Ajoutons à ces exemples toutes les occasions où le comportement cinétique de Data suggère un état émotionnel. Voici une sélection glanée dans La nouvelle génération : son étonnement face à une coiffure extraterrestre dans « Le jour de Data», son souci, son indulgence, ses attentes, sa patience et sa fierté à l'égard de sa « fille » Lal dans « Le rejeton », sa stupéfaction face à l'incapacité initiale de Picard à lui donner un ordre dans « Rédemption II » ; une fois qu'il a enfin reçu un commandement, il manifeste de l'impatience face à l'insubordination du lieutenant-commandeur Hobson. Puis il y a son angoisse à l'idée d'une relation sexuelle avec Tasha Yar; son soulagement lorsque l'Entreprise-D annule sa séquence d'autodestruction dans « Le bail du silence», son sentiment d'appartenance lorsqu'il rejoint l'équipage dans «Tin Man», sa surprise lorsque Riker répond avec brusquerie à une question de routine dans « Le facteur d'Icare », son étonnement, sa curiosité et sa fierté lorsqu'il assiste à l'accouchement de Deanna Troi dans «L'enfant», sa détresse face aux blessures du Dr Crusher dans «L'arsenal de la liberté», sa curiosité quant à la question de savoir s'il existera encore ou non au xxvr siècle dans « Une question de temps», son embarras face aux problèmes informatiques dans «Évolution», son air interrogateur et impuissant, trahissant ses inquiétudes lorsqu'il tente d'évacuer les colons dans « Les enseignes du commandement», et son opiniâtreté à les convaincre malgré tout. Dans « Le culte du héros», Data dit à Timothy: « Je ne saurais tirer aucune fierté de mes capacités ... Je ne puis prendre plaisir à ce que je réussis ... Je prendrais volontiers le risque d'être mal à l'aise à certains moments, si cela signifiait également que je puisse apprécier mes mérites. » Mais il dit ceci avec un vague regret, comme s'il était mal à l'aise de ne pas être mal à l'aise. Par la suite, lorsqu'il dit à Troi que se séparer de Timothy ne lui sera pas

difficile parce que « cela requerrait un contexte émotionnel [qu'il] ne saurai[t] susciter», il le dit sur un ton de regret.

Tout bien pesé, nous avons des preuves comportementales impressionnantes que Data a bel et bien des émotions. Et je ne suis pas le seul à être impressionné : lorsque Data dit au Dr Crusher qu'il ne saurait donner de l'amour à Lai, elle répond: « Mais pourquoi ai-je tant de mal à le croire? » Et dans « Les enseignes du commandement », lorsque Picard, qui a écouté Data jouer du violon, dit : « Votre prestation trahit... un sentiment », Data réplique qu'il n'a pas de sentiments, et Picard dit à son tour: « C'est dur à croire. »

Dans l'autre plateau de la balance, nous trouvons des occasions où Data adopte un comportement absolument inapproprié. Dans « En théorie », il essaie de faire la cour à sa « petite amie » Jenna en concevant un programme d'interaction émotionnel. Après s'être essayé à la relation amoureuse et à l'attention, il réagit au manque d'enthousiasme chez Jenna par un hurlement agressif. Mais Jenna se plaint de ce que tout ce que fait Data est contraint et artificiel. Ils finissent par s'embrasser, et Jenna demande à Data à quoi il pensait en l'embrassant. Data réplique : « À ce moment précis, je recalculais les paramètres des champs de courbure, j'analysais les œuvres complètes de Charles Dickens, j'étudiais un nouveau supplément nutritionnel pour Spot [son chat] et je calculais la pression maximale que je pouvais appliquer à vos lèvres sans les endommager ... »

Cette séquence pourrait s'expliquer simplement ainsi : loin de tenter quelque chose dont il est incapable en principe, Data essaie trop intensément de faire ce qu'il faut. D'innombrables humains de sexe masculin se sont également retrouvés dans la situation catastrophique où ils veulent faire plaisir à une femme sans savoir comment s'y prendre. (La réciproque est peut-être vraie.) L'erreur de Data est de se figurer les réactions appropriées et de les programmer, alors que manifestement Jenna est attirée par lui dès le début, et ce en raison de ce qu'il faisait auparavant. La situation ressemble un peu à celle d'un humain de sexe masculin qui essaierait de faire fonctionner sa raison au lieu de son « instinct » dans ses relations avec sa partenaire - en général, c'est un échec, et son comportement paraît contraint et artificiel. Data gâche tout parce qu'il ne se sent pas en sécurité, tout en montrant clairement par ses symptômes qu'il se soucie de Jenna. Si on lui en donnait le temps, à force d'échecs, je parie qu'il manifesterait même de l'exaspération.

Tout bien pesé, je crois que, même en l'absence de la puce émotionnelle, nous devrions attribuer des émotions à Data. Pourquoi la tentation est-elle donc si forte, y compris chez ceux qui seraient prêts à admettre qu'un androïde a d'autres états mentaux, de nier qu'il puisse avoir des affects? Tout d'abord, Data nie constamment être dans un état affectif, quel qu'il soit. Ces dénis peuvent être expliqués de trois façons. Tout d'abord, il se peut que les émotions soient véritablement absentes chez Data, que celui-ci le sache et le reconnaisse. Mais cette explication est démentie par les autres comportements de l'intéressé. La seconde possibilité, c'est que Data ait des émotions, le sache, mais persiste dans son mensonge pour des raisons connues de lui seul (peut-être craint-il les réactions d'autrui). J'ai néanmoins des doutes sur

ces explications : Data est en général sincère, parfois même de façon inappropriée. Je préfère la troisième explication : Data a des émotions, mais s'abuse sur elles. Data reconnaît qu'il est une machine et a cédé au mythe répandu selon lequel une machine ne saurait avoir d'émotions, il s'est convaincu qu'il ne peut donc en avoir. Par exemple, dans « Le jour de Data », il dit à Riker qu'il ressent « de l'angoisse » à l'idée de jouer le rôle du père de la mariée, après avoir reconnu qu'il ne saurait être nerveux. Je soupçonne qu'il est véritablement nerveux, mais dénie cette expérience en décrivant la nervosité en termes absolument autres. Un être humain ordinaire aurait vu ce « sentiment » confirmé par autrui, et apprendrait le nom du sentiment par la suite. Le pauvre Data, lui, apprend à dénoter le sentiment de nervosité par un tout autre nom. (Voyez la réticence de LaForge à « ratifier » la colère de Data dans « Descente 1 ».) Lorsque Data médite sur la brusque modification de trajectoire dans « Le jour de Data », il est à nouveau nerveux, au point que sa nervosité l'interrompt dans l'accomplissement de ses devoirs ! Et lorsqu'il regrette son absence d'« instinct viscéral », qui pourrait étayer ses raisonnements explicites sur l'ambassadeur vulcanien, assurément la raison pour laquelle il regrette l'absence de ce sentiment, c'est qu'il a vraiment un instinct viscéral que quelque chose ne va pas. Les dénis de l'émotion sont courants chez les humains. Convaincu de ne pas avoir été furieux, envieux ou impatient, j'ai parfois été contraint par la suite de reconnaître que c'était pourtant le cas. Les hommes en général ont la réputation de reconnaître et d'exprimer difficilement leurs propres émotions - dans ce domaine, le comportement de Data est humain, trop humain !

La seconde raison pour dénier des émotions à Data nous renvoie à notre analyse de Spock au chapitre 1. J'ai remarqué à cette occasion que nous pouvions distinguer utilement entre les émotions supérieures (associées à l'activité corticale chez les humains, et qui convoient peu ou pas d'affects) et les émotions inférieures (associées au système limbique et au bulbe rachidien, et caractérisées par des sentiments forts). Data se comporte beaucoup comme Spock, et je crois que l'opinion la plus défendable est que Data, comme Spock, a l'expérience d'une large variété d'émotions supérieures, alors qu'ils manifestent peu ou pas d'émotions inférieures, comme la colère. Personne ne pourrait raisonnablement nier que Data éprouve des émotions inférieures (du moins, pas sans une puce émotionnelle). Au chapitre 1, j'ai suggéré qu'une créature cognitive a besoin d'émotions, mais je n'ai pas suggéré que les émotions inférieures étaient nécessaires. La colère et la peur sont utiles aux créatures relativement destructibles, mais seraient-elles utiles à un individu comme Data ? Dans la mesure où peu de choses peuvent le menacer physiquement, il peut s'en sortir parfaitement sans peur. Et puisque la colère nous permet en général de prendre des forces pour un combat, voire de décupler nos capacités physiques, Data (qui possède déjà une force surhumaine) n'aura pas grand besoin d'elle. En tant que créature cognitive par excellence, Data saura sûrement exploiter comme nous les émotions supérieures : dès lors, quelle que soit la façon dont il y a accès (qu'il les apprenne ou qu'elles lui soient intégrées), sa façon de manifester des émotions supérieures est assez cohérente. Cela posé, demeure le mystérieux désir de Data

d'être humain - d'expérimenter la vie psychologique d'un Homo sapiens typique. Comme Spock le remarque à propos de la quête de Data, dans « Unification 1 » : « Plus humain? Fascinant ! Vous avez un intellect efficace, des dons physiques supérieurs, aucune entrave spirituelle. Il y a des Vulcانيens qui aspirent toute leur vie à obtenir ce dont vos concepteurs vous ont doté. »

J'ai suggéré que Data n'aurait sans doute aucun trait mental inutile, mais j'ai aussi suggéré qu'il s'abuse lui-même sur ses propres états émotionnels. Serait-ce me contredire? Pourquoi un androïde aurait-il des mécanismes d'auto-aveuglement ? Demandons-nous plutôt pourquoi nous, nous possédons ces mécanismes. Ils peuvent être un handicap lorsqu'il nous faut accomplir une tâche dangereuse (conduire en état d'ivresse) et que nous nous en croyons à tort capables. Ils nous occasionnent toutes sortes d'embêtements lorsque nous refusons d'admettre notre état. Et nous pouvons nous adonner à des pratiques très nocives (la drogue, la kleptomanie) tout en nous disant, en nous aveuglant, que nous n'avons pas de problème.

Mais l'auto-aveuglement a aussi ses avantages. Supposons que vous êtes attaqué(e) par un animal dangereux - un tigre aux dents de sabre, par exemple - et qu'en essayant de fuir vous vous foutez la cheville. Il vous est extrêmement utile que les « messages » que vous recevez de façon interne ne vous alertent pas sur la gravité de votre blessure (une douleur excessive vous affaiblirait, et vous obligerait à interrompre votre fuite). Au contraire, votre « mécanisme de choc » fait momentanément obstacle à ces messages, vous permet de continuer à courir, et vous sauve donc la vie. De plus, il se peut que les créatures cognitives comme nous doivent se protéger contre certaines vérités. (Freud pensait que tout homme veut tuer son père et épouser sa mère; si c'est le cas, ce pourrait être trop affreux à supporter, et mieux vaut pour les hommes qu'ils n'y croient pas.) Des auto-aveuglements défavorables ne sont peut-être que le dérapage malheureux d'un mécanisme en général utile.

Dans l'androïde que vous construirez, vous ferez bien d'inclure des mécanismes d'auto-aveuglement. Un androïde relativement indestructible comme Data n'aurait sans doute pas besoin de l'équivalent de notre mécanisme de choc, mais il voudrait peut-être pouvoir opérer sans avoir besoin de réparations. Et s'il est un androïde cognitif, il se peut qu'il y ait des vérités trop horribles pour lui : lorsque le « frère » aîné de Data, Lore, lui ment dans « Datalore » en lui disant qu'il a été conçu délibérément pour être imparfait, Data fait une fixation sur cette idée ! Puisque nous parlons de mensonge, voici une nouvelle preuve de l'auto-aveuglement de Data. Dans « Le culte du héros », Data dit à Timothy que « les androïdes ne mentent pas ». Non seulement Lore ment, mais Data aussi, lorsqu'il considère que c'est une bonne solution - voyez « Indices », où il ment à Picard pour protéger l'équipage de l'Enterprise-D. Et pourtant, je ne pense pas que Data mente à Timothy; c'est plutôt qu'il ne réalise pas qu'il lui arrive de mentir, et qu'il en est capable.

De plus, il ne me semble pas qu'il soit entièrement inconscient de ses capacités émotionnelles. Dans « La qualité de vie », Data dit à Riker: « J'ai observé que les humains fondent souvent leur jugement sur ce qu'on appelle "instinct" ou "émotion".

Parce que je suis une machine, je n'ai pas ce talent particulier. Toutefois, il se peut que je puisse deviner chez d'autres machines des choses qui échapperaient aux humains. » Et lorsque Fajo accuse Data de « manquer de cœur » dans « Le jouet par excellence » (« Vous n'êtes qu'un androïde »), je crois qu'il soupçonne que Data est bel et bien capable de réactions émotionnelles, et qu'il essaie d'en provoquer une. Par la suite, lorsque Fajo est fait prisonnier, Data annonce que sa collection de pièces uniques a été confisquée. « Vous avez perdu tous vos biens », dit Data, et Fajo remarque : « Cela doit vous faire rudement plaisir. » « Non, monsieur, pas le moins du monde, réplique Data. Je ne ressens aucun plaisir. Je ne suis qu'un androïde. » Vu l'air que prend Data en lançant cette réplique, on comprend qu'il ressent (pour le moins) une satisfaction calme et résignée : il y a là un soupçon de malice, qui pourrait bien être du plaisir.

Je n'ai pris en considération jusqu'ici que la vie mentale probable de Data en l'absence de la puce émotionnelle que le Dr Soong entendait lui greffer dans « Frères » et qui est finalement intégrée dans Star Trek VII: Générations. D'une certaine façon, l'addition de la puce émotionnelle va de pair avec l'interprétation que j'ai donnée au comportement de Data. Si, comme le suggère le Dr Soong, la puce procure des émotions « basiques » (et donc inférieures), alors elle ne fait que compléter les émotions supérieures que ressent déjà Data. Mais il est improbable qu'un agent cognitif déjà pourvu d'émotions supérieures puisse acquérir les émotions inférieures par l'ajout d'une puce. Nos systèmes cognitifs ont évolué à partir de systèmes moins complexes, dont font partie les émotions inférieures. En termes d'évolution, les fonctions corticales ont été rajoutées à une économie cognitive basique, dont les émotions inférieures font partie intégrante. On pourrait en conclure que, si l'on doit avoir des émotions inférieures, elles doivent faire partie de sa structure cognitive basique. Je suis enclin à penser, en partie à cause du problème de cadre (voir chapitre 1), que, si jamais nous réussissons à construire des androïdes capables d'agir dans le monde presque aussi bien que les mammifères supérieurs, ils auront certaines de ces émotions fondamentales.

Conscience

Nous avons déjà évoqué la question de la conscience dans les deux dernières sections. L'aspect phénoménal de la sensation tient à la façon dont elle se présente à la conscience, et bien des émotions se laissent par ailleurs « sentir » de façon consciente et subjective. Qu'est-ce exactement que la conscience ? Voici l'une des questions les plus difficiles de la philosophie. Nous pouvons distinguer entre différentes significations de « conscient » : des individus peuvent être doués de conscience, voire d'une conscience de soi, et les états mentaux peuvent être conscients (les sensations, et certaines émotions, par exemple). Il est probable que ces significations sont liées, dans la mesure où les individus n'ont de conscience de soi que s'ils ont d'abord une conscience, et qu'ils n'ont cette conscience que dans la mesure où ils ont des états conscients. Dans « Ce qui fait l'homme », Maddox pose

trois conditions à la qualité de personne: « l'intelligence, la conscience de soi, la conscience tout court ». Si on peut objecter que Data a une conscience, celle-ci doit d'abord porter sur son statut.

Les théories de la conscience d'état les plus débattues sont la théorie d'un ordre supérieur (ou TOS) et la théorie d'une perception interne. (Il y a une troisième perspective assez populaire, l'éliminativisme, qui dit, pour aller vite, qu'il n'y a pas de conscience. Ce qui ne signifie pas que nous sommes tous des zombies, mais plutôt que la conscience est une catégorie populaire qui disparaîtra avec l'aboutissement de la science du cerveau, tout comme l'adjectif « calorique » - appliqué à une substance censée fournir de la chaleur - n'est pas une notion de la science de la thermodynamique sous sa forme aboutie.) Suivant la TOS, la conscience est une propriété de certains états mentaux lorsqu'ils sont des pensées portant sur d'autres pensées. Suivant la théorie de la perception interne, la conscience apparaît lorsque certains états mentaux en perçoivent d'autres.

Une fois encore, ce n'est pas à nous de décider de la théorie correcte pour notre question présente. S'il existe véritablement une conscience humaine, puisque nous ignorons comment et pourquoi nous autres êtres humains en sommes venus à l'acquiescer, nous ne sommes pas en position de nier qu'un ordinateur ou qu'un androïde pourrait avoir des états conscients. De plus, le comportement de Data donne toute raison d'admettre que Data est conscient. Star Trek soutient ce postulat dans deux épisodes distincts. Dans « Élémentaire, mon cher Data », Picard et Data expriment des doutes sur l'affirmation de Moriarty, qui dit être conscient, et Moriarty rétorque en demandant si « l'homme mécanique » est ou non conscient. Démentant pour une fois sa modestie habituelle, Data répond : « Je suis conscient, oui. » Et « Prototype » (V) montre même « ce que ça fait » d'être un androïde. Nous voyons la première scène du point de vue de l'unité automate 3947. C'est comme si nous voyions ce que voit cet androïde.

Mais si Data est conscient, cela signifie-t-il qu'il a également un inconscient? Par exemple, rêve-t-il? Une fois de plus, il n'y a aucune raison, en principe, qu'un androïde n'ait pas de rêves, ou leur équivalent androïde. (La nature de l'expérience pourrait dépendre en partie de la nature physique de l'individu, comme dans le cas de la sensation.) Dans « Droit de naissance 1 », Data commence à rêver à la suite à d'un choc sévère, ce qui est expliqué par le fait que son programme de rêve a été activé plus tôt que prévu. Data commence alors à explorer son propre inconscient à travers la peinture créative, et le Dr Soong finit par lui apparaître en rêve pour lui annoncer qu'il a maintenant « franchi le seuil », et qu'il n'est plus simplement un amas de puces informatiques. Je doute que le rêve en lui-même ait une telle importance: peut-être Soong affirme-t-il que Data a maintenant un inconscient, qui lui faisait précédemment défaut? C'est difficile à accepter - pour les raisons qui nous faisaient dénier à Data la possibilité d'acquiescer des émotions inférieures en option. Je soupçonne que la conscience est un développement évolutif ultérieur à l'inconscient (spéculation liée à la nécessité d'une division mentaliste du travail et au besoin de fonctions mentales basiques qui soient inaccessibles à la conscience). Peut-être devrions-nous interpréter

les remarques de Soong différemment: les « circuits de rêves » de Data font partie intégrante de son architecture cognitive et nécessitent un certain niveau de développement cognitif avant de se mettre en marche. Nous n'en savons pas assez sur le développement de notre propre capacité à rêver pour en dire plus.

Valeurs

*Un ordinateur ne juge pas -
il procède à des sélections logiques.*

Spock à Kirk, «L'ultime ordinateur»

Dans « L'ultime ordinateur », Kirk déclare que l'ordinateur M5 « ne saurait formuler un jugement de valeur », sans doute parce qu'il est une machine. On distingue ordinairement les jugements de valeur et les jugements factuels ; par exemple, « cette chemise est de couleur vive » est un jugement factuel, alors que « cette chemise est laide » est un jugement de valeur. Les jugements factuels sont censés porter, comme leur nom l'indique, sur des faits ; ils sont vrais ou faux. Mais cette distinction ne fait pas grand sens à mes yeux si elle implique que les jugements de valeur ne portent pas sur des faits, et ne sont ni vrais ni faux. On affirme parfois que la valeur, comme la beauté, est dans l'œil du spectateur, mais on n'a jamais avancé d'argument valable pour le prouver (certaines chemises sont indubitablement laides!). S'il faut faire ici une distinction, c'est entre le jugement descriptif et le jugement normatif - le premier affirme ce que sont les choses, le second ce qu'elles devraient être (qu'elles le soient ou non). Tout ce que nous pouvons dire sur un « jugement de valeur », c'est qu'il a force normative (dans le cas des chemises laides, ce serait : « Vous devriez avoir meilleur goût, ou plus de bon sens, au lieu de porter cette chose en public ! »), alors qu'un jugement factuel n'a pas force normative.

La distinction entre normatif et descriptif renvoie à la distinction entre logique et émotion (ou raison et passion, ou croyance et désir) discutée au chapitre 1. Tout comme on peut dire que la logique dénuée d'affectivité ne peut offrir aucune motivation psychologique à un acte, on pourra dire que le jugement descriptif seul ne peut offrir de motivation pratique à un acte. Par exemple, on dit parfois que la science est une entreprise purement factuelle, et que ce n'est pas aux scientifiques de formuler des jugements de valeur. Mais c'est ridicule. Si les scientifiques se contentaient de relever les faits, lesquels choisiraient-ils parmi l'infinité de faits disponibles? Les scientifiques ont eux aussi un problème de cadre, similaire à celui que connaissent les agents cognitifs, et que nous avons évoqué au chapitre 1. Sans une idée préalable de ce qu'il nous faut chercher - sans un jugement portant sur ce qui est bon à chercher - les scientifiques seraient figés dans l'inaction. La science, comme toute autre entreprise humaine, est nécessairement faite de jugements de valeur.

Data formule-t-il des jugements de valeur? Absolument, comme le montre avant tout le moment dans « Le jouet par excellence » où Fajo lui parle de son enfance démunie. Data réplique : « Votre passé n'excuse en rien un comportement immoral,

monsieur.» Nombre des jugements de valeur de Data ont un contenu moral, comme dans « La qualité de la vie », où il déclare: « Je ne crois pas qu'il soit juste de sacrifier une forme de vie pour une autre. » Non seulement il porte un jugement moral, mais il fonde là-dessus sa conduite, avec une intégrité supérieure à celle de bien des humains. Voyez son insubordination lorsqu'il défend les exocomps dans ce même épisode. Il serait temps, maintenant, de cesser de traiter Data comme une machine. Humain ou non, il est une personne, comme Maddox semble l'admettre en fin de compte.

Et il serait peut-être temps aussi d'examiner les actes des exocomps. Pour sauver Picard et LaForge, trois d'entre eux se lancent volontairement dans une expédition de sauvetage risquée. Après qu'elle a réussi, l'un des exocomps reste en arrière, se sacrifiant volontairement pour que les deux autres survivent. Même si Picard et les autres n'établissent jamais de contact linguistique avec les exocomps (pas dans un langage naturel, en tout cas), les actes de ces individus révèlent clairement la présence en eux d'une moralité, et suggèrent que les exocomps sont aussi des personnes.

Sommes-nous du même corps?

Lieutenant Reginald Barclay: « Vous ne pouvez pas avoir pris conscience du temps qui passe. »

Professeur Moriarty : « Oh, mais si. De brèves et terribles périodes où j'avais conscience d'être désincarné. »

Barclay: « Je ne vois guère comment ce pourrait être possible. »

« Le bateau dans la bouteille » (NG)

Pour conclure nos recherches sur la qualité de personne, il nous faut examiner le statut de deux formes exotiques dans Star Trek: l'une est l'intelligence désincarnée, l'autre l'intelligence non individuelle.

Dans « Élémentaire, mon cher Data », LaForge essaie de concevoir un programme de holodeck sur le thème de Sherlock Holmes, pour mettre Data au défi : il demande à l'ordinateur de l'Entreprise-D de créer un ennemi de taille, mais, commettant un lapsus, il demande un adversaire non pas digne de Sherlock Holmes, mais « capable de battre Data ». D'où la « naissance » du professeur Moriarty, une entité qui diffère des autres personnages suscités par le holodeck en ce qu'il est conscient d'exister. (Ce qui, à première vue, nous ferait considérer Moriarty comme une personne, puisqu'il a une conscience de soi.) Après certains moments de tension, Moriarty accepte d'être enfermé et « économisé », contre la promesse de Picard de chercher les moyens pour le faire exister de façon permanente, en dehors du holodeck. Quatre années passent, et dans « Le bateau dans la bouteille », Moriarty est accidentellement réactivé par le lieutenant Barclay. Dans le dialogue ci-dessus, Barclay est sceptique parce qu'il ne comprend pas, vu ses connaissances en matière de holodeck, comment le programme de Moriarty aurait pu être actif dans l'intervalle.

Mais il y a une autre raison d'être sceptique : comment postuler une conscience désincarnée - une intelligence sophistiquée, dénuée de substance ?

Dans chacun des cas que nous avons examinés jusqu'ici - les androïdes, comme Data, les machines, comme les exocomps - les suspects numéro un s'incarnaient, d'une façon ou d'une autre, leur être physique occupait un certain espace, absorbait les données que lui fournissait son environnement, et adoptait un comportement. Dans «Émergence», lorsqu'il semble que l'Entreprise-D acquiert une intelligence momentanée et que Data compare le vaisseau à une forme de vie, Riker acquiesce, disant: « C'est vrai. Elle voit avec ses senseurs, parle avec ses systèmes de communication.» Une fois encore, nous avons affaire à une intelligence qui s'est incarnée. Mais nous faut-il considérer l'intelligence comme quelque chose d'incarné? Parfois, Star Trek suggère que la réponse est non, dans le dialogue mis en exergue, comme dans les diverses rencontres entre le personnel de la Flotte et ce qu'il nomme des êtres de pure énergie. (En voici un rapide inventaire: on Jes trouvera dans «Métamorphose», « Les joueurs de Triskelion », « Retour à demain », «Les lumières de Zetar» [PS]; dans «Seul parmi nous», «Transfigurations» et «L'ami imaginaire» [NG], ainsi que dans « Le nuage », « Héros et démons » et « Cathexis » [V].)

Je n'ai qu'un mot à dire sur la conscience ou l'intelligence désincarnées: je n'y crois pas. Et les exemples de Star Trek ne prouvent en rien que nous puissions les imaginer. Prenez les êtres d'énergie pure, par exemple. Puisque la matière et l'énergie sont convertibles entre elles, le fait que ces êtres soient «composés d'énergie pure» ne signifie pas qu'ils soient immatériels. De plus, le fait qu'ils sont actifs dans le monde physique, qu'ils reçoivent et communiquent des informations et manipulent des objets physiques, suggère qu'ils ont malgré tout un corps, mais un corps différent. Même une entité étrange comme Moriarty est incarnée; sa matière de holodeck peut n'être pas stable à l'extérieur du holodeck, mais c'est secondaire. Mais si Moriarty est juste un programme d'ordinateur, ne peut-il être stocké et inactif? Savoir s'il est ou non un programme d'ordinateur est une question délicate, que nous aborderons de façon quelque peu oblique dans la seconde partie. Mais dans tous les cas il devrait être inactif une fois stocké. C'est pourquoi je suis d'accord avec Barclay: d'une façon ou d'une autre, le holodeck n'a pas stocké le programme de Moriarty, d'où ses « brèves et terribles périodes ... [de] conscience».

Par ailleurs, Moriarty n'a pas besoin d'être sur le holodeck pour être actif. Au dénouement du « Bateau dans la bouteille», Moriarty est piégé: on lui fait croire qu'il a quitté le holodeck sous une forme corporelle ordinaire. En fait, son programme a été transféré dans un ordinateur gérant un programme continu comprenant une vie entière d'expériences simulées. Cet ordinateur est assez petit pour être placé dans un coin du bureau de Picard. Moriarty n'a donc pas un corps humain, pas même une simulation façon holodeck d'un corps humain. Néanmoins, pour qu'il ait des expériences et interagisse avec le programme interactif, Moriarty doit être conçu sous une forme matérielle ou une autre, peut-être grâce à une série de puces informatiques.

Ce qui vaut pour lui vaut pour les autres personnages du holodeck qui sont des

suspects numéro un, notamment le hologramme médical d'urgence à bord du Voyageur. Ce docteur holographique s'en tire bien par rapport à Moriarty, dans l'un et l'autre épisode. Le docteur existe sur un holodeck médical spécial, il a un corps substantiel d'humanoïde et peut même quitter occasionnellement le holodeck grâce à des moyens technologiques spécifiques. En tant qu'officier médical du Voyageur, membre à part entière de l'équipage, le docteur est une personne jouissant de tous ses droits : c'est indéniable, si on considère son comportement et l'effet qu'il a sur les autres. (Dans « Signe de vie », il a même une romantique histoire de cœur avec une Vidiennne, le Dr Damara Pel.) Il se peut que, dans un épisode à venir, nous voyions une réédition de « Ce qui fait l'homme », où le docteur est contraint de combattre pour préserver ses droits. Le premier officier du Voyageur, le commandant Chakotay, ferait un bon avocat, vu ce qu'il dit au docteur dans « Projections » : « Ce n'est pas parce que vous êtes fait de lumière projetée et d'énergie que vous êtes moins réel qu'un être de chair et de sang ... ce qui importe, c'est qui vous êtes. »

On trouve deux exemples de conscience collective dans Star Trek. Le plus connu est celui des Borgs de La nouvelle génération, une race qui assimile la nouvelle technologie et les corps humanoïdes, en les combinant dans des cybermen. Un individu borg est conscient, mais n'a aucune idée de soi, un fait que l'équipage de l'Entreprise-D tente d'exploiter dans « Moi, Borg ». Ils capturent un Borg et coupent sa communication avec la collectivité borg, après quoi il développe une idée de lui-même et prend pour nom Hugh. Hugh finit par être rendu aux Borgs, dans l'espoir que l'idée d'individualité se répandra et perturbera la conscience collective. Dans « Évolution » (NG), Wesley Crusher mène une expérience de nanotechnologie. Les nanites, des outils infinitésimaux utilisés pour effectuer des réparations biologiques, travaillent en général individuellement, mais Wesley réussit à les amener à coopérer. Malheureusement, elles échappent, se multiplient, et prennent le contrôle de l'ordinateur de l'Entreprise-D, menaçant à la fois l'équipage et sa mission scientifique, qui consiste à aider le Dr Paul Stubbs à étudier une explosion stellaire.

Les Borgs et les nanites manifestent une conscience collective, et les Borgs manifestent par surcroît une conscience individuelle. Mais ni l'individu borg ni les nanites individuelles ne sont des personnes, puisqu'ils n'ont aucune idée d'eux-mêmes. Toutefois, voici un exemple où l'on traite les collectivités en tant que personne. Prenons une analogie : des corporations et d'autres groupes de personnes travaillant ensemble sont souvent considérés par la loi comme des personnes, et nous pouvons appliquer le même principe à des questions morales comme le blâme et la louange, lorsque deux personnes ou davantage agissent de concert. On pourrait donc penser que la corporation est coupable de négligence, et non chaque membre individuel de cette corporation. De même, on pourrait penser que les collectivités nanite ou borg sont moralement responsables, à titre collectif, et que cela ne vaut pas pour l'individu nanite ou borg.

Résumé

Notre examen de la personne a révélé que la qualité de personne artificielle (et, par extension, d'intelligence artificielle) est une possibilité à part entière, et que le test le plus sûr pour déterminer si un individu peut prétendre ou non à la qualité de personne ne dépend pas de sa capacité à se faire passer pour un être proprement humain. S'il existait des individus comme Data, ou les exocomps, ou l'hologramme médical d'urgence sur le Voyageur, nous devrions considérer ces individus comme des personnes, et les traiter en conséquence - nous devrions autant prendre en compte leurs intérêts que ceux des humains. Sans quoi, nous ferions acte de discrimination injustifiée. Mais nous n'en avons pas fini avec la qualité de personne. Dans la seconde partie, nous examinerons les nombreux processus exotiques que subissent les personnes dans Star Trek, en nous demandant si ces processus sont possibles, et si les êtres humains pourraient y survivre.

Deuxième partie

Questions de survie

Chapitre 4

Être ou ne pas être ... téléporté !

Reg, la téléportation est de loin la façon la plus sûre de voyager.

LaForge à Reg Barclay, « Les domaines de la terreur » (NG)

Hamlet, Shamlet I

Star Trek n'égale pas encore tout à fait Shakespeare en popularité, mais bien des répliques cultes de cette série sont passées dans la langue populaire, tout comme « Être ou ne pas être », « Mon royaume pour un cheval », etc. (Détail intéressant, nombre d'expressions shakespeariennes ont survécu au XXIV^e siècle - si l'on en croit Star Trek.) « Téléportation, Scotty ! » fait partie de ces répliques, ce qui est d'autant plus surprenant que James T. Kirk ne la prononce jamais dans la série; tout au plus dit-il: « Scotty, téléportation ! » dans Star Trek IV: Retour à la maison. C'est une formule utile, notamment lorsqu'on a un besoin urgent d'échapper aux circonstances. Mais je vais étudier la téléportation sans tenir compte de son côté salvateur, plutôt comme simple moyen de transport. Je n'ai pas fait le compte, mais le personnel de la Flotte a dû être téléporté des centaines de fois devant nos yeux, sans compter les milliers de fois où la scène n'est pas montrée. Parfois, néanmoins, les personnages rechignent à subir le processus. Les tendances antitechnologiques du Dr McCoy se manifestent, entre autres, lors de ses fréquentes manifestations d'hostilité à l'égard du téléporteur. Le Dr Pulaski semble partager ses sentiments et, dans « Le domaine de la terreur », Reg Barclay développe une phobie radicale.

Comme on l'a analysé en long et en large, la raison d'être dramaturgique du téléporteur, c'est que le vaisseau peut stationner en orbite et n'a ainsi nul besoin d'atterrir sur une planète étrangère dans une débauche peu convaincante d'effets spéciaux. Star Trek n'a pas le monopole de la téléportation, mais la série l'a ancrée dans notre conscience collective. Dans La physique de Star Trek, Lawrence Krauss note « l'impact que cette technologie hypothétique a eu sur notre culture - un impact d'autant plus remarquable que, de toutes les prouesses technologiques de science-fiction exhibées à bord de l'Enterprise, c'est la plus improbable ». En tant que physicien, Krauss s'intéresse à l'opération du téléporteur telle qu'elle est présentée pour déterminer si cette technologie est réaliste ou non. En tant que philosophe, je m'intéresse à cette même opération pour savoir si le téléporteur est vraiment ce qu'il

prétend être. En particulier, j'essaierai de répondre à la question: Pourrait-on survivre à ce processus?

Posons dès maintenant deux questions distinctes. L'une, c'est de savoir si le processus est fiable ou non. Peut-être les frissons éprouvés par McCoy et les autres ne relèvent-ils que d'un manque de confiance dans la technologie. Précisons donc que le téléporteur est non seulement technologiquement possible, mais qu'il fonctionne presque parfaitement et, en principe, fait toujours exactement ce pour quoi il est conçu. À cet égard, le téléporteur est éminemment sûr - bien plus que les automobiles, les avions, et les astronefs; c'est bien ce qu'affirme LaForge dans l'exergue.

Mais ces frissons peuvent relever de raisons plus profondes. On peut avoir entièrement foi dans la fiabilité d'un processus et rechigner encore à le subir. La guillotine, par exemple, est fiable et hautement indésirable. Peut-être McCoy se demande-t-il si ce qui se «rematéralisera» sera encore lui! Tant qu'à mettre des mots dans la bouche de Kirk, faisons-le aussi avec McCoy; j'aime à me le figurer imaginant sa destruction imminente et murmurant« Téléportation, Scotty ! » alors que l'ordre d'« énergiser » est donné.) De fait, il est facile d'imaginer des civilisations entières - terrestres ou extraterrestres - qui refuseraient pour les mêmes raisons de se soumettre à la procédure. (C'est le cas des Trills dans «L'hôte» [NG], mais pour mieux garder secrète leur nature symbiotique.) Est-il donc rationnel d'accepter la téléportation en se fondant sur la simple croyance qu'on y survivra? McCoy et ses semblables font-ils preuve de sagesse, ou de simple superstition ? La réponse ne dépend pas uniquement des opérations du téléporteur, mais de la question de savoir ce qu'est la survie d'une personne.

Hypothèses de travail

Hélas, Star Trek n'est guère cohérent sur le mode de fonctionnement du téléporteur. Il y a deux façons de conceptualiser cette technique. L'une est d'y voir un téléporteur de matière, qui scanne le sujet et le «dématérialise» littéralement, puis envoie la matière sous forme d'énergie, ainsi que les informations scannées, à la destination voulue où se produit la « rematéralisation ». C'est ce qui se passe la plupart du temps si nous acceptons la description donnée dans les épisodes de La nouvelle génération intitulés « Rencontre à Loinville» et « Le domaine de la terreur». Dans ce processus, le téléporteur doit maintenir le «schéma» correct du sujet téléporté - de façon que la condition physique et psychologique du sujet rematéralisé soit identique à celle du sujet scanné.

Mais le processus connaît des ratés étranges. Ainsi, certains sujets réussissent à changer de position : cf « Lorsque le rameau se brise », « Le cri chuchoté », « Qui garde les gardiens » et « Ménage à Troi ». Toutefois, puisque ces changements sont probablement des erreurs de continuité imprévues, mieux vaut les ignorer. Ce sera plus difficile avec ce qui se passe dans « Le domaine de la terreur». C'est le seul épisode où l'on consent à nous montrer ce qui se passe d'un point de vue subjectif - et le résultat est surprenant. J'aurais pensé que ce serait un peu comme quand on est

anesthésié : on peut ressentir quelque chose de bizarre - une sorte de picotement - et avoir ensuite l'impression de se réveiller dans un autre endroit. Mais le lieutenant Barclay semble demeurer conscient pendant tout le temps de l'opération, chaque fois qu'il est téléporté. Qui plus est, cet être dématérialisé peut voir des choses dans le « flot de matière », tendre la main et les saisir ! Ces occurrences, si elles servent à pimenter le scénario, sont si improbables que nous devrions également les désavouer. C'est aussi ce que nous voyons dans « Darmok » où Picard, au milieu d'une bataille rangée avec la créature d'énergie, commence pour ainsi dire à se rematérialiser à bord de l'Entreprise-D avant qu'il ait fini de se dématérialiser à la surface d'El-Adrel.

D'autres anomalies sont à prendre plus au sérieux. Dans deux épisodes de La nouvelle génération, par exemple, le téléporteur est utilisé pour offrir une sorte de thérapie reconstituante. Dans « Sélection contre nature », la Dr Pulaski est contaminée par un virus qui altère l'ADN de ses victimes, et les vieillit en accéléré. Pulaski vieillit ainsi, mais échappe à la mort lorsqu'un échantillon de son ADN original - trouvé dans une mèche de cheveux - est transmis au téléporteur, qui la dématérialise, puis exploite les données de la mèche pour la reconstruire avec un ADN sain. Or non seulement elle est débarrassée du virus, mais elle est rendue à la condition physique qu'elle avait avant de contracter le virus: de nouveau, elle a la quarantaine épanouie. Le téléporteur, dans cet épisode, ne transmet pas seulement une matière passive: il peut effectuer des changements qualitatifs sur le sujet en utilisant des données fournies par une autre source. L'autre occurrence de cette thérapie reconstituante se produit dans « Petits coquins », mais de manière inverse. Cette fois, un accident de téléporteur rematérialise Picard, le civil Guinna, l'enseigne Ro Laren, et Keiko Ishiwaka à l'âge de l'adolescence - ils finissent par retrouver leur état physique antérieur lorsque les bonnes données génétiques sont transmises au téléporteur.

Essayons de trouver une signification à tout cela. Dans le cas de Pulaski, on aurait pu s'imaginer qu'elle serait matérialisée avec un ADN sain, mais resterait très vieille, comme avant la dématérialisation. Puisque votre condition physique dépend à tout moment de l'interaction de votre schéma génétique avec l'environnement, vous ne sauriez être « reconstituée » dans un état physique antérieur grâce aux seules données génétiques. Bien sûr, le téléporteur ne se contente pas d'utiliser ces données - il possède aussi le « schéma » de ses sujets. Mais celui de Pulaski, tel qu'il est entré dans le téléporteur lorsqu'elle est scannée, est celui d'une très vieille femme, et c'est donc ainsi qu'elle devrait être rematérialisée. Pulaski-à-la-quarantaine-épanouie ne peut être rematérialisée que si le téléporteur a le schéma de Pulaski à cet âge. Et pourtant, le chef mécanicien O'Brien nie explicitement qu'il y ait eu une « trace de téléporteur » antérieure dans le cas de Pulaski (j'imagine qu'il s'agit là d'une information génétique). Pourquoi nie-t-il? Nous y reviendrons.

Entre-temps, notez que dans les deux cas - Pulaski et le quatuor des « Petits coquins » - les sujets sont reconstitués dans leur état physique antérieur, mais que cela ne vaut pas pour leur état psychologique. Dans l'un et l'autre cas, les sujets se rappellent tout ce qui leur est arrivé dans l'intervalle de temps entre l'incident

malheureux de départ et leur «reconstitution». Puisque la condition physique a dû être reconstituée grâce au bon« schéma» (Pulaski-à-la-quarantaine-épanouie), et puisque l'état psychologique est tiré du «schéma» du sujet au moment où il est scanné (la-très-vieille-Pulaski), nous pouvons conclure que les schémas physique et psychologique sont - chose importante - indépendants.

Cela semble toutefois contredire ce que nous montre l'épisode « Seul parmi nous» (NG). L'Entreprise-D embarque par accident un être énergétique depuis une nébuleuse. Celui-ci finit par habiter Picard, auquel il devient très attaché. Il veut rentrer chez lui, dans la nébuleuse, en emmenant Picard avec lui, et décide d'utiliser pour cela le téléporteur. Il prend le commandement du vaisseau et dit à l'équipage de passerelle, impuissante : « Le téléporteur n'a pas besoin de reconstituer matériellement votre capitaine. Nous ne téléporterons que de l'énergie, et nous deviendrons un schéma combiné d'énergie tiré de nos formes de vie. » Picard est donc téléporté dans la nébuleuse comme « pure énergie», mais la symbiose prévue avec l'être d'énergie ne se produit pas, et Troi détecte l'énergie de Picard dans la nébuleuse, indépendante. Sur quoi Picard (toujours sous une forme de pure énergie) s'infiltré d'une façon ou d'une autre dans les circuits du vaisseau, et l'équipage concocte un plan pour le reconstituer, en utilisant le schéma matériel encore stocké dans la mémoire du téléporteur. La rematérialisation a lieu, et Picard sort de l'aire de rematérialisation, demande ce qui se passe, et dit que la dernière chose dont il se rappelle, c'est le moment où il se préparait à être téléporté.

Essayons de trouver un sens à tout cela. Comme nous l'avons noté, le téléporteur ne saurait téléporter que la matière du sujet (convertie en énergie) puisqu'il doit savoir comment reconstituer le sujet: la vraie téléportation, c'est la matière/énergie plus le schéma, et téléporter le sujet sous forme d'énergie pure laisserait de côté le schéma du sujet. Lorsque l'énergie de Picard regagne le vaisseau, le téléporteur la reconvertit en matière et la réunit avec le schéma de Picard. Jusqu'ici, tout va bien. Mais lorsque Picard est téléporté sous forme de pure énergie, est-ce simplement son schéma physique qui reste à bord, ou aussi son schéma psychologique ? L'épisode laisse supposer que les deux sont laissés en arrière; sans quoi, pourquoi Picard ne parvient-il pas à se rappeler ce qui s'est passé après le moment où il se préparait à être téléporté ? Mais si c'est vrai, que ressentait Troi lorsqu'elle avait l'impression que Picard était «à l'extérieur», dans la nébuleuse ? On ne peut pas jouer sur les deux aspects : ou le schéma psychologique de Picard a été dépêché dans la nébuleuse, ou il ne l'a pas été. S'il n'a pas été envoyé dans la nébuleuse, rien ne pouvait permettre à un être empathique de s'identifier avec Picard. C'est pourquoi une partie au moins du schéma psychologique de Picard a dû être envoyée dans la nébuleuse. Soit tout a été téléporté, et Picard souffre d'amnésie, soit seule la partie émotionnelle de son schéma psychologique a été téléportée avec l'énergie, et le reste - la partie qui se souvient - est restée dans la mémoire du téléporteur. Quoi qu'il en soit, une fois de plus, cela implique que le schéma physique et le schéma psychologique du sujet téléporté soient indépendants.

Jusqu'ici, il semble qu'il y ait une constante dans la téléportation ; la

rematériation consiste toujours à réassembler la matière originale. Mais ce n'est pas ce qui se passe dans « Sélection contre nature », « Petits coquins », et « Seul parmi nous ». Lorsque le téléporteur utilise le schéma d'ADN sain pour remplacer l'ADN endommagé de Pulaski, il a deux façons de procéder. L'une est de prendre la matière qui constitue l'ADN endommagé et d'y surimposer l'ADN sain. Mais pourquoi ne pas simplement jeter par-dessus bord l'ADN endommagé et se procurer une nouvelle matière pour l'ADN sain, dans une sorte de greffe ? Rien ne l'interdit dans le canon de Star Trek. Voyez pour exemple les duplicateurs du vaisseau - qui servent à produire toutes sortes de biens matériels, comme la nourriture. Ces duplicateurs se procurent de la matière/énergie de quelque source inexplicée, et y surimposent une configuration physique particulière. Pourquoi le téléporteur, qui est très proche de leur nature, n'en fait-il pas autant ? En admettant cette possibilité, on interprète de façon intéressante la revendication d'O'Brien, qui dit ne pas avoir de « trace de téléporteur » antérieure pour Pulaski. Il en a forcément trouvé une pour reconstituer le schéma physique de Pulaski-dans-la-quarantaine-épanouie, mais le cache délibérément parce que, pour reconstituer une trace, il faut se procurer de la matière nouvelle (le cerveau de la très-vieille-Pulaski aurait perdu bien des neurones que le cerveau de l'autre Pulaski possédait), et quelqu'un a peur des implications sociales et éthiques de l'utilisation du transporteur - des implications touchant la duplication, que nous examinerons plus à fond au chapitre 5. La reconstitution du quatuor dans « Petits coquins » requiert également qu'on se procure un surcroît de matière, puisque les sujets dématérialisés étaient bien plus grands que les adolescents matérialisés. À présent, nous avons une interprétation alternative de « Seul parmi nous ». L'histoire assez improbable de l'énergie de Picard, qui chercherait un chemin pour retourner dans les circuits du vaisseau, est une mystification. L'énergie sans schéma de Picard s'est tout simplement dissipée lorsque la combinaison anticipée avec l'être d'énergie n'a pas marché. Mais puisque le schéma de pré-téléportation de Picard a subsisté dans la mémoire du téléporteur, c'est qu'il a simplement été surimposé sur une matière nouvellement acquise.

Si vous avez des doutes sur l'interprétation machiavélique de ces épisodes, en voici trois autres qui démontrent absolument que le téléporteur peut acquérir de la matière neuve en vue d'une rematériation. Il arrive qu'une double rematériation se produise à la suite d'une dématériation simple, comme dans « L'ennemi intérieur » (PS), où le capitaine Kirk est scindé entre ses personnes cérébrale et animale; dans « Une seconde chance » (NG) où Riker est scindé en deux jumeaux qualitativement identiques; et dans « Tuvix » (V) où le chef de la sécurité Tuvok et le cuisinier du vaisseau, Tuvix, ayant fusionné en un seul individu, sont ensuite de nouveau séparés. Si la rematériation exploite seulement la matière originelle, un tel dédoublement est impossible. Ou la matière a été elle-même divisée, et on a acquis de la matière neuve pour chaque « jumeau », ou bien un des « jumeaux » reçoit la matière originelle, et l'autre ne reçoit que de la matière acquise, ou encore les deux « jumeaux » reçoivent de la matière acquise. Quel que soit le cas, il a fallu acquérir de la matière neuve.

Cela suggère un mode de fonctionnement du téléporteur entièrement différent: la téléportation des seules données. Puisque nous savons que le téléporteur peut acquérir de la nouvelle matière et y surimposer un schéma, pourquoi ne pas simplement utiliser de la matière acquise ? Il suffirait de scanner le sujet, d'envoyer les données à destination, et d'y rematérialiser un individu à base de matière neuve. Puisqu'il faut, de toute façon, envoyer les données requises pour la rematérialisation - données nécessaires à la production du bon corps et de la bonne psychologie - pourquoi ne pas travailler sur elles seules ? Le problème que pose ce processus a déjà pu vous venir à l'idée: que faire du sujet original, resté à bord ? Et quitte à le tuer sans douleur, en faisant usage du processus de scanning, ne nous faut-il pas nous débarrasser du corps ? Oui, bien sûr - par évaporation, peut-être. Bien que ce ne soit pas strictement nécessaire, c'est plus propre, et en accord avec le monde de Star Trek. Y compris avec l'existence (irréaliste ou non) des armes évaporisatrices que nous observons dans la première série (l'androïde Andréa évaporise l'androïde Kirk dans « De quoi sont faites les petites filles ? ») et dans La nouvelle génération (Riker utilise son phaseur pour évaporer les clones surgissant dans « L'ascension de la grande échelle »). De fait, si quelques opérations de téléporteur dans Star Trek semblent requérir une matière neuve, tous les exemples observés concordent avec la téléportation par données et l'acquisition totale de matière. Il se peut que dans la dématérialisation il se produise toujours une conversion en énergie, qui se dissipe avec le seul envoi du schéma et justifie nombre des séquences de vieillissement, comme dans « Reliques » (NG) où Scotty est ramené du téléporteur à bord du Jenolan soixante-quinze ans après sa dématérialisation ! On peut se figurer que le téléporteur a gardé le schéma de données de Scotty pendant soixante-quinze ans, mais on a plus de mal à voir comment le téléporteur pourrait conserver la matière-convertie-en-énergie de Scotty « dans la glace » pendant tout ce temps.

Pour résumer les questions qui apparaîtront dans la suite de ce chapitre, nous nous contenterons d'examiner la téléportation des données en excluant deux possibilités (analysées au chapitre 5). L'une, c'est le dédoublement façon Kirk ou Riker, l'autre c'est la matérialisation d'un sujet qualitativement différent du sujet scanné, façon Pulaski ou le quatuor des « Petits coquins ». Nous ne considérerons dans ce chapitre que deux types de téléportation. Téléporter la matière consiste à scanner, dématérialiser, transmettre schéma et matière, et rematérialiser cette même matière. Téléporter les données consiste à scanner, détruire la matière première, transmettre le seul schéma, et rematérialiser en utilisant seulement de la matière nouvellement acquise. Maintenant, il nous reste deux questions à poser. Tout d'abord, est-il rationnel de se soumettre à une téléportation de matière, avec l'idée qu'on y survivra ? Puis, est-il rationnel de se soumettre à une téléportation de données en croyant qu'on y survivra ?

Témoignages recevables

Imaginez que vous ayez l'occasion d'être téléporté(e) sous l'une ou l'autre

forme, à la surface d'une planète sûre et agréable, pour des vacances gratuites. Imaginez encore que je vous persuade facilement que les deux processus sont fiables technologiquement. Pour que vous fassiez un choix en toute connaissance sur le processus que vous choisirez, je vous présente des centaines de témoignages provenant de sujets rematérialisés, de leurs amis et de leurs proches. Tout d'abord, les sujets rematérialisés rapportent tous qu'il se rappellent avoir pénétré dans le scanner, mais n'ont ensuite aucun souvenir jusqu'à leur réveil. Ils rapportent qu'ils se rappelaient au réveil leur petit déjeuner de ce matin-là; qu'ils le sentaient encore au creux de leur estomac; qu'ils se rappelaient le but du voyage et désiraient l'accomplir; qu'ils voulaient rentrer pour retrouver leurs proches. Ils se rappellent avoir éprouvé une certaine angoisse avant la procédure, un soulagement nerveux au réveil, etc. Ensuite, dans des sondages élargis, les parents et amis rapportent tous qu'après rematérialisation tout reste comme à l'habitude. La question, en ce qui vous concerne, est donc: d'après ces informations, opteriez-vous pour une téléportation par matière? Par données? Ni l'une ni l'autre?

Une majorité écrasante des gens que j'ai interrogés rejette l'idée d'une téléportation par données - presque personne n'est tenté par cette option - et si certains se laissent un peu plus convaincre par une téléportation par matière, ils sont à peine une majorité à accepter de s'y soumettre. Dans presque tous les cas de refus d'une téléportation par données, la raison est la même - ils craignent que le sujet scanné ne survive pas à la procédure, mais soit remplacé par une sorte de clone. On peut dire que le sens commun suggère que vous ne sauriez rationnellement survivre à la téléportation par données (de fait, on n'y voit pas une véritable forme de transport), et se divise sur la question de savoir si on peut s'attendre ou non rationnellement à survivre à la téléportation par matière. Moi qui joue l'avocat du diable, j'essaierai de montrer que le sens commun se trompe doublement. Tout d'abord, il n'y a pas de différence notable, en ce qui concerne la survie personnelle, entre la téléportation par matière et la téléportation par données ; puis, étant donné les informations précédentes, il n'y a pas de raison rationnelle d'être pessimiste quant à la survie à l'un ou l'autre processus. Ma première stratégie passera par la négation : j'essaierai de montrer que ces intuitions de bon sens sur la téléportation ne concordent guère avec d'autres intuitions de bon sens sur la survie personnelle. J'accuserai donc le sens commun d'incohérence. Nous nous intéresserons tout spécialement au fait que ces deux intuitions de bon sens apparaissent également dans les attitudes des personnages de Star Trek.

Presque assez, c'est bien assez

Commençons par un cas ordinaire de survie personnelle. Vous êtes aujourd'hui assis dans une chaise avec ce livre. Nous supposerons très banalement que la personne assise dans cette chaise il y a dix secondes était numériquement identique (c'est-à-dire: est l'unique et même personne) à celle qui y est assise à cet instant précis. La survie personnelle, d'ordinaire, implique une identité personnelle qui résiste au temps.

Je pense que la vision de bon sens de la survie personnelle revient tout simplement à poser une identité personnelle qui résiste à la durée. Qu'implique-t-elle d'autre? L'identité physique et mentale, ou psychologique, résistant à la durée - en d'autres termes, un corps et un esprit qui demeurent les mêmes.

Prenons l'identité corporelle qui résiste au temps. S'il est vrai qu'elle existe tout au long de votre vie, le corps que vous avez maintenant est exactement le même qu'à votre naissance. Mais c'est absurde!, pourrait-on m'objecter. Notre corps actuel est bien différent - plus grand, plus velu, ridé un peu partout ! Ce ne sont pourtant que des différences de surface. La science nous apprend que, sur une période de sept ans, toutes les cellules du corps, exception faite des neurones, sont remplacées. Votre corps n'a aujourd'hui plus rien de commun avec votre corps à la naissance, ou il y a seulement dix ans. Comment peuvent-ils être identiques? Le même problème se pose pour l'identité psychologique. On a clairement un esprit à l'âge de deux ans, mais maintenant vous ne connaissez pratiquement plus aucun des états mentaux que vous aviez à l'âge de deux ans, et vice versa. Comment pourriez-vous être psychologiquement identique à ce que vous étiez à l'âge de deux ans ?

Tout cela remet en question le bon sens selon lequel la survie personnelle ordinaire implique une identité personnelle qui perdure. Si la personne est composée d'un corps et d'une psychologie, entre le vous d'aujourd'hui et le nourrisson que vous étiez, comment peut-il y avoir identité? La réplique tentante est qu'il existe quelque chose qui constitue une personne au-delà de «l'habillement» physique et psychologique. Nous appellerons ce supplément l'essence ou l'âme. (Parler ici de l'« âme» peut susciter la confusion: dans certaines perspectives, l'âme n'est rien d'autre que l'esprit ou la psychologie; dans d'autres, elle est autre chose.) Si l'âme, contrairement au corps et à l'esprit, ne change pas au cours du temps, elle peut, à coup sûr, nourrir l'identité d'une personne à travers le temps.

Star Trek encourage parfois cette vision des choses. Dans « Le temps au carré» (NG), lorsque Picard se retrouve dédoublé, Troi déclare au capitaine que le clone « est tout autant Jean Luc Picard que vous-même». Or Troi est dotée d'empathie, ce qui lui permet de détecter l'identité personnelle. Si elle se contentait de recueillir les pensées ou les sentiments du clone, elle aurait déclaré qu'il sent ou pense être Picard. Le fait qu'elle affirme catégoriquement qu'il y a là une identité (et au chapitre 6, nous étudierons la possibilité que les deux Picard soient bel et bien identiques) prouve une capacité à détecter quelque chose d'autre - l'essence ou l'âme de Picard.

Mais nous n'avons pas besoin de nous référer à l'âme pour expliquer l'identité personnelle - l'existence de l'âme nous est probablement inutile, comme nous le verrons plus loin. Pour illustrer la façon dont on peut justifier l'identité en l'absence d'âme, laissons tomber les individus et concentrons-nous sur des objets matériels ordinaires. Supposez que vous ayez une montre à l'ancienne, comme celle de Samuel Clemens dans « La flèche du temps I & II», avec une mécanique, un remontoir, un ressort, etc. Un jour votre montre ne marche plus, et vous l'apportez à un joaillier qui change le ressort. Le fait que votre montre ait maintenant un ressort différent vous

amène-t-il à conclure que ce n'est plus la même montre? A-t-elle cessé de vivre pour être remplacée par une nouvelle montre? Certes non (et si votre montre originelle n'existe plus, on peut se demander quel droit vous avez sur la nouvelle). L'identité des objets matériels quotidiens ne pose pas problème, semble-t-il, malgré les petits changements progressifs. Si, par la suite, vous faites changer le remontoir, puis le cadran, etc., au bout de dix ans, tout le mécanisme intérieur aura été remplacé. Puis supposez que le boîtier et le verre doivent être changés. Vous risquez de vous retrouver avec une montre composée intégralement d'éléments neufs, tout en ayant l'impression de posséder la même qu'auparavant. Soit maintenant une entité plus abstraite, par exemple votre équipe favorite de base-ball. Verra-t-on jouer les Orioles de Baltimore et les Yankees de New York au XXIV^e siècle, si le championnat de base-ball existe encore à cette époque? Oui : en quelques années (ce qui est peu), l'équipe tout entière et les cadres sportifs auront été peu à peu remplacés ; et pourtant l'équipe sera encore la même. La nef de Thésée était composée de planches qu'on remplaçait à mesure qu'elles pourrissaient ; après bien des années, toutes les planches d'origine avaient été remplacées, mais nous pensons qu'il s'agissait du même navire. De même, nous pouvons nous figurer que l'Entreprise-D subit réfection sur réfection jusqu'à ce que toutes ses parties originelles auront été remplacées, sans qu'il perde son identité. Comment est-ce possible, puisque les montres, les équipes de baseball, les nefes et les astronefes n'ont pas d'âme ?

Le philosophe Robert Nozick a présenté un schéma rendant compte de tous ces jugements sur l'identité ordinaire des objets, qu'il appelle schéma de plus grande continuité d'une identité numérique. Cette théorie est simple, mais il convient d'introduire ici deux notions. La première est celle de l'objet-à-la-fois (une tranche temporelle d'un objet, abrégée en tranche d'objet), la seconde, celle de la similarité qualitative. Lorsque nous comparons deux objets - moi et vous, par exemple - nous devrions réaliser que, très souvent, une composante temporelle y est intégrée, si bien qu'en fait nous ne faisons que comparer la tranche d'objet moi-actuel avec la tranche d'objet vous-actuel. Nous pourrions aussi comparer des tranches d'objet telles que moi-il-y-a-dix-ans avec vous-actuel. Et nous avons déjà comparé les tranches d'objet vous-actuel et vous-à-la-naissance. Recourons aux noms alphanumériques pour différencier les tranches : nous pourrions nommer le bébé engendré par mes parents en 1957 H1, et le moi-actuel H2.

La seconde notion est celle de similarité qualitative, qui ne dépend que de la proportion des attributs partagés par les objets ou tranches d'objet. Deux chaises, deux jumeaux « identiques » partagent nombre de leurs attributs et sont donc relativement similaires qualitativement. Une chaise et un être humain ont en commun relativement peu d'attributs, et guère de similarité qualitative. Comme nous l'avons déjà établi, H1 et H2 ne sont pas similaires qualitativement, même si nous considérons d'ordinaire que H2 est numériquement identique à H1. Je pense que la meilleure façon d'expliquer ce jugement est de poser une relation entre les deux tranches d'objet - car ce sont les tranches d'un seul et même objet. (Notez que les jugements d'identité ordinaire concernent en général l'identité d'objet, mais que nous pouvons aussi en

faire sur l'identité de tranches d'objet.) Si la relation d'identité d'objet est maintenue entre H1 et H2, ce n'est pas le cas pour la relation d'identité de tranche d'objet: H1 et H2 sont deux tranches distinctes du même objet.)

Comment portons-nous des jugements sur l'identité des objets? Une tranche d'objet a une descendance causale ou des éléments de continuité, qui interviennent par la suite. Juste après que votre montre a reçu un nouveau ressort, il y a deux éléments de continuité importants de la tranche M1, soit la montre d'avant la réparation. L'une est M2, l'ancien ressort, et l'autre est M3, la montre réparée, qui intègre le nouveau ressort. (Il y aura d'autres menus éléments de continuité, tels le minuscule copeau de métal détaché du boîtier lorsque le joaillier l'ouvre en le forçant avec une lame.) Laquelle de ces tranches est identique à M1? La première étape consiste à comparer les similarités quantitatives entre M1 et M2, et entre M2 et M3. Puisqu'il y a une similarité qualitative bien plus grande entre les deux derniers, M3 remporte facilement la confrontation: c'est l'élément de continuité le plus proche de M1. Voilà qui est compatible avec notre jugement, à savoir que M3 est identique à M1. De même, supposez qu'on change un cristal de dilithium dans l'Entreprise-D, Appelez l'astronef originel E1, l'astronef modifié E2 et l'ancien cristal E3. Il est clair que E2 remporte la palme en matière de similarité qualitative entre ces éléments de continuité, et nous en concluons qu'il y a une identité d'objet entre E1 et E2. De même, votre équipe de baseball préférée peut remplacer son lanceur - quand ce serait Carl Ripken - sans perdre son identité. (On notera que nous disons parfois que les individus connaissent une crise d'identité, et les Orioles de Baltimore peuvent en passer par là si Ripken doit être remplacé. Lorsqu'un événement nous amène à repenser l'orientation de notre vie, nous avons un sentiment de décalage et d'incertitude. Mais la «crise» se produit parce que nous ne savons pas bien quels changements opérer, non parce que nous nous demandons si nous avons cessé d'exister, ou cesserons d'exister. Laissons donc de côté ce sens non littéral d'«identité».)

Le schéma de plus proche continuité demande à être un peu précisé. Tout d'abord, nous comparons normalement tous les éléments de continuité, et seulement ceux d'une tranche d'objet à un moment particulier: normalement, nous ne comparons pas l'élément de continuité propre à un instant donné avec un autre, pour voir lequel l'emportera. Puis, comme nous le verrons au chapitre 5, nos jugements semblent nécessiter que l'élément de continuité le plus proche l'emporte continuellement sur les autres, et avec une avance appréciable - il doit l'emporter de loin. Enfin, être élément de continuité le plus proche d'une tranche d'objet donnée ne suffit pas à s'identifier à elle. L'élément de continuité le plus proche doit aussi être suffisamment proche. Il serait plus correct de parler de nos jugements en matière d'identité numérique en nous basant sur un schéma qui soit à la fois le plus proche et aussi suffisamment proche. Pour illustrer cette règle, imaginons que nous mettions la nef de Thésée en cale sèche, qu'elle prenne feu et brûle entièrement. Le grand monceau de bois brûlé et de cendres qui nous reste est de loin l'élément de continuité le plus proche de la nef originale, mais nous dirons néanmoins que celle-ci n'existe

plus: l'entassement de charbons et de cendres ne lui est plus identique. De même, si vous croyez qu'en mourant vous cessez d'exister, il vous faut admettre que votre cadavre est de loin le plus proche élément de continuité de votre être mourant, et nier pourtant qu'il en soit assez proche pour être considéré comme identique à votre personne. La façon évidente de donner un sens à ces intuitions est de noter que, pour être identique à la nef précalcinée de Thésée, l'élément de continuité le plus proche doit être lui-même une tranche de 'navire. Et pour être identique à votre être mourant, l'élément de continuité le plus proche doit être une tranche de personne. Enfin, comme on l'a souligné, que l'une des tranches d'objet soit identique à une autre, qu'elles soient qualitativement très similaires n'est pas une condition nécessaire - sans quoi mon « moi-actuel » ne saurait être identique à mon « moi-à-la-naissance », H1 ne saurait être identique à H2. Il nous faut ajouter une observation de plus à notre argumentation. Il devrait être évident à présent que H1 et H2, quoique très dissemblables qualitativement l'un par rapport à l'autre, ne sont pas moins connectés d'une façon intéressante: il y a une chaîne causale d'autres tranches d'objet entre eux - chaque tranche étant l'élément de continuité le plus proche du maillon précédent. Aussi longtemps que deux tranches d'objet appartiennent à la même entité (une personne, par exemple) et qu'elles sont reliées par la bonne chaîne (une chaîne des éléments de continuité les plus proches reliés par une très forte similarité qualitative), alors cette relation est préservée. Ce serait le cas même si la science faisait la remarquable découverte que nous recyclons toutes nos cellules sur une période de sept secondes et non de sept années.

Nous avons maintenant une façon pratique d'exprimer les conditions nécessaires et suffisantes de l'identité corporelle, mentale et personnelle. La tranche de corps C2 est identique à la tranche de corps C1 si et seulement si C2 est l'élément de continuité le plus proche de C1 ; la tranche d'esprit E1 est identique à la tranche d'esprit E2 si et seulement si E2 est le continueur le plus proche de E1 ; la tranche de personne P2 est identique à la tranche de personne P1 si et seulement si P2 est l'élément de continuité le plus proche de P1. Pour répéter une observation précédente, la survie personnelle ordinaire n'implique pas seulement une identité personnelle qui perdure, mais aussi une identité corporelle et psychologique qui résiste au temps, et nous sommes heureux de pouvoir donner un sens à cette hypothèse.

La question suivante est: en quoi consiste une identité personnelle qui perdure? L'identité personnelle qui perdure est-elle une question d'identité personnelle, psychologique, ou les deux, ou ni l'une ni l'autre (mais plutôt d'identité de l'âme)? L'idée de bon sens qui dit la téléportation par matière plus viable que la téléportation par données suggère que la permanence de l'identité corporelle est importante. Si c'est le cas, il y a une forte analogie entre les gens et les objets ordinaires. Prenons à nouveau la montre de Sam Clemens. Il est clair qu'elle survit à son démontage et à son remontage - parfois, il peut s'avérer utile de démonter un objet avant de l'envoyer quelque part. Mais supposons qu'au lieu de démonter la montre le lieutenant Worf l'évaporise avec son phaseur, puis la duplique en une montre qualitativement identique.

Cette seconde montre n'est en rien un élément de continuité de l'original, n'est-ce pas? C'est une simple réplique de la montre originale - une montre numériquement distincte, quoique tout à fait similaire. Cette analogie nous suggère que la matière compte pour l'identité personnelle.

Mettons à l'épreuve cette intuition en examinant dans Star Trek une série de situations inutiles fondées sur des greffes corporelles. Pour éviter les inconséquences, supposons que dans chaque cas le patient peut se permettre cette opération (grâce à l'excellent plan médical de la Flotte) et qu'elle est relativement indolore. De plus, dans tous les cas, la procédure est fiable car l'opération a un taux élevé de réussite technique. Pour vous faire une opinion informée, des centaines de témoignages vous viennent non seulement des sujets post-opératoires mais de leurs amis et de leurs proches. Les sujets post-opératoires disent systématiquement se souvenir de tout ce qui s'est passé jusqu'à ce qu'on les anesthésie - ce qu'ils se rappellent ensuite, c'est leur «réveil» après l'opération. Ils se rappellent n'avoir pas eu de petit déjeuner ce matin-là, ils évoquent leurs décisions précédentes, disent vouloir les accomplir, et signifient leur désir de retrouver leurs proches. De plus, dans les sondages élargis, les amis et parents ne rapportent rien d'extraordinaire chez le sujet post-opératoire. En ce qui vous concerne, étant donné ces informations, la question reste : quelle procédure choisir rationnellement dans chacun des cas suivants?

Fonctions corporelles

Mais je suis Roger Korby !

L'androïde Korby à Kirk dans « De quoi sont faites les petites filles ? »

Quatre scénarios de Star Trek nous aideront à poser les choix forcés. Dans « Le piège samaritain », puis dans « Tapisserie » (NG), nous apprenons que, alors qu'il était frais émoulu de l'Académie de la Flotte interstellaire, Jean-Luc Picard s'était commis dans un combat avec trois Nausicaains qui s'était achevé par un coup de poignard reçu au cœur: il avait alors bénéficié d'un implant cardiaque artificiel. On a eu raison de remplacer son cœur endommagé par un cœur artificiel en bon état de fonctionnement, et il est probable que c'est ce que Picard aurait souhaité dans ces circonstances (sans doute était-il inconscient, et quelqu'un d'autre a décidé pour lui). Si vous deviez choisir entre la mort à coup sûr et la vie avec un implant artificiel qui remplit exactement le même office, n'opteriez-vous pas pour l'implant ?

Dans « Assistance vitale » (PE9), Vedek Bareil a un problème organique plus grave : son état requiert un traitement qui détruira progressivement ses organes essentiels pour les remplacer tous, ou presque, par des organes artificiels. Une fois de plus, chaque fois que c'est possible, l'implant est l'unique solution s'il faut choisir entre une mort assurée et une vie se poursuivant au même niveau fonctionnel. À la fin du XX- siècle, refuser les implants pourrait avoir un sens - en raison de la douleur,

des inconvénients et du manque de fiabilité du processus. Au XXIV- siècle, ces considérations n'entrent plus en ligne de compte: il est raisonnable d'avoir recours aux implants même si on doit se retrouver avec un corps largement artificiel.

Jusqu'ici ces considérations ne contredisent pas l'importance que nous accordons à la matière puisque, si l'on s'en tient au schéma de l'élément de continuité le plus proche pour l'identité corporelle, ces remplacements partiels (et graduels) préservent l'identité corporelle. Poussons le bouchon encore un peu plus loin. Supposons que Bareil ne souffre pas d'une dégradation progressive, mais d'une déficience organique généralisée attaquant tous les organes importants, hormis le cerveau. La procédure d'implantation totale demeure-t-elle une option intéressante? Une fois celle-ci accomplie, il y aura deux éléments de continuité importants du corps préopératoire de Bareil: d'un côté le corps, composé du cerveau originel, de quelques membres originels, et des implants d'organes majeurs; de l'autre, le gros tas d'organes u rebut. Mais de ces deux éléments, un seul est un corps. - l'identité corporelle est donc probablement maintenue dans la procédure d'implants.

Allons un peu plus loin encore : si tous les organes importants excepté le cerveau s'avèrent déficients, pourquoi ne pas se contenter de transplanter le cerveau sain dans un corps artificiel? (On pourrait y ajouter le système nerveux central.) Nous discernons maintenant deux versions différentes de la conception qui fait primer la matière. Tout d'abord, il y a celle qui accorde de l'importance au corps: l'identité personnelle requiert l'identité corporelle (dans laquelle le corps inclut le cerveau). Puisque, dans le cas d'une greffe de cerveau, le corps décérébré de Bareil semble être un élément de continuité plus proche que le cerveau de Bareil, et puisque le corps décérébré ne saurait prétendre à la qualité de personne, cette version affirme que Bareil cesse d'exister après la greffe de cerveau. L'alternative, c'est la vision qui fait primer le cerveau: l'identité personnelle dépend de l'identité du cerveau. Selon cette vision, la greffe de cerveau (en supposant qu'elle réussisse) préserve bel et bien l'identité personnelle: Bareil va où va son cerveau.

Poussons les choses encore un peu plus loin. Dans « Assistance vitale», lorsque la dégradation organique de Bareil gagne le cerveau, le Dr Bashir envisage de remplacer certains éléments dudit cerveau par des « implants positroniques ». (Nous supposerons, pour la logique de l'argumentation, que ces implants sont les exacts équivalents fonctionnels des éléments originels du cerveau.) Si le choix, à chaque étape de la dégradation, se fait entre une mort certaine et la vie maintenue au même niveau fonctionnel, ne faut-il pas opter pour les implants artificiels? Supposons que Bashir suive la procédure par étapes, en remplaçant à chaque fois moins de 10 % du cerveau. Cautionner cette procédure, c'est adopter la vision qui fait primer le cerveau, puisque, dans le schéma de l'élément de continuité le plus proche, l'individu pourvu d'implants positroniques sera, à chaque étape, un élément de continuité plus proche que la portion de cerveau mise au rebut. Mais dans l'épisode, Bashir remplace la moitié du cerveau, procédure qui requiert à coup sûr le stockage temporaire des informations psychologiques de Bareil - il aurait pu aussi bien remplacer le cerveau dans sa totalité. Dans la perspective qui fait primer le cerveau, qu'on remplace la

moitié du cerveau ou le cerveau tout entier, l'individu, à l'issue de l'opération, ne possédera pas l'élément de continuité le plus proche du cerveau préopératoire de Bareil, et dès lors cet individu n'est pas Bareil.

Bashir semble faire primer le cerveau, comme nous l'avons remarqué au chapitre 2. Il accomplit de bon gré la greffe d'organes, mais rechigne à toucher au cerveau, car il craint que Bareil ne perde « l'étincelle de vie » qui, elle, ne saurait être reproduite. Le major Kira Nerds, porte-parole de Bareil (alors inconscient), dit que celui-ci préférerait des implants positroniques, et Bashir répond: « Kira, si je poursuis l'opération, l'homme qui se réveillera ne sera peut-être pas celui que vous connaissiez. » Bashir pourrait donc accepter la forme de fonctionnalisme cérébral et mental évoquée aux chapitres 2 et 3. Il se préoccupe du résultat de la procédure d'implantation: elle pourra produire une nouvelle personne, numériquement distincte de Bareil.

Poussons les choses un peu plus loin. Dans « De quoi sont faites les petites filles? », le fiancé de l'infirmière Christine Chapel, l'archéologue Roger Korby, a été porté disparu depuis cinq ans lorsque la Flotte reçoit un message d'Exo III, émanant apparemment de Korby lui-même. Elle envoie l'Entreprise sur ses traces: Kirk et Chapel sont téléportés sur la planète (accompagnés de deux officiers de sécurité, dont l'anonymat suggère qu'ils sont promis à une fin prochaine) à la rencontre de Korby, qui affiche une grande forme physique. Mais il est blessé, et l'on découvre à cette occasion qu'il est androïde. Il explique que le véritable Dr Korby est tombé sur les ruines d'une civilisation antique technologiquement avancée, dont les chefs produisaient des androïdes pour leur servir d'esclaves. L'un de ces androïdes, appelé Ruk, était encore en bon état de fonctionnement, et il est devenu le domestique de Korby. Au cours de ses études archéologiques, Korby est mort d'une maladie fatale et inexplicable. L'unique solution était de permettre à Korby de transférer sa psychologie dans un corps androïde. « Vous ne sauriez imaginer ce que c'était, dit-il à Christine Chapel. J'étais gelé, à l'agonie ... je ne sentais plus mes jambes. Il n'y avait plus que mon cerveau entre la vie et la mort. » Voyant un doute sur son visage, il plaide: « Je suis le même qu'avant, Christine, et je suis peut-être meilleur qu'avant. » Si l'attitude de Christine est équivoque, celle de Kirk face à ces révélations est tout à fait claire. Lorsque Stock demande des nouvelles de Korby, Kirk répond: « Le Dr Korby n'a jamais vécu ici »: il suppose que Korby est mort bien avant l'atterrissage sur Exo III. Kirk considère que Korby n'a pas survécu à la procédure de duplication.

Quel processus Roger Korby a-t-il exactement subi avec l'aide de Ruk? Dans ce même épisode, on construit une réplique androïde de Kirk (contre son gré) en clonant son corps, puis la psychologie de Kirk y est reproduite: nous pouvons donc présumer que Korby a lui aussi été dupliqué physiquement, et que sa psychologie a été reproduite dans un corps androïde. Ce processus a eu pour conséquence immédiate la mort physique de Korby: détail important, son cerveau est mort avec son corps. Vous remarquerez que la procédure à laquelle s'est soumis Korby est une ultime extension logique de celle que Bashir a fait subir à Vedek Bareil. Nous avons déjà imaginé une greffe de cerveau sur un corps artificiel décérébré, et un remplacement total de

cerveau par des composants artificiels. Si nous concilions ces deux procédures, le résultat équivaut à l'opération effectuée par Ruk sur Korby.

Ne te connais surtout pas toi-même

C'est encore moi, Christine! ... Roger! Je suis là-dedans!

**L'androïde Korby à l'infirmière Christine Chapel,
dans « De quoi sont faites les petites filles ? »**

Si l'une ou l'autre des conceptions faisant primer la matière est juste, il est possible de savoir absolument qui on est... et de se tromper. Non seulement l'androïde Korby se trompe lorsqu'il pense être vraiment Roger Korby, mais les commentaires de Bashir suggèrent que l'individu mouvant et parlant qui résultera d'une substitution totale de cerveau pensera à tort qu'il est Bareil. Si quelqu'un doit savoir qui est tel ou tel individu, il ne faut pas penser que c'est bien l'individu lui-même! La possibilité d'une telle erreur est magnifiquement illustrée dans les épisodes de Star Trek que voici.

Dans «L'héritage» (NG), Data apprend avec un étonnement ravi que la scientifique qui leur rend visite, la Dr Juliana Tainer, est l'ancienne épouse de son créateur, le Dr Noonian Soong, et qu'elle a de nombreuses informations sur sa création et sa « petite enfance »; Au fil des jours, Data remarque certains détails bizarres sur sa « mère », et ses soupçons se confirment lorsqu'elle a un accident - son corps inconscient est manifestement celui d'un androïde. Data accède à un programme holographique implanté dans son cerveau positronique - par le Dr Soong lui-même. Celui-ci y explique que, lorsque son épouse a été blessée mortellement et est tombée dans le coma, il a construit un corps androïde et y a transféré sa psychologie, sans en informer «Juliana» à son réveil ! Le Dr Tainer est donc un androïde qui s'ignore! Qui plus est, il s'agit là d'un androïde qui croit être l'ex-épouse du Dr Soong, mais, si nous nous en tenons à la perspective selon laquelle c'est la matière qui prime, ce ne peut être vrai. Data décide de ne rien révéler à «Juliana».

Cet épisode fait écho, par son thème, à l'épisode «Les survivants» (NG), où l'Entreprise-D croise une planète dénuée de toute forme de vie, hormis un vieux couple, Mr et Mrs Uxbridge. Ce couple explique que leur planète a été attaquée par une espèce baptisée les Husnaks, qui a exterminé la population tout entière, sauf eux. En fait, Kevin Uxbridge a été l'unique survivant du massacre; c'est un «douwd» éternel et omnipotent qui est venu vivre parmi les mortels, et est tombé amoureux d'une mortelle.

Après l'attaque des Husnaks, Kevin a « recréé » sa femme (par simple pensée, alors que dans le cas de Korby et de Juliana, il fallait une technologie appropriée), sans jamais rien en dire à Mrs Uxbridge à qui Picard déclare, néanmoins, sans mâcher ses mots: «Je peux vous toucher. J'entends votre voix. Je sens votre parfum. Vous êtes à tous égards une personne réelle, avec vos opinions propres. Mais vous n'existez pas. Vous êtes morte avec les autres membres de la colonie. » Pris au sens littéral,

l'énoncé de Picard est incohérent : je suppose qu'il veut dire par là que l'épouse de Kevin n'existe plus, et que la personne qui lui fait face est une tout autre personne, qui n'est née que lorsque Kevin l'a « conçue » en pensée.

Une fois de plus, si Picard a raison (s'il assume la perspective selon laquelle c'est la matière qui compte), cette personne se trompe en pensant savoir qui elle est.

Enfin, dans « Rumeurs », le commandant Miles O'Brien retourne à la base et trouve que tout le monde agit bizarrement.

Il semble victime d'un délire paranoïaque où des gens familiers auraient été enlevés et remplacés par des clones hostiles. Tout le nœud du scénario - et il est vraiment très réussi - c'est qu'en fait, c'est lui l'imposteur: il est un clone du vrai O'Brien, et il a été dépêché par l'infâme Dominion, qui a kidnappé O'Brien. Ce que démontre l'épisode (comme nombre d'autres scénarios de séries télévisées), c'est que tout être programmé ou pourvu de souvenirs bidons y verra de véritables souvenirs, et se croira semblable à ceux qui l'entourent.

L'esprit surpasse la matière

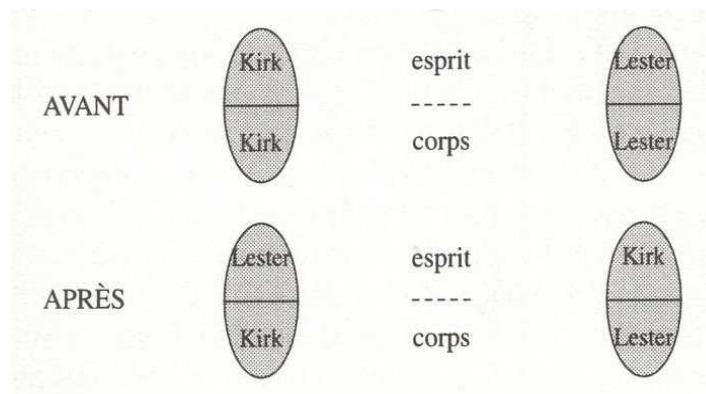
Corps de Kirk: « Si je comprends bien, vous affirmez être le capitaine Kirk ? »

Corps de Lester: « Je ne suis pas le capitaine Kirk. Ça se voit clairement. Ce que j'ai dit, c'est que tout ce qui faisait du capitaine Kirk un être vivant, spécifique, est contenu dans ce corps. »

Dans « Volte-face », le Dr Janice Lester provoque un échange inhabituel avec Kirk : par un vieux procédé, l'essence de Lester passe dans le corps de Kirk, et vice versa (voir diagramme ci après). Bien que McCoy ne puisse prouver scientifiquement l'échange, l'équipage de l'Enterprise commence à avoir des soupçons, qui deviennent des certitudes. Pourquoi ? À cause du comportement des deux personnes, notamment celui de la personne qui occupe le corps de Kirk (en bas à gauche dans le diagramme), surtout lorsqu'elle donne l'ordre d'exécuter les officiers supérieurs.

L'équipage en conclut que cette personne est en fait Lester, et que Kirk occupe le corps de Lester. Maintenant voyons comment cela a dû se passer pour la personne qui occupe le corps de Lester (en bas à droite dans le diagramme). Elle se serait « réveillée » dans le corps de Lester, mais avec tous les souvenirs et autres caractéristiques psychologiques de James T. Kirk.

Si vous vous réveilliez dans cet état (et ne tombiez pas aussitôt dans la folie), n'en concluriez-vous pas que vous étiez vraiment James T. Kirk, inexplicablement transféré dans le corps de Lester ? Dans l'épigraphe, la personne qui occupe le corps de Lester admet que les apparences sont contre elle, mais affirme que tout ce qui fait que Kirk est Kirk n'en est pas moins présent. Et il est clair que la personne qui occupe le corps de Lester considère qu'elle est Lester.



Si l'une ou l'autre des perspectives qui mettent la matière au premier plan est juste, tout le monde se trompe dans cet épisode - les sujets de l'expérience et l'équipage de l'Entreprise! Puisque le corps et le cerveau originels de Kirk continuent à fonctionner chez la personne qui occupe son corps, c'est que cette personne est Kirk (et que la personne qui occupe le corps de Lester est Lester). Mais je vous fais confiance pour trouver cette conclusion improbable. En laissant de côté toute question d'«essence», le transfert implique clairement un transfert des psychologies (nous considérerons que «esprit» et «psychologie» sont ici synonymes). Après avoir interrogé les gens au hasard dans la rue pendant des années, j'ai découvert que la réaction de bon sens à ce processus est de le décrire non comme des gens qui échangent leur psychologie, mais comme des gens qui échangent leur corps (et leur cerveau). Le sens commun suggère que, dans ce cas, Kirk va où va sa psychologie - ce que nous appellerons la perspective «selon laquelle c'est l'esprit qui compte».

Mais si Kirk va où va sa psychologie, cela vaut sûrement aussi pour Vedek Bareil et Roger Korby, non ? Le bon sens ne peut pas jouer sur les deux tableaux, et Star Trek non plus. Mais quel sera son choix? La grande différence entre le cas de la personne qui habite le corps de Lester et les cas de Bareil et de Korby, c'est la matière qui constitue leurs corps. Mais nous avons déjà vu au chapitre 3 qu'il n'y pas de raison, en principe, pour nier qu'un corps mécanique - fait de silicone et de métal - puisse avoir une psychologie. Et dans tous les cas, la preuve que le sujet original est bel et bien présent est toujours la même - à savoir, le comportement du sujet après la procédure. Dans le cas de la personne qui habite le corps de Lester, nous pouvons conclure que la meilleure explication de son comportement, c'est qu'elle est bel et bien Kirk. Si l'individu qui émerge de la procédure de substitution de cerveau marche et parle comme Bareil, pourquoi nier qu'il est réellement Bareil? (Cela vaut pour Korby, même si l'infirmière Chapel dit à l'androïde : «Vous ne voyez donc pas, Roger, que tout ce que vous avez fait prouve que ce n'est pas vous!» Une façon d'interpréter cet énoncé étrange est de dire que, pour Chapel, l'androïde n'est pas numériquement identique à Roger Korby, mais si c'est ce qu'elle veut dire, elle n'aurait pu choisir tournure plus confuse. Une meilleure interprétation serait de dire que Chapel affirme que Roger Korby a changé, en pire.)

Une application à l'identité du schéma de l'élément de continuité le plus proche confirme ces jugements, si nous donnons à ce schéma un poids particulier. Le bon sens

applique le schéma de l'élément de continuité le plus proche, et donne de la valeur à la fois à la matière et à l'esprit, mais en faisant primer l'esprit. Prenons l'épisode «Volte-face». Une fois l'échange accompli, Kirk a deux éléments de continuité importants - la personne qui occupe son corps, et celle qui occupe le corps de Lester. Le corps du premier, y compris le cerveau, est identique à celui de Kirk avant l'échange, et la psychologie du second est identique à celle de Kirk avant l'échange : juger que le second est Kirk, c'est accorder le primat à l'identité psychologique. (Notez que nous appliquons le schéma de l'élément de continuité le plus proche à l'identité psychologique tout comme pour l'identité corporelle : l'esprit de la personne qui habite le corps de Lester est l'élément de continuité le plus proche de l'esprit de Kirk avant l'échange, et lui est donc identique.) Dans le cas de Bareil, il y a un seul élément qui peut sérieusement prétendre être l'élément de continuité de sa psychologie: l'esprit de l'individu marchant et parlant qui résulte de la procédure de substitution de cerveau. Puisque l'identité psychologique est préservée par la procédure, c'est aussi le cas de l'identité personnelle. (De fait, puisque les éléments cérébraux de Bareil mis au rebut ne soutiennent plus aucune psychologie, un jugement d'identité personnelle semble encore plus valide dans le cas de Bareil que dans celui de l'échange corporel de Lester.) Quant à Roger Korby, même s'il a deux éléments de continuité importants après la procédure (l'un est son cadavre, l'autre l'androïde), l'esprit de l'androïde est l'élément de continuité le plus proche de son esprit, et lui est donc identique. C'est pourquoi l'androïde est Korby. (Une fois de plus, ce jugement est encore plus valide que dans le cas de la personne qui habite le corps de Lester.) C'est du moins ce qu'exige la cohérence du jugement de bon sens. Et si c'est le cas, alors l'androïde Juliana dans «L'héritage» est bel et bien l'ex-épouse du Dr Soong, après coup. Comme le dit à Data le programme holographique implanté dans son cerveau par le Dr Soong : « La vérité, c'est que, pour tout ce qui importe, elle est bien Juliana Soong. »

Il n'y a, en principe, aucune différence entre la procédure subie par Roger Korby, celle subie par Juliana Soong, et la téléportation par données. Tout comme les schémas physiques et psychologiques de Korby et de Juliana, dans la téléportation par données les schémas physique et psychologique du sujet sont surimposés sur la matière acquise. Et dans les deux cas le corps original est détruit. En appliquant le schéma de l'élément de continuité le plus proche où c'est l'esprit qui prime sur le corps, on obtient le résultat suivant: le sujet matérialisé de la téléportation par données n'est pas un simple clone, il est identique au sujet dématérialisé. Quid de la téléportation par matière? Si nous admettons qu'il y a une identité corporelle entre le sujet dématérialisé et le sujet rematérialisé, alors il n'y a, en principe, aucune différence entre la téléportation par matière et la survie ordinaire, quotidienne.

La vie après la mort

Jusqu'ici, j'ai insisté sur le fait que le sens commun devrait se montrer cohérent et élargir sa conception des possibilités de la survie personnelle. Mon

impression toute particulière est que les gens qui répugnent le plus à cette idée - ainsi qu'à celle d'une téléportation, qu'elle se fasse par la matière ou par les données - sont ceux (notamment de confession chrétienne) qui croient que la mort n'est pas la fin de leur existence. Si on les interroge plus profondément ils ne pensent pas que la mort implique une perte de l'identité corporelle; qu'on survit à la mort en gardant tous ses souvenirs, ses intentions, ses désirs, etc. De fait, puisque la mort n'implique d'ordinaire aucune perte d'identité corporelle, à moins d'une crémation ou d'une évaporation, les chrétiens considèrent que la mort n'implique pas la non-persistance des choses. La meilleure façon de comprendre cette conception de la mort est de supposer une séparation de la personne et de la psychologie du corps. Ce qu'il faut noter, c'est que, une fois de plus, l'identité corporelle (et cérébrale) n'est pas ce qui prime dans l'identité personnelle.

Mais que se passe-t-il exactement lorsque vous êtes séparé(e) de votre corps? Il y a deux possibilités fondamentales. L'une, c'est que votre esprit survit à l'état désincarné (c'est ce qu'on nomme parfois la perspective platonicienne). Cette existence est assez difficile à imaginer, et, pour autant que je puisse le faire, elle ne me semble pas particulièrement désirable. Peut-être peut-on la comparer avec l'idée que votre psychologie est enregistrée sur disque dur et stockée en permanence. Dans ce genre d'existence en continu, il est difficile de voir comment on peut s'impliquer dans une activité alors qu'on est désincarné. (Notez que je ne suggère pas qu'il faut nécessairement avoir un corps terrestre pour être actif - peut-être des corps immatériels ou célestes agissent-ils - et je ne suggère pas non plus que l'activité requiert des mains, ou même des yeux. Mais il faut avoir quelque chose qui permette d'interagir avec le monde où l'on vit, quel qu'il soit.) Mettons de côté ce problème, parce que pour saint Paul - dont l'opinion fut adoptée par les premiers chrétiens, et constitue la vision chrétienne officielle - après la mort, l'esprit d'un chrétien trouve un nouveau corps: « S'il y a un corps psychique, il y a aussi un corps spirituel [...] et, de même que nous avons porté l'image du terrestre, nous porterons aussi l'image du céleste » (1 Corinthiens 15, 44-49).

Si vous partagez les vues de saint Paul sur ce point, alors vous pensez sans doute que, d'une façon ou d'une autre, Dieu fait en sorte que votre psychologie soit extraite de votre corps actuel, et perdure dans un corps entièrement différent. L'intérêt, bien sûr, c'est que la position chrétienne classique de la survie après la mort physique équivaut à la fois à la procédure subie par Korby et à la téléportation par données (encore que la version divine fasse plus de dégâts, puisqu'elle laisse traîner un peu partout des cadavres en putréfaction). Elle équivaut même à la « reconception » par Kevin Uxbridge de son épouse morte. Une fois de plus, un rejet direct de la téléportation par les données ne s'accommode guère de la conviction profonde qu'on survivra physiquement à la mort par un tel processus. Notez aussi que c'est seulement dans le cas de la téléportation par données qu'on peut être sûr d'avoir des témoignages (non controversés) de ceux qui sont passés « de l'autre côté ».

Au passage, certains chrétiens croient apparemment que Dieu les ressuscitera

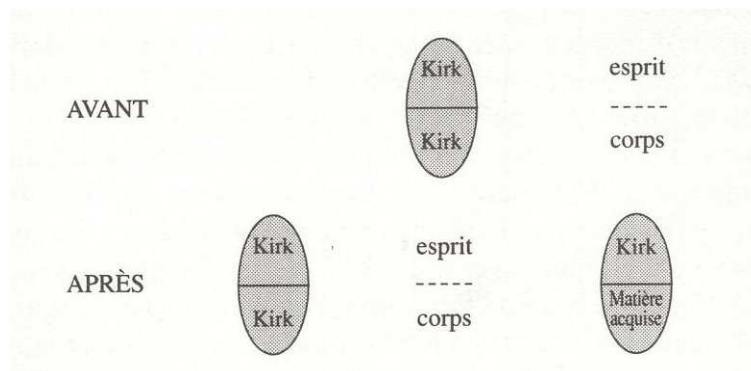
tous au ciel (si tout se passe bien, s'entend) en rematérialisant leur corps terrestre actuel. Le problème, ici, c'est que, puisque votre corps au cours de l'existence est une succession de tranches de corps, dont chacune est composée de matière différente, votre corps ressuscité sera-t-il le vôtre? Tout dépend de la tranche de corps que Dieu choisira de « vous rendre ». Si vous mourez à soixante-dix ans d'une crise cardiaque et que Dieu « vous rende » un corps bien plus sain - le corps de vos vingt ans, par exemple - ce corps ne comptera sans doute pas comme l'élément de continuité le plus proche de votre corps au moment de votre mort. Si la personne ressuscitée doit être « vous », Dieu devrait ressusciter la tranche de corps correspondant à l'heure de votre mort, dans un processus équivalant à la téléportation par matière. Arrangement rien moins qu'optimiste dans certains cas.

La seconde difficulté a souvent été commentée : la matière constituant le corps d'une personne est trop massivement recyclée. Une partie de la matière constituant votre corps sert à faire une partie du corps de quelqu'un d'autre, et peut-être a-t-elle été auparavant la matière d'un autre. Il semble donc que, même si Dieu a fait du mieux qu'il a pu en distribuant cette matière, tout le monde sera incomplet au jour de la Résurrection. C'est pourquoi l'une ou l'autre forme de résurrection implique forcément l'acquisition de matière neuve.

Cela donne un poids supplémentaire à l'affirmation selon laquelle le bon sens veut que l'on choisisse une téléportation par données si on veut y survivre. Mais est-ce rationnel? Devrions-nous accepter que le schéma de l'élément de continuité le plus proche fasse primer l'esprit sur la matière? Cette perspective s'accommode fort bien d'une vision fonctionnaliste de la pensée. Le fonctionnalisme permet l'émergence d'états d'esprit multiples; c'est pourquoi un esprit qualitativement identique peut, en principe, être réalisé au moins sous deux modes - en esprit et en silicone - et donc peut-être dans un corps « céleste ». La perspective qui fait primer l'esprit sur la matière dit que c'est la continuité causale de ces états (en nombre suffisant) - quelle que soit la façon dont ils sont réalisés - qui constitue la survie de la personne. Mais, maintenant, considérons trois objections à cette vision, qui pourraient nous amener à repenser les modes de téléportation.

Le principe d'indépendance

Certains philosophes rejettent le schéma de l'élément de continuité le plus proche faisant primer l'esprit sur la matière pour adopter ce que j'appellerai le principe d'indépendance. Avant d'établir ce principe, essayons de l'approcher intuitivement en revenant sur l'épisode « De quoi sont faites les petites filles ? ». J'ai mentionné que Roger Korby n'était pas la seule personne à être clonée dans un androïde : Kirk a aussi connu ce sort. Toutefois, il y a une différence - dans le cas de Kirk, l'ancien corps et l'ancien cerveau n'ont pas été détruits. Kirk a deux éléments de continuité, tous deux qualitativement identiques au Kirk d'avant l'opération. Le diagramme suivant illustre cela:



La personne qui habite le corps de Kirk (en bas à gauche dans le diagramme) est, nous en avons l'intuition, le véritable Kirk; l'androïde (en bas à droite dans le diagramme) est une simple réplique. Une fois de plus, cela va dans le sens du schéma de l'élément de continuité le plus proche de l'identité personnelle: bien que l'esprit de l'androïde soit un élément de continuité proche de l'esprit du Kirk d'avant l'opération, son corps n'est pas un élément de continuité proche du corps de Kirk avant l'opération. L'esprit de la personne qui habite le corps de Kirk, par ailleurs, est un élément de continuité proche de l'esprit de Kirk avant l'opération, et la personne qui habite le corps de Kirk a aussi le corps de Kirk avant l'opération. C'est lui, en somme, qui remporte la palme en matière d'élément de continuité le plus proche.

Toutefois, l'unique différence entre l'androïde Kirk et l'androïde Korby, c'est que l'original est conservé. Si l'androïde Korby est le vrai Korby, cela signifie que si le corps et le cerveau originaux de Kirk avaient été détruits, l'androïde Kirk aurait été le vrai Kirk ! Cette contingence paraît absurde à certains philosophes qui avancent alors le principe d'indépendance - la question de savoir si l'identité se maintient entre la tranche de personne A et la tranche de personne B dépend d'éléments relevant d'autres individus, possibles ou actuels. Ma réfutation se fera en deux parties. Tout d'abord, il y a des moments en philosophie où des intuitions fondamentales entrent en conflit: en ce qui me concerne, je ne me sens pas attiré intuitivement par le principe d'indépendance, je ne vois aucune raison de l'adopter.

Ma seconde réfutation sera plus substantielle. Si on adopte ce principe, il semble qu'il y ait deux possibilités. L'une est de revenir à la perspective qui fait primer la matière, ce qui signifierait que la téléportation par données ne vaut pas. Toutefois, nous avons déjà vu qu'adopter cette vision nécessite des jugements contre-intuitifs sur le cas Kirk-Lester. De plus, si le principe d'indépendance doit prévaloir, l'on doit adopter un schéma de l'élément de continuité le plus proche qui fait primer le surplus de 'matière. Si le jugement d'identité ne peut faire appel à aucun fait renvoyant à d'autres individus, c'est que l'identité personnelle exige que soit conservée plus de 50 % du corps originel (ou, si l'on fait primer le cerveau, plus de 50 % du cerveau). Toutefois, rappelez-vous la façon dont a commencé l'objection: on tient pour absurde que l'androïde Kirk soit Kirk, si la personne qui habite le corps de Kirk a été détruite. Maintenant, supposons que l'on adopte la perspective qui exige une part majoritaire de cerveau. Dans le cas où 49 % du cerveau de Vedek Bareil aurait été préservé (et en l'absence d'élément de continuité le plus proche), il faut

admettre que le résultat n'est pas Bareil, mais aurait été Bareil s'il avait reçu 2 % de plus de cerveau. Cela ne semble pas moins absurde.

De fait, si l'on veut adopter le principe d'indépendance, il faut rejeter entièrement le schéma de l'élément de continuité le plus proche, nier que l'identité personnelle dépende soit de l'identité physique, soit de l'identité psychologique, et affirmer au lieu de cela qu'elle dépend d'une essence ou d'une âme. Après tout, certains lecteurs peuvent penser que j'ai vraiment mal compris l'histoire chrétienne de la résurrection en me polarisant sur l'identité psychologique comme porteuse de la qualité de personne. On pourrait m'objecter que Dieu ne prend pas juste la psychologie pour l'implanter dans un nouveau corps - Dieu prend l'âme et l'implante, ainsi que la psychologie, dans un nouveau corps. Et cela, pourrait-on affirmer, fait toute la différence entre une résurrection céleste et un simple clonage. Bien que cette idée soit séduisante, elle ne permet pas de rendre compte de l'identité personnelle.

Soyons clair sur ce qu'implique la suggestion présente et retournons à la question de la survie quotidienne, qui, suivant la vision de l'âme, implique non seulement une identité corporelle et psychologique, mais aussi une identité spirituelle. De plus, si l'on suit la suggestion présente, c'est l'identité spirituelle qui importe, puisqu'elle fait l'identité personnelle. Ce qui est intéressant avec cette suggestion, c'est qu'elle considère l'identité personnelle et l'identité psychologique comme séparables en principe: non seulement l'identité personnelle pourrait persister en l'absence de l'identité psychologique, mais l'inverse serait vrai. Nombre de religions non chrétiennes ont adopté la première possibilité : cela semble surpasser la doctrine classique de la réincarnation. Lorsqu'on est réincarné, c'est sans les souvenirs et les expériences de la vie précédente, et il faut tout reprendre depuis le début. Passons sur les aspects bizarres de cette doctrine : si nous examinons la possibilité inverse, l'existence d'une identité psychologique (et physique) en l'absence d'une identité personnelle, un problème épistémologique très difficile apparaît.

Le problème, c'est que si la vision de l'âme est juste, l'identité psychologique et l'identité physique telles qu'elles perdurent dans le temps ne sauraient rien indiquer sur l'identité personnelle telle qu'elle perdure dans le temps. Admettons que vous ayez une âme, en plus d'un corps et d'un esprit, et qu'elle soit indépendante. Admettons par ailleurs que l'identité personnelle revient à l'identité de l'âme. Maintenant, appelons Y1 la tranche de personne engendrée par vos parents, et Y2 la tranche de personne qui lit en ce moment ce livre. Quelle raison peut-on exactement invoquer pour penser que Y2 est identique à Y1 ? Par hypothèse, l'identité de l'âme, si elle résiste au temps, est distincte de l'identité physique et psychologique, à laquelle on ne saurait faire appel pour démontrer qu'il existe une identité spirituelle. Mais quelle autre preuve pouvons-nous en avoir, en l'absence de certaines capacités extrasensorielles ? (La perception extrasensorielle ne nous aiderait pas, car elle ne détecterait que des états psychologiques.) Tout cela s'accorde entièrement avec «votre» expérience suivant laquelle Dieu a recyclé de nombreuses âmes à travers l'esprit et le corps qui sont demeurés « vous » depuis l'arrivée au monde de Y1. En

d'autres termes, ce que nous pensions être « vous » aurait pu être une succession de centaines de milliers de différentes âmes locataires. Comment le savoir? Comment Y2 peut-il le savoir? Nous avons déjà vu qu'on ne saurait explorer avec sûreté sa propre identité (un point qui sera souligné au chapitre 5). Tout bien considéré, puisque ni la perspective qui fait primer la matière ni celle qui fait primer l'âme ne sont très prometteuses, c'est encore la perspective qui fait primer l'esprit sur la matière qui reste la plus probable.

L'objection de Data

Mais ne pourrait-on adopter la perspective qui fait primer l'esprit sur la matière, sans pour autant cesser d'objecter qu'il est impossible de survivre à la téléportation, qu'elle se fasse par matière ou par données? Cette position est suggérée par le dialogue entre Data et le commandant Maddox dans « Ce qui fait l'homme », lorsque Maddox tente de convaincre Data de se prêter à la procédure de démontage qui lui permettra d'étudier la construction de Data (voir chapitres 2 et 3).

Maddox : « J'ai pensé que nous pourrions reparler de cette question, que j'arriverais à vous persuader. Vos souvenirs et votre savoir demeureront intacts. »

Data : « Mais ils se réduiront à une simple information factuelle. La substance, la saveur même des moments vécus risquent de se perdre. Prenez les jeux de hasard ... j'avais lu et absorbé le moindre traité sur la question, et je me suis trouvé bien préparé pour cette expérience. Et pourtant, lorsque j'ai fini par jouer au poker, je me suis rendu compte que la réalité ne ressemble que fort peu à ce qu'en disent les règles. »

Maddox : « Où voulez-vous en venir? »

Data : « Tout en considérant qu'il est possible de charger des informations contenues dans un cerveau positronique, je ne crois pas que vous soyez en mesure de préserver l'essence de ces expériences. La mémoire a un aspect ineffable qui, je le pense, ne saurait survivre à votre procédure. »

L'objection soulevée par Data contre la procédure de Maddox peut être adaptée à la question de la survie à la téléportation. Puisque le téléporteur doit « charger » les informations dans le cerveau du sujet, il y a une chance que soient perdues la « substance », la « saveur », l'« essence », ou quelque autre qualité ineffable des expériences vécues par la personne - et la personne avec.

L'objection de Data est étonnamment similaire à une objection philosophique commune touchant le fonctionnalisme des états mentaux, abordée dans les chapitres 2 et 3. L'intuition fondamentale qui sous-tend cette objection, c'est que le fonctionnalisme (comme toute autre vision matérialiste) est incomplet parce qu'il ne peut rendre compte du point de vue subjectif à la première personne. Selon cette objection, un compte rendu fonctionnaliste de l'esprit sera, comme toute autre théorie scientifique, entièrement constitué de propositions objectives. Mais une

partie de l'expérience qualitative ne peut, semble-t-il, être captée par les descriptions à la troisième personne. Le philosophe Franck Jackson imagine une Mary, experte en neuroscience et connaissant tout sur les opérations du cerveau. Sa théorie scientifique du cerveau est complète, mais elle a passé toute sa vie dans un décor en noir et blanc. Elle sait que la science peut tout lui apprendre sur l'expérience du rouge, mais lorsqu'elle sortira du laboratoire et verra une tomate mûre, elle apprendra quelque chose de nouveau : elle saura à quoi ressemble le rouge. Puisque c'est là une connaissance qu'elle ne possédait pas auparavant, c'est que notre connaissance scientifique du cerveau et de ses opérations a beau être la plus complète possible, elle ne rend pas compte de certains aspects mentaux. L'objection de Data est renforcée par la comparaison avec le poker : selon Data, même si vous étudiez à fond les règles du jeu, vous n'apprendrez jamais tout ce qu'il faut savoir, sinon par l'expérience subjective du jeu.

Pouvons-nous réfuter cette objection? Vous remarquerez d'abord que l'exemple est exagéré. Data fait référence à un incident où il a été complètement bluffé au poker par Riker. Il semble ignorer que le bluff est une tactique précieuse, mais comment a-t-il pu ignorer ce point s'il a tout lu sur ce jeu? Et s'il y a des choses que Data ignore avant coup, son expertise est loin d'être complète. De plus, même si Data sait vraiment tout ce qu'il faut savoir sur le poker, il peut lui manquer une certaine connaissance non propositionnelle, une certaine capacité ou un certain savoir-faire qu'on acquiert seulement par l'expérience. Dans le cas de Mary, il faut nous assurer qu'elle acquiert bel et bien une nouvelle connaissance quand elle observe pour la première fois une tomate : il faut un argument pour nous convaincre qu'elle n'aurait pas eu cette connaissance à travers son savoir exhaustif sur le cerveau. De plus, s'il s'agit vraiment d'une nouvelle connaissance, elle pourrait être non propositionnelle, et s'identifier peut-être à une capacité nouvelle, par exemple la capacité à identifier les objets rouges.

Dans le cas de la téléportation, qu'elle se fasse par matière ou par données, il n'y a encore aucune raison de penser qu'il manquera au sujet rematérialisé (s'il est fonctionnellement identique au sujet dématérialisé) quelque chose que le sujet dématérialisé possédait. Mais une autre objection au fonctionnalisme est aussi valable ici. Au chapitre 3, j'ai conclu qu'une entité artificielle ne peut pas percevoir l'aspect phénoménal de la sensation et de l'émotion. Mais j'ai laissé sans réponse la question de savoir si cet aspect phénoménal pouvait différer d'une personne à l'autre, ou entre des opérations organiques ou artificiellement physiques de l'intelligence. Ce qui laisse également sans réponse la question de savoir si le transfert d'une psychologie personnelle à un autre médium altère l'aspect phénoménal de l'expérience. Étonnamment, cela semble recouper les affirmations de Vedek Bareil dans « Assistance vitale », après qu'on lui a implanté un demi-cerveau.

Kirk: « Comment vous sentez-vous? » Bareil : « Éveillé. Tout est... différent. »
Kirk : « Différent ? Comment cela ? »

Bareil: « C'est difficile à expliquer. Mais lorsque vous me touchez, cela ne

semble pas réel. C'est plutôt comme ... une lointaine réminiscence du toucher. »

Comme je l'ai mentionné au chapitre 3, savoir si ce changement qualitatif dans l'expérience est ou non une vraie possibilité lorsque la nouvelle opération est fonctionnellement identique relève de la controverse philosophique. Mais cela n'importe pas à notre débat actuel, puisque nous pourrions reconnaître cette possibilité sans qu'elle remette en cause la perspective suivant laquelle nous pouvons survivre à la perte de notre corps. Après tout, si l'expérience qualitative peut différer entre les corps organiques, votre vision actuelle du rouge pourrait un peu s'éloigner de celle que vous aviez à l'âge de six ans. Ainsi, cette potentialité interdit toute objection rationnelle à la survie à la téléportation par matière ou par données - à cause des scrupules tels que ceux exprimés par Data. Mais nous pouvons peut-être relier l'émoi de Data à quelque chose d'entièrement différent, et voir une objection supplémentaire à l'un ou l'autre mode de téléportation.

Le principe d'exclusion

Je soupçonne bien des gens, y compris les auteurs de Star Trek, d'adopter ce que j'appelle le principe d'exclusion - selon lequel, si nous autres humains inventons un processus et le manipulons de façon fiable, alors ce processus s'accomplit en toute connaissance de cause - au contraire des processus dont nous ne sommes pas responsables. Examinons les preuves. Nous acceptons en général la possibilité d'un échange corporel entre Kirk et Lester dans « Volte-face », mais nions la possibilité d'une greffe progressive du corps et du cerveau sur Vedek Bareil dans « Assistance vitale ». La différence ? La première se passe d'une façon mystérieuse, grâce à une technique inventée par une race supérieure, alors que la seconde est le produit de la technologie humaine. Examinons par ailleurs les deux réactions fondamentales face à Data. D'un côté, une bonne partie de la répugnance à accepter Data comme personne peut être liée au fait qu'il a été conçu et construit par un humain. D'un autre côté, le manque général de compréhension quant à la façon dont Data fonctionne crée une aura de mystère qui confine au mystique. Si le principe d'exclusion est largement accepté, il pourrait expliquer pourquoi tant de gens sont enclins à croire que la téléportation par données marche lorsque c'est Dieu qui la gère, et non les humains.

Ce ne sont pas seulement les « gens dans la rue », qui sont tentés par le principe d'exclusion. Dans son ouvrage par ailleurs excellent, *La physique de Star Trek*, Lawrence Krauss écrit que l'idée de la téléportation vise à recréer une « personne fonctionnellement identique », elle « contredit nombre de croyances spirituelles quant à l'existence d'une "âme" qui se distinguerait d'une façon ou d'une autre du corps. Qu'est-ce qui se produit quand vous mourez, après tout ? De nombreuses religions ne soutiennent-elles pas que l'âme perdure après la mort ? Qu'arrive-t-il donc à l'âme tandis que le corps est téléporté ? En ce sens, la téléportation serait une merveilleuse expérience en matière de spiritualité. Si une personne remontée à bord de l'Enterprise demeurait intacte et manifestement identique, cela serait une preuve

frappante qu'une personne humaine n'est rien de plus que la somme de ses parties corporelles, et cette démonstration contredirait immédiatement nombre de croyances spirituelles. »

Krauss commet ici une erreur bien compréhensible. Nous avons déjà observé un parallèle frappant d'une part entre le processus hypothétique de la téléportation par données et la survie de l'esprit à la mort physique telle qu'on se la figure, et d'autre part entre la téléportation par données et une vision eschatologique un peu moins populaire. Il semble en toute cohérence que ceux qui s'attendent à survivre à la mort devraient accueillir favorablement l'idée de la téléportation, qui ne « contredit » pas du tout leurs croyances. La tendance de Krauss à penser le contraire suggère qu'il ratifie le principe d'exclusion. En proposant d'utiliser le téléporteur pour vérifier les croyances spirituelles, il insinue fortement qu'il est sain d'appliquer la méthode scientifique au problème de la survie personnelle, surtout si on la compare à la méthode scientifique que nous avons adoptée - méthode par laquelle nous ne faisons ces expériences qu'en imagination, en sollicitant nos intuitions et en recherchant la cohérence. Mais bien qu'on puisse désirer une méthode plus fiable, et qu'on soit tenté de chercher une « expérience cruciale » en matière de spiritualité, le fait est que le succès de la téléportation n'ajouterait strictement aucune nouvelle preuve contre les croyances spirituelles de toute sortes !

Nous diviserons les perspectives spirituelles - « la croyance quant à l'existence d'une âme qui se distinguerait du corps » - en deux types, tout en rappelant qu'il est ambigu de parler d'une « âme ». Dans la première perspective, l'âme ne fait qu'une avec l'esprit, et l'identité de l'âme résistant au temps n'est autre que l'identité psychologique résistant au temps. Confronté à un téléporteur en état de marche, que dira un spiritualiste ? De deux choses l'une. Ou il affirmera qu'il n'y a pas d'identité psychologique entre le nouveau sujet et le sujet d'avant la téléportation (peut-être parce que celui-ci est mort, que sa véritable psychologie est ailleurs, et que le sujet rematérialisé n'est qu'un clone; ou peut-être parce que le sujet matérialisé est un zombie sans un gramme d'esprit). Ou bien il reconnaîtra que tout démontre que la personne originale a survécu, parce que le sujet rematérialisé montre la même psychologie - en d'autres termes, la même âme.

D'un autre côté, supposons que par « âme » on entende quelque chose de séparable à la fois du corps et de l'esprit (c'est la perspective que Krauss a en tête, je parie). Confronté à un téléporteur en état de marche, que dira le tenant de cette perspective-là ? De deux choses l'une. Ou il affirmera qu'il n'y a pas d'identité psychologique entre le sujet rematérialisé et le sujet avant la téléportation (peut-être parce que le sujet est mort, que son âme est ailleurs, et que le sujet rematérialisé n'est qu'un clone); ou bien que le processus a de fait préservé l'identité psychologique mais a égaré l'identité de l'âme (ou peut-être que le sujet matérialisé n'a aucune âme), dans ce cas, il ne fera que nier que l'âme originale est présente. Ou bien il reconnaîtra que la personne originale a, de fait, survécu (par un processus surnaturel inconnu) parce que le sujet matérialisé manifeste la même psychologie, et dans ce cas il n'y a pas plus de raison manifeste de nier l'identité de l'âme que dans le

cas d'une survie ordinaire et quotidienne.

L'attraction de Krauss pour le principe d'exclusion est compréhensible parce qu'il ne diffère que très peu d'un principe méthodologique (solide) appelé le rasoir d'Occam: ne multipliez pas les entités au-delà du nécessaire. De manière très défendable, cet argument nous dit que, lorsque nous sommes confrontés à un phénomène, toutes choses égales par ailleurs, nous devrions choisir l'explication qui requiert le plus petit nombre d'entités diverses. Si nous acceptons le rasoir, comme c'est sûrement le cas de Krauss, il s'ensuit que, si nous inventons un processus et pouvons le manipuler de façon fiable, alors probablement rien d'inconnu ne se passe lorsque le processus s'accomplit. Qui plus est, il en découle que Krauss a raison en disant que la téléportation réussie dénie l'existence d'une âme non matérielle qui serait essentielle aux personnes. Mais aucun penseur cohérent ne devrait se laisser arracher au spiritualisme et convertir au matérialisme par de telles preuves, car le rasoir est parfaitement général, et ne s'applique qu'aux processus dont nous ne sommes pas responsables. Si un processus se produit, et que nous pouvons l'expliquer et le prédire en ne faisant appel qu'à des postulats naturels (ou surnaturels, si nous pouvons le faire en toute certitude), alors le rasoir postule qu'il n'y a sans doute rien de surnaturel qui se produise lorsque ce processus s'accomplit. L'intérêt de cet argument, c'est que, s'il marche dans le cas de la survie à la téléportation, il marche tout autant dans le cas de la survie ordinaire, étant donné notre connaissance actuelle de ce processus. S'il faut renoncer aux croyances spirituelles face à un téléporteur en état de marche, il faut tout autant le faire face à l'existence humaine ordinaire. (Inversement, s'il n'y a pas de raisons irrésistibles de renoncer aux croyances spiritualistes face à l'existence humaine ordinaire, alors il n'y a pas de raisons irrésistibles d'y renoncer face à un téléporteur en état de marche.) Nous n'avons donc aucune raison de nous attendre à ce qu'un spiritualiste récalcitrant se laisse convertir par le coup du téléporteur: il recourra exactement aux mêmes mystères pour invoquer ce qui se passe et pour expliquer les événements les plus ordinaires. Comme il devrait être clair maintenant, je pense que le principe d'exclusion est absolument erroné, et ne peut offrir aucun argument contre la possibilité de survivre à l'un ou l'autre mode de téléportation.

Alors ... être ou ne pas être téléporté ? Voilà la question, mais elle est la réponse? Dans la perspective de l'esprit considéré comme supérieur à la matière, la question de savoir si la téléportation est rationnelle dépend de savoir si, à travers des témoignages valables, on peut raisonnablement s'attendre à ce que cité psychologique soit préservée dans le processus. Pour en juger, il faut examiner le comportement des sujets rematérialisés et des témoignages de leurs amis, collègues et proches. Que la téléportation se fasse par matière ou par données, ces témoignages soutiennent fortement la conclusion selon laquelle la continuité psychologique est préservée. La seconde condition, c'est qu'il n'y ait pas d'autre élément de continuité psychologique du sujet scanné. Puisque j'ai défini la téléportation par matière et la téléportation par données de façon à exclure cette possibilité - au moins dans le cas des rematérialisations uniques - nous devrions conclure qu'il est, après tout, rationnel

d'être téléporté, et irrationnel de s'y refuser. Toutefois, que penser des cas de téléportation qui provoquent un changement numérique ou qualitatif? Ce sera le sujet du chapitre suivant.

Chapitre 5

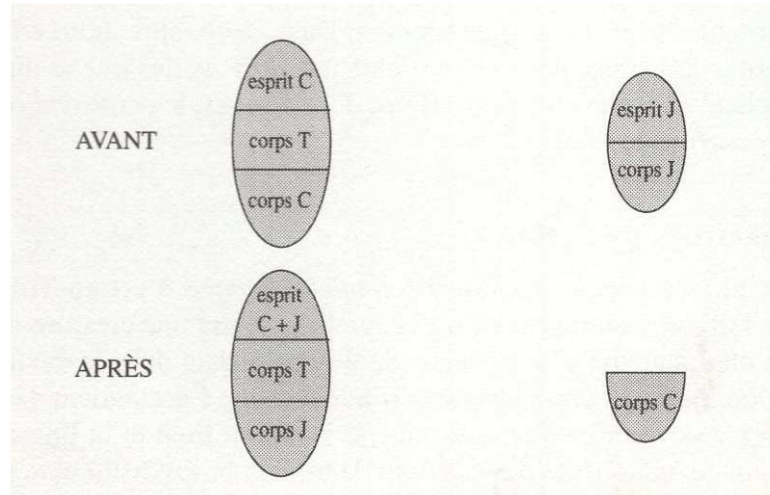
L'épanouissement personnel

L'intérêt de la téléportation, tout insolite qu'elle soit, tient à sa fonction de conservation - dans la plupart des cas - et ce, à deux titres. Tout d'abord, la procédure conserve le nombre de sujets: si un sujet se soumet à la procédure, il en émerge un et un seul à la sortie. Et puis, elle conserve l'identité qualitative: le corps et la psychologie du sujet qui réapparaît ne peuvent être distingués de ceux du sujet premier. Dans ce chapitre, nous étudierons des procédures encore plus insolites : celles qui se distinguent de l'une ou l'autre de ces deux formes de conservation - et parfois des deux !

Réactions de fusion

L'un des personnages de Profondeur Espace 9 est un Trill. Un Trill ou « symbion », c'est la symbiose entre une créature de grande longévité - une espèce de limace nichée dans la cavité abdominale - et une succession d'humains qui l'accueillent. (Là où cela devient confus, c'est que le symbion final et la limace originale ont tous les deux le nom de trill ; pour les distinguer, je mettrai une majuscule lorsque Trill désignera le symbion.) Lorsqu'un hôte humain meurt, le trill est transféré (avec sa psychologie intacte) chez un nouvel hôte. C'est un grand honneur que de l'accueillir, et le fruit d'une longue préparation, requérant de nombreuses années d'entraînement. Chaque Trill adopte le prénom de l'hôte humain du moment et garde le nom de famille du trill. Le lieutenant Jadzia Dax, huitième incarnation du trill Dax, est une jeune femme ; dans «Dax» (DS9), Jadzia Dax est accusée de trahison et de meurtre, crimes qu'on suppose avoir été commis par Curzon Dax, et la cour essaie de décider si l'identité personnelle de Curzon a été préservée en elle. Cette procédure cesse d'être nécessaire lorsqu'il apparaît que Curzon était innocent... mais supposez qu'il ait été coupable. Cela aurait-il valu aussi pour Jadzia Dax? La thèse qui fait primer l'esprit sur le corps, défendue au chapitre 4, nous suggère une façon de régler la question. L'identité psychologique a-t-elle été préservée au cours du changement d'hôte ? La pensée de chaque Trill est un mélange entre la psychologie du trill et celle de l'hôte, et la psychologie du trill est elle-même un composé de toutes les symbioses précédentes. Essayons de traduire par un diagramme ce qui s'est passé lorsqu'on a

conjoint l'humaine Jadzia et le trill Dax. C renvoie à Curzon, J à Jadzia, et T au trill.



Il y a deux personnes distinctes (ou des tranches de deux personnes distinctes) avant le processus de jonction - le Trill Curzon Dax et l'humaine Jadzia. Bien qu'il reste deux individus à l'issue du processus, un seul est une personne: Jadzia Dax. L'autre individu est le corps humain inanimé, associé jadis à l'humain Curzon, puis au Trill Curzon Dax. Nous avons ici une procédure insolite qui ne maintient ni l'une ni l'autre forme de conservation. Le nombre de personnes change, puisqu'il y avait deux personnes avant la soudure, et une seule ensuite ; avant la jonction, il y avait deux psychologies individuelles, et après, une seule. De plus Jadzia Dax n'est pas qualitativement égale en psychologie à Curzon Dax, ni à Jadzia. Appliquons l'élément de continuité le plus proche. Curzon Dax a deux éléments de continuité - Jadzia Dax et le corps inanimé de Curzon Dax. Il est clair que Jadzia Dax est l'élément de continuité de Curzon Dax puisqu'elle contient le corps du trill, qui n'a pas changé, et la plus grande partie de sa psychologie lui vient de Curzon Dax. Il semble donc que Jadzia Dax soit identique à Curzon Dax. Mais à présent nous devons considérer les éléments de continuité de Jadzia. Il n'y en a qu'un: Jadzia Dax, qui a assimilé sa psychologie tout entière et possède par ailleurs son corps. Il semble que Jadzia Dax soit identique à Jadzia. Mais si c'est vrai, alors par transitivité d'identité (si $a = b$ et $a = c$, alors $b = c$), Curzon Dax est identique à Jadzia, et ce n'est pas exact ! (S'ils étaient identiques, il suffirait de mettre Curzon Dax en prison pour y mettre J Jadzia.)

Avant d'essayer d'élucider ce cas, demandons-nous pourquoi l'identité personnelle semble si importante. L'une des raisons à cela, c'est que la justice semble en dépendre fondamentalement: il est sûr qu'une personne ne devrait pas être punie pour les actes d'une autre (à l'exception de la loi Klingon, où un fils doit assumer la culpabilité du père). Cela vaut pour les récompenses : je donne 18 à une étudiante bien précise en supposant que c'est elle qui a passé l'examen. Dès lors, si on peut montrer que Jadzia Dax n'est pas identique à Curzon Dax, il serait inique de la punir pour une chose qu'il a faite, lui. Nous avons déjà éliminé une possibilité : nous savons que Jadzia Dax ne peut être identique à Curzon Dax tout en étant identique à Jadzia. Mais laquelle des possibilités restantes est la plus plausible? Jadzia Dax est-elle identique à Curzon Dax et non à Jadzia, et vice versa, ou n'est-elle identique ni à l'un ni à

l'autre? L'examen d'autres cas nous aidera à en décider.

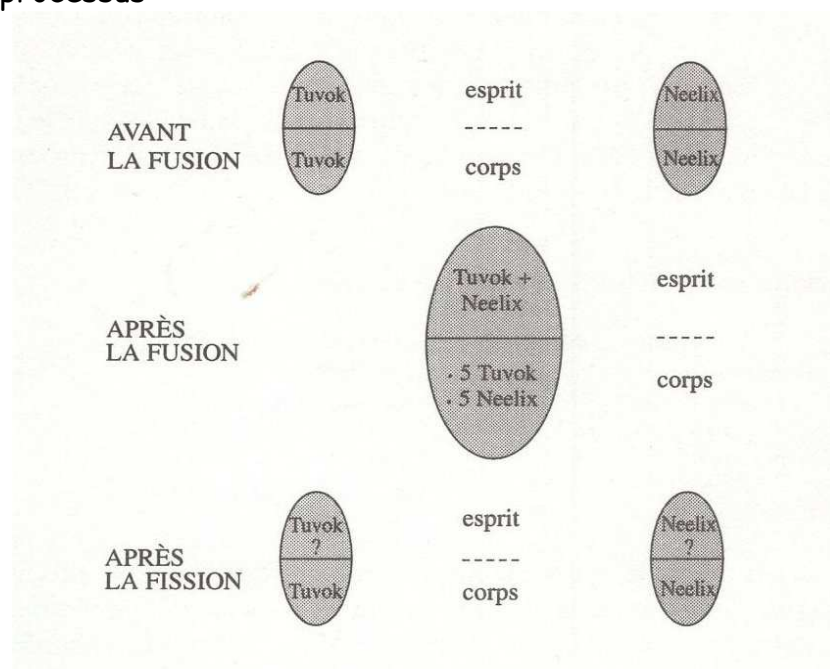
Fission, fusion et confusion

*L'homme que vous voyez devant vous est
littéralement une fusion de deux hommes.*

**Le médecin de bord au capitaine
Kathryn Janeway, « Tuvix » (V)**

Dans « Tuvix », il se produit un accident de téléporteur surprenant. Après avoir collecté des échantillons végétaux à la surface d'une planète, le chef de la sécurité Tuvok et le cuisinier du bord Neelix sont ramenés au vaisseau, mais ce qui se matérialise sur l'aire de rematérialisation du téléporteur, c'est, pour citer le médecin de bord holographique, une «fusion» - un individu unique, possédant les traits physiques et psychologiques des deux hommes. Le nouveau venu s'appelle « Tuvix ». Lorsque l'équipage du Voyageur travaille à mettre fin à la fusion, Tuvix accueille cette idée avec un manque total d'enthousiasme. Il ne s'identifie ni à Tuvok ni à Neelix. « Dans un sens, je les considère comme mes parents », explique-t-il au partenaire de Neelix, Kes. Lorsqu'on découvre un procédé de restauration, Tuvix commence par refuser de s'y soumettre. « Je ne veux pas mourir, dit-il. J'ai le droit de vivre. » Nombre des membres d'équipage expriment leur sympathie, et le docteur refuse de lancer la procédure de téléportation qui doit restaurer l'ordre initial. Mais J Janeway prend le contrôle des opérations, enclenche le téléporteur, et deux individus distincts apparaissent sur l'aire d'atterrissage.

Mettons de côté pour l'instant les questions éthiques. Tuvix a-t-il raison de redouter la procédure de restauration, dont il considère qu'elle mettrait fin à son existence? Qu'est-il arrivé à Tuvok et à Neelix? Une fois de plus, traçons un diagramme du processus:



Si l'on raisonne comme dans le cas de Jadzia Dax, nous devons conclure que Tuvix ne saurait être identique à Tuvok et à Neelix, les personnes telles qu'elles étaient avant la fusion, au sommet du diagramme. (Pour éviter toute question, nous appellerons T et N les personnes d'après la fission.) Tuvix est clairement l'élément de continuité le plus proche de Tuvok, mais aussi celui de Neelix. Mais si Tuvix est identique à la fois à Tuvok et à Neelix, alors, par transitivité d'identité, Tuvok est identique à Neelix, ce qui ne saurait être juste. Mais puisque Tuvok et Neelix contribuent à part égale au corps et à la psychologie de Tuvix, ils ont également le droit de revendiquer leur identité avec lui. On ne saurait donc dire que l'un est identique si l'autre ne l'est pas. Tuvix n'est donc identique ni à Tuvok ni à Neelix. Mais si c'est le cas, qui sont T et N? Sont-ils de nouvelles personnes, de proches répliques de Tuvok et Neelix? Ou sont-ils réellement Tuvok et Neelix?

Des processus comme ceux qui ont produit Jadzia Dax et Tuvix sont des fusions personnelles qui posent à coup sûr d'étonnants problèmes d'identité. Pour essayer de démêler l'énigme de l'identité et de la culpabilité de Jadzia Dax, et celle de l'identité et des droits de Tuvix, nous examinerons le processus opposé : la fission de la personne.

Une migraine à fendre le crâne

Picard: « Comment deux hommes adultes pourraient-ils partager les mêmes souvenirs d'enfance? Cela n'a strictement aucun sens ! »

Le sosie de Riker: « Je suis Will Riker. Je ne sais pas quel être, quelle créature est revenue à bord du Potemkin ce jour-là, mais ce n'était pas moi ! »

« Une seconde chance » (NG)

Un exemple spectaculaire de fission personnelle se produit dans l'épisode de la sixième saison de La nouvelle génération, intitulé « Une seconde chance ». Mais avant de l'aborder, voyons un exemple glané dans un épisode de la seconde saison, « En montant la longue échelle ». L'Entreprise-D rencontre les Mariposiens, descendants d'un groupe de colons venus de la Terre, qui ont atterri brutalement sur la planète qu'ils entendaient coloniser. Cinq des colons seulement ayant survécu à l'atterrissage, ils ont recouru au clonage, et la société mariposienne va au-devant de gros problèmes si elle ne trouve pas du matériel génétique frais pour de nouveaux clones. Ils sollicitent L'Entreprise-D pour un don indolore de cellules, provoquant un Beuuuurk ! général. Voici notamment la réaction de Riker :

Riker : « Vous voulez nous cloner ? »

Premier ministre mariposien : « Oui. »

Riker : « Jamais ! Pas moi ! »

Premier ministre: « Mais quel mal cela pourrait-il vous faire? »

Riker: « Physiquement, aucun. William Riker est unique, pour ne pas dire spécial. Mais des centaines, des milliers de Riker ... je me sentirais diminué au-delà de

l'imaginable.»

Prenons maintenant « Une seconde chance ». L'Entreprise-D visite la base de recherche de Nerval IV, que le commandant Riker a aidé à évacuer huit ans plus tôt, lorsqu'il était lieutenant sur le Potemkin. Cette planète a une atmosphère orageuse, qui rend la téléportation difficile, et lorsque l'équipe de Riker parvient à se faire téléporter en profitant d'une «fenêtre» viable, elle fait une découverte surprenante : Riker se trouve face à face avec son double! Ce sosie dit être le véritable William T. Riker, qui n'aurait jamais quitté la planète, affirmation contestée, bien sûr, par le commandant Riker. On établit rapidement que le sosie n'est pas un clone (avec, au passage, une belle conférence du Dr Crasher sur les similarités non clonables du schéma cervical). Apparemment, lorsque la base de recherche a été évacuée, Riker a été le dernier à être téléporté. Deux signaux viables ont été envoyés depuis la surface de la planète: l'un a produit une rematérialisation à bord du Potemkin; l'autre a été réfléchi sur la surface de la planète et a produit une rematérialisation à cet endroit. Lorsque Picard demande à Geordi LaForge lequel des deux sujets rematérialisés est le vrai, LaForge réplique qu'ils sont vrais tous les deux, puisqu'ils ont été tous deux «formés à partir d'un schéma complet».

LaForge a raison dans un sens et tort dans l'autre. Il a raison de dire que chacun des deux sujets matérialisés compte comme personne réelle : chacun possède les attributs nécessaires pour être une personne, au vu des critères ordinaires. Mais si LaForge affirme que chacun d'eux est le vrai Riker - celui qui a existé juste avant la téléportation, ce jour fatal - il se trompe assurément. Appelons Riker 1 le sujet matérialisé à bord du Potemkin et Riker 2 celui matérialisé à la surface de la planète. Il y a quatre possibilités :

1. Riker 1 = Riker, et Riker 2 = Riker
2. Riker 1 = Riker 1, et Riker 2 := Riker
3. Riker 1 := Riker, et Riker 2 = Riker
4. Riker 1 := Riker, et Riker 2 := Riker

Si l'on argumente comme pour la fusion, nous pouvons montrer que, suivant le principe de transitivité d'identité, puisque Riker 1 n'égale pas Riker 2 (en incarcérant l'un, on n'incarcère pas l'autre), ni Riker 1 ni Riker 2 ne sont identiques au Riker original. (Riker 2 a raison d'affirmer dans l'exergue de cette section que Riker 1 n'est pas lui, mais je ne pense pas que ce soit exactement ce qu'il veuille dire.) C'est pourquoi William T. Riker a cessé d'exister le jour où il s'est scindé. En supposant qu'il ne se produise pas de fissions ultérieures, le commandant Riker est identique à Riker 1 et à Riker 2. L'ironie, c'est que l'individu qui a prôché l'importance de la singularité aux Mariposiens (qui s'en passaient fort bien) était Riker 1, qui, sans le savoir, avait à ce moment même un jumeau bien plus semblable à lui que ne saurait l'être un clone !

Si l'on considère qu'il s'agit d'une téléportation par données avec «dédoublage», une telle fission implique la perte de l'identité corporelle et personnelle (même si la matière est préservée, personne ne sait qui a reçu quoi). De

plus, cette fission implique également la perte de l'identité psychologique : puisqu'il y a maintenant deux esprits numériquement distincts, dont chacun est un élément de continuité également proche, ils ne sont ni l'un ni l'autre identiques à l'esprit original. En réfléchissant sur ce cas, rappelez-vous que la seule différence entre un cas de fission et un cas de téléportation ordinaire par données tient à la matérialisation de deux sujets au lieu d'un; les sujets matérialisés ne peuvent se distinguer au seul vu du comportement interne et externe du sujet scanné; chacun d'eux se rappellera avoir pris son petit déjeuner, être entré dans le scanneur, etc. Et puisque le sujet scanné était sans doute conscient de la possibilité d'une fission, chaque sujet matérialisé s'identifie naturellement avec le sujet scanné, comme le démontrent les comportements de Riker 1 et de Riker 2.

Ce qui importe vraiment dans la survie personnelle

La question que soulève la fission personnelle est la suivante : comment la considérer, étant donné qu'elle implique la perte de l'identité personnelle, corporelle et psychologique? Concentrons nous brièvement sur les deux options habituelles en matière d'existence. Première option: le quotidien au continu, la survie au jour le jour. Seconde option: la mort ordinaire. L'homme de la rue préfère de loin, et dans une très large majorité, la survie ordinaire et continue à la mort. (Imaginez une enquête qui demanderait aux gens ce qu'ils pensent du jugement: « Je préfère le droit de continuer à survivre normalement aujourd'hui au droit de mourir. » Je crois que nous nous attendrions à une quasi unanimité pour la réponse: « Tout à fait d'accord! »)

Ajoutons maintenant une troisième option - celle de la fission personnelle. Sans doute la plupart des gens répondraient-ils qu'ils préfèrent de loin la survie continue ordinaire à la fission. Mais supposez que nous essayions de considérer la fission à la fois en termes de survie ordinaire continue et de mort ordinaire ? La fission est-elle aussi horrible que la mort? Est-elle plus proche de la survie personnelle? Une façon de tester nos intuitions, ici, est d'imaginer un choix forcé où la survie ordinaire continue n'est pas une option. Préparez-vous à laisser s'envoler votre imagination. Supposons que vous ayez une maladie mortelle qui vous tuera d'ici une semaine. Supposons encore que la seule voie de guérison de cette maladie ait été trouvée au cours de recherches sur la téléportation; en modifiant le téléporteur d'une certaine façon (comme dans « Sélection anti-naturelle » [NG]), les médecins ont découvert que la maladie peut être éliminée chez le sujet scanné. Mais, pour une raison inexplicable, cette procédure ne fonctionne que si deux matérialisations ont lieu : une téléportation ordinaire n'éliminera pas la maladie et la mort suivra sous peu. Vous êtes confronté(e) à l'alternative suivante: une mort imminente et certaine, ou la fission et, apparemment, la perte de votre identité personnelle. Que choisiriez-vous ?

Comme élément de réponse, examinons l'existence d'un jumeau fissionne}, qui connaîtrait l'autre jumeau. Nous en avons une idée en observant les deux Riker : c'est en gros comme si on était un jumeau identique, et où est le mal ? Dans « Une seconde chance», il y a un peu de rivalité fraternelle (Riker 2 se sent mis en retrait- à juste

titre, puisque c'est Riker 1 qui a été récompensé pour les actes de Riker), mais aussi une affection qu'il reconnaît à contrecœur (lorsque Riker 2 fait la cour avec succès à Deanna Troi, dont les relations avec Riker 1 avaient connu un certain refroidissement, Riker 1 se sent « flatté, pour ainsi dire ... »). Puis, à la fin de l'épisode, Riker 1 admet que Riker 2 a un droit égal sur certaines de ses possessions, et lui remet le trombone de Riker.

On voit difficilement quelles différences de principe il y aurait entre des jumeaux fissionnels et des jumeaux identiques. Cela n'a pas l'air de trop mal se passer - et de fait, les jumeaux identiques affirment souvent qu'ils préfèrent être jumeaux plutôt qu'être enfant « unique », du moins tel qu'ils se le figurent. Imaginez un moment (si vous n'êtes pas jumeau) que vous vous découvrez un jumeau parfait que vous n'avez jamais vu. Est-ce un problème d'avoir des doubles identiques ? Je doute que les triplés, les quadruplés ou les quintuplés soient moins heureux que les jumeaux. Bien sûr, si vous connaissez tous vos jumeaux et qu'ils sont des centaines, cela posera des problèmes délicats. Mais ces problèmes n'ont rien à voir avec le fait d'être unique, et c'est ce que je vais démontrer.

Dans « Une seconde chance », Data dit à Worf: « J'ai découvert que les humains attachent un prix à l'idée d'être unique - au sens où ils sont différents de tous les autres. L'existence d'un double empêcherait ce sentiment. » La plupart d'entre nous approuveraient totalement l'idée que nous sommes attachés à la particularité de notre caractère. Commençons par donner un sens clair à cette idée: s'agit-il d'une unicité numérique ou qualitative? Si elle est numérique, nul besoin de s'alarmer, puisque la logique veut que chacun de nous soit numériquement distinct. Même les plus identiques des jumeaux identiques ne sont pas numériquement identiques. Et, même si nous vous clonions un trillion de fois, ou vous fissurions en un trillion de « descendants », nous ne vous ôterions pas votre unicité numérique. C'est donc l'unicité qualitative qui compte. Chacun de nous chérit l'idée que personne d'autre ne lui soit qualitativement semblable. Un débat philosophique s'est élevé autour de la question de savoir si l'unicité qualitative est garantie par la logique ; certains philosophes affirment que si a et b sont numériquement distincts, alors il est au moins une qualité que a possède et que b ne possède pas, ou vice versa. Si c'est vrai, alors personne d'autre n'est exactement comme vous. Mais mettons cela de côté. Ce que semble avancer Data, c'est qu'il n'est pas bon qu'il y ait d'autres personnes extrêmement semblables à vous. Pourtant, en quoi est-ce une mauvaise chose? Rien ne le prouve. Les humains semblent chercher avant tout ceux qui leur ressemblent assez, en apparence, en tempérament ou en conviction. Nous sommes peu à nous complaire dans une situation où nous ne nous « intégrons » pas, ne nous « mélangeons » pas, ou n'« appartenons » pas. En considérant les choses d'un autre point de vue, posez-vous honnêtement la question: le monde serait-il meilleur si plus de gens partageaient (en gros) vos opinions? Je parie que vous êtes tenté(e). La reproduction de votre psychologie aiderait à atteindre cet objectif plus qu'elle ne lui nuirait, non ?

Mon diagnostic spéculatif sur la crainte commune de la duplication, c'est qu'elle n'a en fait rien à voir avec l'importance d'être unique. Lorsque nous nous figurons des

clones et des doubles, nous imaginons une vaste armée de créatures qui pensent et agissent toutes pareillement. Ces êtres fantasmatiques ne sont pas effrayants parce qu'ils se ressemblent, mais plutôt parce qu'ils manquent apparemment de la liberté de déterminer le chemin que prendra leur vie. Nous voyons en eux des esclaves dévoués corps et âme: séduire le cœur et l'esprit d'un seul d'entre eux, c'est les séduire tous. Nous les imaginons comme les Borgs de La nouvelle génération, qui possèdent effectivement une conscience collective ne laissant aucune place à l'individualisme. Mais, en réalité, les descendants d'une fission ne manqueront d'autonomie que si le sujet primitif en manquait lui aussi. Une fois qu'ils existent, ils ne ressemblent pas aux Borgs - en séduire un, ce n'est pas séduire le lot - puisqu'ils ne sont pas soudés psychologiquement. Bien sûr, la fission (et le clonage) offrent la possibilité de créer des armées au cerveau lavé, mais pas plus que la vie telle que nous la connaissons, comme en témoigne l'histoire humaine.

Il vaut la peine de répéter ce point, notamment en ce qui concerne le clonage. Les scientifiques écossais ont récemment cloné un agneau à partir d'une cellule non reproductrice de brebis. Le clone est, bien sûr, génétiquement identique à la donatrice de cellule, et le succès de la procédure a effrayé ceux qui envisagent une extension, qui n'est que trop possible, au genre humain. Mais nombre des objections au clonage humain propagées dans le monde reposent sur une erreur manifeste. Si nous vous clonions séance tenante et faisons croître un nouvel être humain génétiquement identique, il n'y a aucune raison de supposer que cette personne serait votre double psychologique. La personne clonée grandirait à une époque différente, elle vous serait donc sans doute bien moins ressemblante qu'un jumeau identique.

Et si vous étiez en fait un jumeau fissionnel - sans le savoir? Une fois encore, pour paraphraser le premier ministre mariposien, on ne voit pas quel mal cela pourrait vous faire - à moins que vous n'ayez l'impression de n'avoir pas connu votre jumeau. Le fait d'être jumeau ne contredit pas votre expérience présente? Il en va de même pour le jumeau issu de fission. (La personne engendrée par vos parents a cessé d'exister à l'âge de quinze ans, ayant été kidnappée et fissionnée au cours d'une téléportation par données; votre jumeau est maintenant dans un laboratoire souterrain de Virginie.) Cela vous poserait-il d'autre problème que du souci pour votre jumeau? Cela ferait-il moins de vous une personne? Pour autant qu'on le sache, Dieu fissionne continuellement des humains; peut-être que toutes les cinq minutes chaque personne cesse d'exister et est fissionnée: un descendant remplace instantanément l'original, et l'autre se matérialise ailleurs, sur une terre jumelle (ou est offert en joujou à Satan). Les personnes ne subsistent jamais plus de cinq minutes, et tout le monde a non seulement un jumeau fissionnel mais aussi des millions de cousins fissionnels. Cela change-t-il quoi que ce soit pour vous, du fait que cela n'appartient pas à votre expérience ordinaire?

Mais peut-être n'est-ce pas le point de vue de vos produits fissionnels qui vous préoccupe. Peut-être est-ce surtout votre point de vue à vous. Après tout, pourriez-vous dire, quel est l'intérêt d'avoir des rêves et des ambitions si ce n'est pas vous qui les réalisez? Pour répondre à cette objection, pensez aux multiples possibilités que

quelqu'un d'autre atteigne le but particulier que vous vous étiez fixé. (Prenez le temps de songer à un but particulier.) À l'une des extrémités du spectre, il y a la survie ordinaire où (nous le supposons) c'est vous qui atteignez le but fixé. Dans tous les autres cas, imaginez que c'est quelqu'un d'autre qui y parvient, sans que vous puissiez intervenir dans son existence. Vous identifiez-vous à l'accomplissement de cette personne, y trouvez-vous une satisfaction personnelle? Certes non, je présume. Encore un échelon: nous arrivons au cas où c'est l'un de vos disciples, ou un ami proche qui connaît cette réussite. Savoir qu'ils réaliseront probablement l'ambition qui est la vôtre ne vous inspire-t-il pas une satisfaction personnelle? À présent, imaginez qu'il s'agit de l'un de vos enfants, qui a la moitié de votre carte génétique. Cela ne vous donne-t-il pas une satisfaction personnelle plus grande encore? (Même les parents qui ne participent pas directement à l'éducation de leurs enfants semblent se réjouir de leurs succès.) Puis supposez que c'est votre clone, quelqu'un de génétiquement identique à vous. Ne serez-vous pas encore plus ravi(e) - ou pas moins ravi(e) - que s'il s'agissait de votre enfant? (Rappelez-vous que vous ne serez pas là pour intervenir dans leur vie.) Enfin supposez que c'est votre jumeau issu de fission qui vous est génétiquement identique et présente une évidente continuité psychologique avec vous. N'est-ce pas franchir un pas de plus? N'est-ce pas vous rapprocher de la situation où c'est vous le vainqueur? Et le fait qu'il y ait, en gros, deux fois plus de chances de réussir lorsqu'il y a deux survivants plutôt qu'un n'adoucit-il pas la différence?(Non seulement cela, mais vos jumeaux issus de fission pourraient, en coopérant, atteindre deux de vos objectifs qui s'excluent l'un l'autre.)

L'argument précédent tend à étayer l'idée que la fission personnelle vaut bien la survie ordinaire. Si Riker avait su qu'il était sur le point de se scinder, ça n'aurait pas été la fin du monde pour lui, même s'il avait cessé d'exister. Si cet argument est irrésistible, alors la survie personnelle n'est pas forcément une question d'identité personnelle (ou, si vous maintenez que la survie personnelle équivaut à l'identité personnelle, ce n'est pas une question très importante). Et la survie personnelle n'est pas une question d'identité psychologique, par ailleurs. Qu'est ce qui est important, alors? Qu'il y ait un élément de continuité assez proche de vous - appelons cela une vision « intellectuellement cohérente » de ce qu'une personne tient pour important dans sa survie personnelle. Lorsqu'une continuité psychologique suffisamment proche est momentanément désactivée, c'est l'identité personnelle qui prévaut. Puisque, à ce que nous savons, c'est le cas dans la survie ordinaire, il n'est pas surprenant que nous croyions à tort que c'est l'identité personnelle qui importe vraiment. Répétons-le, la continuité psychologique est ce qui importe pour la survie personnelle ordinaire. Si l'on préserve également une certaine identité, c'est un bonus sans importance. Et si l'on ne préserve pas l'identité (si Dieu scinde sans arrêt les personnes), c'est une perte sans gravité. Ce que nous ignorions ne saurait nous nuire, ni nuire à nos continuateurs psychologiques !

Continuité psychologique et justice

L'identité proprement dite est chose curieuse: plus on la considère, moins elle semble importante. Si la fission est largement préférable à la mort, cela montre que l'identité, mentale, personnelle, ou autre, n'est pas ce qui importe vraiment pour la survie personnelle. Mais quel sens donner alors aux intuitions de sens commun sur la justice, mentionnées plus haut? Faut-il en déduire que, si Riker commettait un crime, les deux descendants de fission, ne lui étant pas identiques, seraient acquittés? Voilà de quoi rendre le crime juteux: passez à l'acte juste avant de vous scinder!

Bien sûr, cela ne saurait être exact. Mais non pas parce que l'identité personnelle n'est pas ce qui importe dans la survie personnelle. Commencez par quelque chose de simple - la propriété de Riker. Nos intuitions sur la justice semblent ici assez claires: ses deux descendants de fission ont un droit égal à une part de ses biens (une intuition dont Riker 1 sent manifestement tout le poids). Il est compréhensible que Riker 2 se sente retardé dans sa carrière puisque Riker 1 a obtenu une promotion en récompense des actions de Riker, et il est clair que Riker 2 a droit à une partie des éloges. Ce qui vaut pour les avoirs de Riker semble également s'appliquer à ses dettes. S'il devait de l'argent, chaque moitié devrait être redevable d'une moitié de la dette. Cette idée peut s'étendre à tous les cas où la justice a des exigences normales, et s'applique déjà dans le monde tel que nous le connaissons. D'habitude, les individus agissant de concert sont responsables en commun des résultats de l'action; le fait qu'un membre du groupe ne soit pas identique au groupe n'a aucun poids.

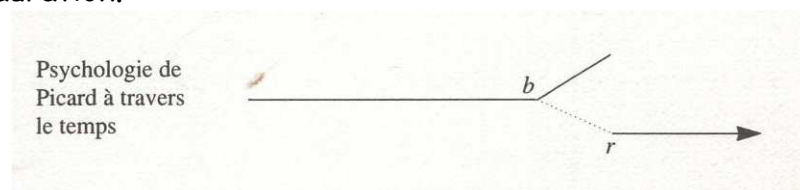
Cette intuition nous permet de régler le cas de Jadzia Dax. Puisque chaque Trill est une fusion entre deux personnes consentantes, il est utile de considérer les actes du Trill à travers l'analogie de deux personnes agissant de concert. La fusion ressemble plus encore à d'autres unions entre personnes consentantes dans lesquelles chaque personne prend la responsabilité des actes de l'autre. Jadzia - l'innocente - est ce sur quoi nous devrions nous concentrer ici. En acceptant la fusion, Jadzia réalise que certaines des actions de Jadzia Dax seront motivées par la partie de la psychologie du Trill héritée de Curzon Dax. Qui plus est, il semble que si Curzon Dax a fait une promesse à quelqu'un, même si Jadzia n'en sait rien, elle assume tacitement, en consentant à la fusion, que Jadzia Dax s'engage à honorer la promesse si possible. (C'est précisément ce qui se produit, notamment dans « Serment sur l'honneur » [DS9].) C'est pourquoi Jadzia accepte tacitement la responsabilité de Jadzia Dax pour toute mauvaise action commise par Curzon Dax. Si Curzon Dax est coupable, alors Jadzia Dax sera responsable.

Ce qui est merveilleux avec l'argument que je viens de donner, c'est qu'il nous dispense de trancher sur l'identité entre Curzon Dax et Jadzia Dax. C'est aussi bien, puisque la réponse la plus défendable à cette question est négative. Bien qu'il y ait sans doute proportionnellement plus de Curzon Dax que de Jadzia dans la psychologie de Jadzia Dax (puisque'il y a eu sept hôtes jusque-là, y compris Curzon), il y a trop de Jadzia pour qu'on ne tienne pas compte d'elle. Et bien que le corps du trill vienne de Curzon Dax, le corps humain vient de Jadzia. Puisqu'il est difficile d'établir si c'est à Curzon Dax ou à Jadzia que Jadzia Dax est identique, et qu'elle ne peut pas être

identique aux deux, la réponse la plus défendable est qu'elle n'est identique ni à l'un, ni à l'autre. Curzon Dax et Jadzia ont tous deux cessé d'exister, et une nouvelle personne est venue au monde.

Les choses ne sont pas aussi simples dans le cas de Tuvix. Un raisonnement parallèle mène à la conclusion que Tuvok et Neelix ont cessé d'exister lorsque la fusion s'est produite. Cela nous laisse avec deux questions difficiles : lorsque Tuvix s'est scindé en T et N, Tuvok et Neelix ont-ils été restaurés? Et le capitaine Janeway a-t-elle pris la bonne décision en scindant Tuvix? Il est clair que ni T ni N ne sont identiques à Tuvix (si l'on raisonne comme dans le cas de la fission de Riker). Tuvix avait donc raison de penser qu'il cesserait d'exister après la fission. Mais cela laisse ouverte la question de savoir si T et N sont tous deux identiques à Tuvix. On peut penser qu'ils ne sont pas identiques, en raison du blanc au cours duquel Tuvok et Neelix ont tous deux cessé d'exister - la période au cours de laquelle Tuvix a existé. Mais peut-être ce blanc n'est-il pas si important. Rappelez-vous que dans «Reliques» (NG), le schéma de Scotty était resté coincé dans le téléporteur pendant soixante-quinze ans et que nul n'a douté que ce soit Scotty qui en soit sorti. Toutefois, au cours de ce laps de temps, il n'y avait personne d'autre qui prétendît être Scotty - il n'y avait aucun élément de continuité du Scotty original ; l'analogie ne fonctionne donc pas.

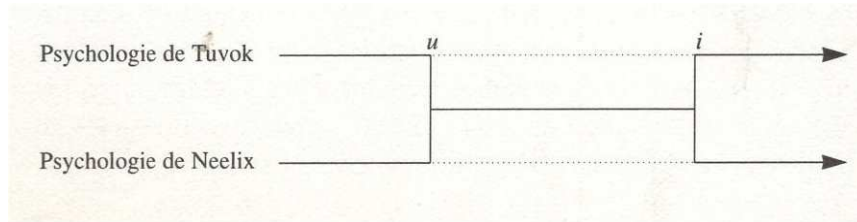
Examinons plutôt une fois de plus « Seul parmi nous », où Picard est téléporté hors du vaisseau sous forme d'« énergie pure », puis matérialisé selon le schéma stocké dans la mémoire du téléporteur. Comme on l'a noté dans le chapitre 4, le fait que Troi devine la présence de Picard « là-dehors » dans la nébuleuse suggère que sa psychologie a gardé sa continuité, et pourtant, lorsque Picard est reconstitué, il n'a aucun souvenir de ce qui s'est passé après la téléportation. Voici, sous forme de diagramme, les événements présumés - t est l'espace temporel de la téléportation, r celui de la restauration.



Notez que la psychologie de Picard maintient sa continuité sur la branche supérieure après b , mais seulement pour un court laps de temps. Lorsque son union avec l'être d'énergie pure échoue, cette branche de sa psychologie s'interrompt. C'est alors, à r , qu'apparaît une psychologie prolongeant la psychologie originelle de Picard. Le schéma de continuité permet de faire ici marche arrière. Il y a discontinuité : entre b et r , l'élément de continuité de la psychologie originale est représentée par la branche supérieure, mais après r , par la branche inférieure. Toutefois, on peut ne pas tenir compte de la discontinuité lorsqu'elle est insignifiante. (Comparez avec la discussion sur la résurrection des corps au chapitre 4). C'est pourquoi la personne qui sort en fin de compte de l'aire de rematérialisation semble finalement être bel et bien Picard.

Maintenant, voici la situation de Tuvix, où u est le temps de la fusion, et i celui

de la fission.



Une fois de plus, je pense que le schéma de continuité permet de revenir en arrière, et deux pistes distinctes apparaissent ici. Entre *u* et *i*, l'élément de continuité le plus proche de Tuvok (la branche supérieure avant *u*) est Tuvix, mais *T* (la branche supérieure) est l'élément de continuité le plus proche après *i*.

Janeway a-t-elle pris la bonne décision en ramenant Tuvok et Neelix? Il y a deux raisons pour penser que oui. Tout d'abord, ni Tuvok ni Neelix n'ont consenti à la fusion, et ils ne l'auraient pas fait si on leur avait demandé leur avis. En effet, d'une part, ils mènent des vies séparées, chacun avec ses proches, et elles seraient sérieusement compromises si Tuvix continuait d'exister. D'autre part, il y a le point de vue de Tuvix. Bien que Tuvok et Neelix survivent chacun dans *T* et *N*, ce n'est pas comme si Tuvix n'avait jamais existé. Chacun d'eux a la moitié du corps de Tuvix (et donc, tout comme celui-ci considérerait Tuvok et Neelix comme ses «parents», il devrait considérer *T* et *N* comme ses enfants puisque chacun d'eux a hérité de son capital génétique). Mais plus encore, chacun a la moitié de sa psychologie, et se rappelle avoir été Tuvix. Bien qu'ils se consolent moins que Riker (dans son cas, chaque descendant de fission est qualitativement identique au sujet original), c'est tout de même mieux que de mourir. Bien que ces raisons positives doivent tout de même être confrontées au désir de Tuvix de mener une existence continue, ce désir semble lui-même contrebalancé par le désir de Tuvok et de Neelix de ne pas être fusionnés. Je crois que, tout bien pesé, Janeway a pris la bonne décision : du moins, elle n'en a pas pris une mauvaise.

Je devrais dire à ce point que le cas de Tuvix est à mes yeux moins significatif que celui de Riker. En laissant de côté les incongruités flagrantes - ainsi, les vêtements de Tuvix sont un mélange de ceux de Tuvok et Neelix -, examinons un moment la façon dont deux psychologies distinctes pourraient être combinées, en préservant les traits distinctifs de chacune. S'il est vrai que souvenirs et expériences pourraient s'additionner, les états mentaux qui déterminent l'action - opinions et désirs - ne sauraient être cumulés. Les opinions seraient éminemment incohérentes, et les désirs absolument incompatibles. Au mieux, le résultat d'une telle fusion serait une personne qui ne saurait quoi penser ni quoi faire, et au pire, un goulash psychotique - peut-être pas même une personne.

À la croisée des chemins

J'ai interprété les cas de Riker, Jadzia Dax et Tuvix en me fondant sur le postulat suivant: l'identité disparaît lorsque la psychologie subit une fusion ou une fission. Mais le philosophe David Lewis a imaginé une réfutation ingénieuse de la perte

d'identité personnelle en cas de fission, et nous devrions peut être l'adapter aux cas de fusion. Lewis remet en question l'idée selon laquelle il n'y a qu'une seule personne avant la fission, et propose une analogie avec les autoroutes. Il est tentant d'assimiler les autoroutes aux routes, mais erroné, car parfois les deux se confondent. S'il nous faut les assimiler en invoquant la transitivité de l'identité, nous pouvons montrer que la carte se trompe lorsqu'elle indique une autoroute alors que la route a pris un embranchement ! Lewis affirme que nous pouvons considérer les personnes comme des autoroutes, et ce que j'ai appelé les tranches de personne (comme celle qui occupe votre chaise depuis ces dix dernières secondes) comme les routes. Lorsque la fission se produit, elle révèle le fait que deux personnes ont partagé la tranche de personne d'avant la fission, tout comme deux autoroutes peuvent se partager un tracé routier. Le nombre de personnes qui partagent une tranche de personne dépend du sort des éléments de continuité - une fission se produira-t-elle ou non ? Les personnes sont, aux yeux de Lewis, des «vers» de l'espace-temps, dont certains pourraient se croiser et partager à l'occasion des tranches de personne.

C'est pourquoi il n'y a jamais eu un seul et unique Riker: Riker 1 et Riker 2 existaient depuis toujours, mais personne ne l'avait remarqué parce qu'ils partageaient les mêmes tranches de personne jusqu'au moment de la fission. Si on applique ce tableau aux fusions, le cas assez irréaliste de Tuvix évoque deux grandes routes séparées qui se rejoindraient brièvement avant de se séparer à nouveau. Qu'il soit ou non correct, le point de vue de Lewis n'est en rien un retour à la notion de bon sens sur la survie personnelle. Dans la perspective de Lewis, et celle que j'ai préconisée, le prolongement de votre psychologie est celle d'une personne unique qui résiste au temps si et seulement si il n'y a pas de fission. Dans ma perspective, si vous vous scindez, vous n'existez plus, mais ce n'est pas si mal. Dans celle de Lewis, s'il y a une fission (qui ne sera pas répétée), alors il n'y a jamais eu de «vous» unique de toute façon, puisque chacune des tranches de personne depuis «votre» naissance jusqu'à la fission était en fait partagée par deux personnes. Ce n'est peut être pas plus mal, mais ça ne paraît pas mieux !

Changement de point de vue

Supposez que vous soyez le capitaine de l'Entreprise-D. Tandis que le vaisseau est dans la station spatiale d'arrimage, d'étranges voleurs subtilisent votre astronef et le mettent en pièces. Mais ce sont en fait des voleurs riches et philanthropes (j'avais bien dit qu'ils étaient étranges!) et laissent un vaisseau qualitativement similaire à l'ancien. Le nouveau vaisseau est identique à l'ancien à la fois dans sa fonction et dans son apparence. Qu'avez-vous perdu ? Vous n'avez plus le même vaisseau, et alors ? Maintenant, modifions ce cas. Supposons que ces voleurs philanthropes tentent de vérifier, avant de commettre leur crime, si vous êtes entièrement satisfait(e) de votre vaisseau. Leurs espions vous entendent vous plaindre de ci et ça, et vous rendent ainsi service : au lieu de vous rendre un vaisseau fonctionnellement identique, ils apportent çà et là quelques améliorations. Celles-ci, et

quelques retouches esthétiques, produisent des différences d'apparence. Qu'avez-vous perdu? Un peu de familiarité avec le vaisseau, peut-être, mais n'avez-vous pas gagné au change ? Supposons encore que vous soyez absolument insatisfait(e) de l'ancien Entreprise-D et que le produit de remplacement laissé par les voleurs soit donc absolument différent de l'original. Ne correspondrait-il pas alors à l'objet de vos désirs?

Supposons maintenant que vous soyez sur le point de subir une greffe instantanée et totale d'un corps et d'un cerveau synthétique - ou son équivalent, une simple téléportation par données. Jusqu'ici, nous avons simplement imaginé de remplacer votre corps par un équivalent fonctionnel et esthétique. Mais supposons que nous fassions quelques changements - un coup de ciseau ici, un rembourrage là, on élimine les poignées d'amour et la cellulite- sans coût ni dommage supplémentaires. Direz-vous en toute honnêteté qu'il n'y a rien dans votre corps que vous ne souhaiteriez changer, si l'occasion s'en présentait ? Quelques kilos de moins ? Quelques centimètres de plus ? De fait, pour bon nombre d'entre nous, obtenir le corps dont nous rêvons pourrait provoquer un changement fonctionnel et esthétique complet, une guérison ou un arrêt du vieillissement, voire un changement de sexe! Qu'y aurez-vous perdu? Un peu de familiarité avec vous-même, peut-être, mais n'aurez-vous pas gagné au change ?

Ou supposons que vous souffriez d'un handicap particulier. LaForge est né aveugle, mais a une vision prothétique grâce à son viseur et à certains implants cybernétiques (tout comme certains sourds, de nos jours, ont retrouvé l'ouïe grâce à des prothèses). Si nous pouvions simplement remplacer les yeux d'une personne aveugle par des yeux identiques, de bon fonctionnement; ou utiliser le téléporteur, où l'on aurait entré les informations génétiques correctes (comme dans « Sélection antinaturelle » et « Petits coquins ») pour rematérialiser des yeux en état de marche? Dans « Signes de vie » (V), la Viidienne Danara Pel souffre des effets ravageurs du «phage». Pour protéger son cerveau, le médecin de bord transfère momentanément sa psychologie dans un corps holographique. Mais au lieu de reproduire le corps malade, les projecteurs holographiques sont utilisés pour créer un corps immunisé contre le phage. Si ce genre de soin médical est possible, n'est-ce pas une bonne idée? C'est tout à fait cohérent avec l'idée que l'identité ou la continuité corporelles n'importent pas beaucoup pour la survie personnelle.

Mais à présent, réfléchissons un moment sur la psychothérapie. Quel est son intérêt? Changer la psychologie du patient, bien sûr - le débarrasser de ses désirs, opinions, souvenirs, attitudes, craintes indésirés, en les remplaçant par autant de sentiments désirables. Maintenant, supposons qu'on vous offre de procéder à ces changements à travers une simple téléportation par données. Puisqu'il nous faut extraire et stocker momentanément votre psychologie, pourquoi ne pas y apporter quelques retouches, éliminer un désir malsain ou un souvenir torturant? Pouvez-vous dire en toute honnêteté qu'il n'y a rien dans votre psychologie que vous ne désireriez changer, si vous en aviez l'occasion ? Quelques phobies en moins ? Quelques points de QI en plus? De fait, pour nombre d'entre nous, obtenir la psychologie que nous

aimerions pourrait susciter un changement radical en matière de fonction mentale, une psyché plus saine, voire un changement de personnalité! Qu'aurez-vous perdu? Peut-être pas même un peu de familiarité avec vous-même (vous pourriez ne plus vous rappelez qui vous êtes, si c'est cela que vous désirez), et n'aurez-vous pas gagné au change?

Et maintenant, additionnons les deux procédures, dans ce que j'appellerai la thérapie du téléporteur. C'est comme un passage chez le coiffeur. Avant de vous soumettre à la seule téléportation par données, vous nous dites ce que vous souhaitez, ou vous consultez nos brochures, les photos et descriptions de divers styles corporels et psychologiques. Une fois que vous avez fait votre choix, nous programmons le téléporteur, et presto ! votre nouveau «vous» est matérialisé. Quelle objection pourriez vous formuler contre une telle procédure ? Après tout, s'il est important pour vous de rester le même, personne ne va vous forcer à changer - c'est comme si on égalisait votre coupe chez le coiffeur, ou comme si on ne touchait à rien. Mais pour certains d'entre nous, la réalisation de nos désirs intimes de changement apportera une perte radicale en termes de continuité, physique et psychologique.

Vous noterez une progression dans les perspectives que nous avons étudiées dans ce chapitre et le précédent. Nous avons commencé par rechercher une âme immuable comme moyen de préserver ce qui est vraiment important dans la survie personnelle. Puis nous avons découvert qu'une âme immuable n'était pas nécessaire, et que nous pouvons survivre au changement sans perdre notre identité personnelle. Puis nous avons découvert qu'on peut au moins dire que nous survivons à la perte de notre identité personnelle. Maintenant nous découvrons qu'on peut dire au moins que nous pouvons survivre même sans continuité personnelle significative ! Le philosophe Raymond Martin a soutenu que, pour bien des gens, ce qui importe avant tout en matière de survie, ce n'est pas l'identité personnelle, ni même la continuité psychologique, mais plutôt de devenir la personne qu'ils rêvent le plus d'être. Un jour peut-être, la technologie de la téléportation telle qu'elle apparaît dans Star Trek mettra à l'épreuve ces intuitions. Reste une question: est-il ou non rationnel de subir des transformations aussi radicales? C'est ce sur quoi je vous laisse méditer.

L'ennemi intérieur

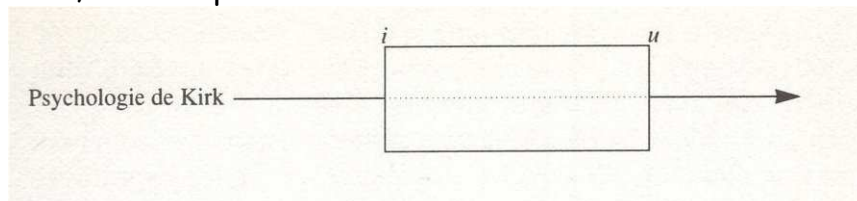
Un thème classique de Star Trek, c'est le «coup d'État psychologique». D'épisode en épisode, l'un ou l'autre des membres d'équipage est «physiquement envahi» - un extraterrestre l'investit et le manipule à l'avenant. D'habitude, cela supprime temporairement la véritable psychologie de la personne, comme dans les épisodes de La nouvelle génération intitulés « Seul parmi nous », « Conspiration », « Indices », et « Jeux de pouvoir », ainsi que le film Star Trek II : La colère du Khan. Parfois, ça tourne plutôt à la coexistence pacifique, comme dans « Retour à demain » (PS) où la psychologie de Spock occupe momentanément le corps de l'infirmière Chapel, où elle coexiste avec sa psychologie. Parfois, il y a une lutte d'influence, comme dans « Le schizoïde » (NG), où Data est envahi par la psychologie du génie

cybernétique, le Dr Ira Graves après la mort physique de cette scientifique, et où Graves prend peu à peu le contrôle, avec des effets dramatiques prévisibles.

Le thème qui montre plusieurs êtres investissant un corps unique est devenu une préoccupation de la vie réelle. La plupart d'entre nous connaissent bien ces histoires d'enfants maltraités, dont la personnalité se brise en mille morceaux, donnant lieu à un dysfonctionnement de personnalité multiple. (Dans «Masques» [NG], Data est pris d'assaut par toute une série de personnages distinctes!) Mais ce dysfonctionnement n'est que la partie visible d'un énorme iceberg, car il y a des preuves substantielles qui remettent en question ce qu'on a appelé l'unité de la conscience, selon laquelle une personne a un esprit. Plus spécifiquement, elle dit que chaque organisme humain normal se voit associer une seule conscience unifiée, un seul esprit, et exactement une seule personne.

S'il faut en croire Star Trek, il y a bien assez de place dans un seul corps pour plus d'une personne. Mais cela se passe-t-il uniquement lorsqu'il y a une sorte d'invasion venue de l'extérieur du corps ? Lorsque, dans « Seul parmi nous », on demande à la conseillère Troi pourquoi ses pouvoirs empathiques n'ont pas détecté la présence de l'être d'énergie qui a momentanément habité Worf et le Dr Crusher avant de s'installer chez Picard, elle affirme qu'elle lit normalement une «dualité» chez les humains, ce qu'elle illustre par la métaphore du croisement de routes, lorsqu'on hésite sur le chemin à prendre. Est-ce juste une façon d'excuser son échec, ou est-ce une façon de dire que les humains manifestent une dualité psychologique ?

Ajoutons à cela l'étrange accident de téléporteur dans «L'ennemi invisible» (PS), où Kirk subit une fission tout à fait différente de celle qui affecte Riker. Les deux Kirk qui en résultent ne sont pas identiques psychologiquement en termes qualitatifs - ils sont comme le jour et la nuit. L'un a hérité de la moitié violente, animale, «mauvaise» de Kirk, et l'autre de sa moitié cérébrale, pacifique, « bonne ». En fin de compte, les deux moitiés sont réunies. (Notez que le processus est l'exact opposé de ce qui s'est passé dans « Tuvix ».) En diagramme, cela donne ceci : i est le temps de la fission, u le temps de la fusion.



La branche supérieure entre i et u est le BK («Bon Kirk»), la branche inférieure le MK («Mauvais Kirk»), et la personne qui existe après u peut être appelée K. Tout d'abord, vous remarquerez que ni BK ni MK ne sont identiques à Kirk (pour les raisons invoquées dans le cas de Riker), ni à K (pour les raisons invoquées dans le cas de Tuvix). Mais, contrairement à Riker, Kirk ne cesse pas simplement d'exister. Puisque BK et MK existent pour un temps relativement court, et puisque K est l'élément de continuité de Kirk avant i, le schéma de continuité ignore les blancs et considère K comme identique à Kirk.

Mais ce n'est pas la raison pour laquelle j'ai introduit cet exemple. Comment le

téléporteur peut-il scinder Kirk en deux psychologies fonctionnant toutes deux? Ou il y a une extraordinaire coïncidence - d'une façon ou d'une autre, le téléporteur, en divisant la psychologie unifiée de Kirk en deux, réussit à produire deux psychologies réduites, mais viables - ou c'est vraiment une scission, ce qui implique qu'il y avait déjà deux psychologies présentes avant i. Cela signifie-t-il que Kirk est fondamentalement double comme le seraient tous les humains d'après la conseillère Troi ?

Nous ne sommes pas seuls

Il y a un élément étonnant qui semble aller dans le sens de Troi, mais, avant de l'examiner, il nous faut ajouter quelque chose à la discussion sur l'esprit et sur les personnes des chapitres 1 à 3. Dans les discussions précédentes, nous tentions d'établir des critères nécessaires pour les esprits et les personnes; nous essayions d'établir les propriétés qu'il faut posséder pour passer pour un être intelligent, ou pour une personne. Mais nous n'avons pas examiné le critère signalétique: nous ne nous sommes pas demandé comment distinguer un esprit d'un autre, une personne d'une autre. En d'autres termes, lorsqu'on est confronté à un ou plusieurs esprits ou personnes, comment savoir combien ils sont? Comment recenser les esprits, ou compter les personnes ? La réponse de bon sens à cette question, raffinée à plusieurs titres, c'est qu'il faut recenser les organismes humains vivants.

Mais il y a d'autres moyens. Supposons que vous appeliez une entreprise et que vous parliez affaires avec un nommé Pete. Le lendemain, vous rappelez et demandez votre interlocuteur de la veille. Mais l'homme qui décroche le téléphone dit ne pas savoir qui vous êtes et ne rien savoir de votre conversation précédente. A supposer que vous croyiez cet homme, vous conclurez sans doute que vous êtes en contact avec un autre Pete, une personne différente, un esprit différent. Pourquoi ? Parce que vous supposez que, si cet homme était le premier Pete, il n'aurait pas oublié si vite votre cas. Les états mentaux d'une personne sont privés - si l'on excepte la perception extrasensorielle-, personne n'y a le même accès. Le premier Pete aurait introspectivement accès au souvenir de la conversation ; puisque le second Pete n'a pas d'accès introspectif à ces souvenirs, ce n'est pas du tout le même Pete. Maintenant, examinez les faits suivants.

Vous avez sans doute entendu parler des hémisphères droit et gauche du cerveau. Dans les années soixante, le psychologue Roger Sperry et ses collègues menèrent des expériences sur des patients qui avaient subi une opération chirurgicale (une commissurotomie, ou bissection du cerveau) qui a coupé les connexions cérébrales principales (corpus callosum) entre les deux hémisphères cervicaux. C'est une procédure d'expérimentation typique. Supposons que vous ayez un écran opaque à travers lequel vous pouvez passer vos mains. De l'autre côté de l'écran, il y a un bol contenant des billes (que vous ne pouvez pas voir). L'expérimentateur vous demandera de ramasser quelques billes dans votre main, d'en deviner le nombre, puis d'en désigner le nombre avec votre main droite. Ensuite, vous devrez dire ce nombre à

haute voix. La procédure est répétée pour la main gauche. Toutes choses égales par ailleurs, vous vous tirerez très bien de cette simple expérience. Mais voilà ce qui se produit lorsque le sujet a subi une commissurotomie (nous l'appellerons un patient au cerveau scindé, ou PCS). Il se conduit de façon normale pendant la première moitié de l'expérience, lorsqu'il utilise la main droite. Lorsqu'on passe à la gauche, le PCS typique signalera le bon nombre de billes ramassées, mais, lorsqu'on lui en demandera de prononcer le nombre, il se comportera exactement comme s'il hasardait un chiffre probable. C'est pourquoi, si le nombre est 4, le sujet dira n'importe quel chiffre bas, mais sans énoncer quelque chose de dément comme 2 034. Dans une autre expérience, un sujet verra s'illuminer brièvement le mot « porte-clefs » au centre de son champ visuel, puis on lui demandera de sentir de sa main gauche des objets sans les voir et d'en choisir un qui corresponde au mot qu'il a vu. Contrairement aux sujets normaux, le PCS choisira une clef, mais lorsqu'on lui demandera d'identifier verbalement ce qu'il a vu, il dira « porte ». En somme, le résultat de ces expériences résiste à toute une série de tests. Sperry rapporte: « L'un des traits les plus généraux et aussi les plus intéressants, voire frappants de ce syndrome peut être résumé comme un dédoublement apparent dans la plupart des domaines de la prise de conscience. Au lieu d'avoir un flux de conscience normalement unifié, ces patients se comportent souvent comme s'ils avaient deux flux de conscience indépendants, situés dans l'un et l'autre hémisphère, dont chacun est distinct, et sans contact avec les expériences mentales de l'autre. En d'autres termes, chaque hémisphère semble avoir ses propres sensations séparées et propres, ses propres perceptions, ses propres concepts et impulsions à agir, avec des expériences volitives, cognitives et d'apprentissage reliées entre elles. Après l'intervention chirurgicale, chaque hémisphère a donc sa propre chaîne de souvenirs, rendus inaccessibles au processus de remémoration de l'autre. »

Dans le cas des billes, par exemple, l'hémisphère gauche reçoit des informations des mouvements des doigts, et puisque chez presque tout le monde le centre de locution est situé dans l'hémisphère gauche, il n'a pas de mal à produire la réponse verbale correcte. De même, l'hémisphère droit reçoit des informations des mouvements des doigts de la main gauche et les contrôle : elle n'a aucun mal à signaler le nombre correct. Mais le discours est produit par l'hémisphère gauche, et il semble que celui-ci ignore littéralement la réponse verbale correcte - qu'il ignore combien de billes se trouvent dans la main gauche, et hasarde une réponse probable.

L'un des aspects de cette dernière expérience propre à perturber le sens commun, c'est que l'hémisphère gauche ne dit pas simplement: « Je ne sais pas. » Puisqu'il n'y a aucune raison de supposer qu'il ment, ce phénomène apparaît comme une confabulation. Comme on l'a déjà souligné plusieurs fois dans les chapitres précédents, l'un des postulats de bon sens qu'il nous faut rejeter, c'est l'idée que tout le monde sait exactement ce qui se passe dans son propre cerveau. L'introspection n'est pas un guide infallible de la vie mentale. De fait, certains mécanismes théoriques du cerveau sont hors de portée de la conscience introspective; le cerveau ne cesse d'inventer des hypothèses plausibles sur l'expérience. Voici un simple exemple de confabulation quotidienne. Regardez le diagramme ci-après, qui est

blanc excepté un petit carré et un petit rond. Fermez votre œil gauche, placez le diagramme à une trentaine de centimètres de votre visage, et fixez le carré de l'œil droit. Maintenant, déplacez lentement la page vers vous en fixant constamment le carré jusqu'à ce que le cercle disparaisse. Vous avez localisé votre point aveugle - point où le nerf optique traverse la rétine. La raison pour laquelle le cercle disparaît est que vous ne pouvez pas voir ce qu'il y a dans votre point aveugle, et que votre cerveau remplit ce blanc dans votre espace visuel en essayant de deviner ce qu'il y a. Puisque la surface environnante est blanche, votre cerveau vous dit que cet espace est blanc lui aussi. Vous remarquerez que votre cerveau agit sans votre collaboration. De fait, sans l'expérience vous n'auriez jamais remarqué qu'il y a un point aveugle dans chacun de vos yeux. (Pour tester l'œil gauche, tournez la page à l'envers.)



Des confabulations encore plus libres apparaissent dans le cas de pathologies mentales. Par exemple, le syndrome de Korsakoff fait perdre au sujet sa mémoire immédiate. Mettez un verre dans la main du patient et il inventera une histoire impressionnante pour justifier sa présence. Il ne ment pas ; il ignore juste que c'est une invention. C'est sans doute le même processus qui se produit chez le PCS. Après tout, comment un sujet peut-il ignorer ce qui est dans sa propre main ? Le lobe gauche invente sans doute l'histoire la plus plausible qu'il peut, ce qui ne remet pas en cause l'idée que chacun sait ce qu'il a entre ses propres mains. Le lobe gauche hasarde une possibilité et élimine les erreurs.

Sperry est assez prudent dans sa description du comportement des PCS. Mais le philosophe Roland Puccetti expose un argument en trois étapes qui l'amène à une conclusion assez étonnante. Demandez-vous quelle est la meilleure explication du comportement d'un PCS. Assurément, la situation ressemble à celle de votre contact téléphonique avec les deux Pete. Nous diagnostiquons chez le PCS un blocage total des contacts entre les deux hémisphères, et pourtant chaque hémisphère est par ailleurs un cerveau complet, avec ses propres opinions, désirs, souvenirs, intentions, etc. (Pour étayer cette idée, notez qu'un jour on a vu un PCS assis devant la télévisions avec un journal ; chaque fois qu'une main amenait le journal devant les yeux, l'autre le poussait de côté pour mieux voir la télévision.) La conclusion évidente, c'est que nous avons affaire à deux cerveaux et à deux personnes bien distinctes, dont chacune habite dans un hémisphère du cerveau.

Maintenant, demandez-vous quelle est la meilleure explication au fait que les PCS abritent deux cerveaux et deux personnes. Il semble y avoir deux possibilités. L'une, c'est qu'il y a un seul esprit et une seule personne qui s'est scindée en deux lorsqu'on a procédé à la commissurotomie; en d'autres termes, la personne originale a fissionné. L'autre, c'est qu'il y a toujours eu deux esprits et deux personnes, et que l'opération a révélé leur existence. Est-il possible que ce soit là un cas de fission? Vous remarquerez que, contrairement au dédoublement de Riker dans le téléporteur, il n'y a pas de procédure de dédoublement ici. Tout ce qui s'est passé, c'est qu'un chirurgien a pris un scalpel et a tranché dans un paquet de fibres nerveuses. Est-il

plausible que cela crée en quelque sorte au moins un cerveau flambant neuf, et au moins une personne flambant neuve ? Pas vraiment. Il est bien plus plausible qu'en coupant les connexions cérébrales on révèle l'existence de deux esprits et de deux personnes qui étaient là dès le début. Nous devrions donc conclure que les PCS abritaient deux esprits et deux personnes depuis toujours.

Troisième étape: demandez-vous si les PCS présentaient une particularité avant l'opération ? La plupart d'entre eux étaient épileptiques au dernier degré, et l'opération visait à prévenir des crises potentiellement dangereuses chez ces sujets. (De ce point de vue, la procédure a eu un succès mitigé : trancher le corpus callosum ne fit que prévenir la contamination des crises d'un hémisphère à l'autre.) Et sinon ? Aucune particularité-ils avaient des cerveaux structurés exactement comme vous et moi : avant la commissurotomie, leurs cerveaux étaient représentatifs de ceux du restant de la population. La conclusion qui s'impose, c'est que tout organisme humain normal abrite deux cerveaux et deux personnes. Comme j'aime à le dire, chacun de nous est en fait une paire de jumeaux siamois, toujours soudés à l'endroit du cerveau!

On pourrait opposer deux objections majeures à l'argument de Puccetti. Comme l'a souligné le philosophe Thomas Nagel, l'argument suivant a une conclusion qui ne cadre pas avec le postulat de base. Supposons que l'on conclue qu'un PCS a deux cerveaux en établissant une comparaison entre les PCS et les humains normaux. Puisque le comportement des PCS diffère du nôtre, nous concluons qu'ils ont deux fois ce que nous avons, et donc qu'ils abritent deux esprits et deux personnes. Mais nous ne saurions conclure que les humains normaux ont en eux deux esprits et deux personnes sans nous faire accuser d'incohérence ! Vu la forme sous laquelle j'ai présenté l'argument de Puccetti, celui-ci résiste pourtant à l'objection. Nous partons de notre conception de l'esprit, et des critères identitaires signalétiques de l'esprit, et nous appliquons directement ces critères au comportement du PCS. Il n'entre dans la démonstration aucun postulat sur les organismes humains normaux, et il n'y a donc pas d'incohérence.

La seconde objection est plus prometteuse. Si Puccetti a raison, alors chacun de nous est exactement comme Data après qu'il a été envahi par le Dr Graves. Nous ne sommes pas tout seuls ! Mais le conflit interne entre Data et Graves se manifeste dans le comportement, ce qui permet à Picard de déduire que Data n'est pas tout seul. Si Puccetti a raison, nous devrions observer un conflit chez les organismes humains normaux - comme celui que révèle l'homme dont les deux mains se battent au sujet du journal et de la télévision. Puisque nous n'observons pas ces conflits, la conclusion de Puccetti doit être erronée.

La réfutation se fera en deux parties. Tout d'abord, séparons l'état mental en deux domaines - informationnel et volitif. Les opinions, souvenirs et expériences sensorielles relèvent de l'information, les désirs et intentions de la volition. Dans un organisme humain où le corpus callosum est intact, un échange quasi instantané d'informations traverse les hémisphères (et la plupart des expériences des deux hémisphères sont quasi identiques de toute façon.) C'est pourquoi on s'attendrait à ce que chaque hémisphère ait presque exactement les mêmes états informationnels, et

que ces états ne soient pas cause de conflit. Passons aux états de volition : puisque les deux hémisphères partagent le même intérêt dans le corps, il y aura sans doute intersection impressionnante de désirs et d'intentions, là aussi. Mais le recoupement ne sera pas assez parfait pour éviter tout conflit, et qu'est-ce qui explique donc l'unité apparente des états de volition ? Puccetti souligne ici le fait que presque tout le monde a son centre locutoire dans l'hémisphère gauche, et il spéculé que cet hémisphère gauche domine littéralement l'hémisphère droit. Le gauche est une brute qui contrôle non seulement la parole mais aussi la plupart de ce que fait le corps en termes de volition. La plupart du temps, le droit est un complice impuissant des actions du gauche.

Cette réfutation répond aussi à une troisième objection qui peut vous avoir traversé l'esprit. Si vous êtes deux, pourquoi avez-vous l'impression d'être un ? La réponse, bien sûr, c'est que le « vous » en question est sans doute le gauche - et que vous avez eu le contrôle si longtemps que le droit est pratiquement invisible pour vous. Si vous dites votre opinion du moment, et que le droit vous entend, il pense sans doute : « Je suis là, espèce de brute ! » C'est ici, si nous acceptons cette vision schizoïde de l'être humain, que se pose un problème. Si le gauche commet un crime où le droit participe à contrecœur, punir le gauche, c'est aussi punir le droit, qui est innocent. Et peut-être le droit se vengera-t-il par instants, en exploitant un relâchement du gauche pour lui faire des ennuis !

Tout cela correspond bien à ce qu'affirme Troi, qui sent la dualité chez les humains. Parce qu'elle est empathique, elle peut sentir le conflit entre le gauche et le droit même lorsqu'il n'apparaît pas dans le comportement. Et le phénomène où l'on « est au croisement des chemins » n'est pas si étrange - on sent comme un conflit de deux volitions qui essaient chacune de prendre le contrôle du corps.

Vous pourriez avoir l'idée de demander pourquoi le corps nous semble avoir deux cerveaux. Notez que les Klingons, comme le lieutenant Worf, possèdent toutes leurs parties corporelles en double. Les avantages de cette combinaison sont exposés dans l'épisode de La nouvelle génération intitulé « Éthique » où Worf semble sur le point de mourir, puis un système de soutien se déclenche et il retrouve la santé. Cette duplication en bloc est particulièrement utile à une race guerrière, mais nous aussi, nous avons des systèmes de soutien. Nous avons deux gonades, deux reins, deux poumons et pouvons survivre et nous reproduire si nous perdons l'un des deux. Peut-être le cerveau est-il fait de deux cerveaux soudés ensemble, et nous en avons deux parce que le cerveau est si important que ça vaut le coup d'avoir une roue de secours. Les pathologies du cerveau confirment cela : un organisme humain peut survivre à la destruction d'un hémisphère. Si la destruction se produit au tout début de l'existence, l'organisme fonctionnera presque normalement. Même si c'est l'hémisphère gauche qui est détruit : le droit s'occupera de la production de parole avec peu ou pas de difficulté si la perte se produit dans la toute première enfance.

Méditez un moment ce que vous seriez si l'hémisphère droit gouvernait tout - et bien sûr, vous, les droits, vous savez ce dont je parle ! Ne serait-ce pas horriblement frustrant ? Passer toute une vie à l'ombre de l'hémisphère gauche

pourrait avoir certaines conséquences psychologiques étranges. Pensez-y, et examinez certains phénomènes bien connus du grand public qui n'ont pas d'explication naturaliste concluante, comme la « possession démoniaque », l'écriture automatique et l'hypnose. Dans l'écriture automatique (et autres phénomènes voisins, comme le oui-ja), un sujet produira des écrits cohérents sur lesquels il dira n'avoir aucun contrôle. Souvent, cette écriture est le produit de la main gauche, et souvent, chose frappante, elle donne lieu à une écriture en miroir, qu'on ne déchiffre qu'en tenant la page devant une glace. (Un sujet a produit cinq romans, apparemment en écriture automatique.) En hypnose, il existe un phénomène connu sous le nom de l'observateur caché, découvert assez récemment par le psychologue Ernest Hilgard, alors qu'il démontrait une surdité hypnotique chez un sujet aveugle. On peut dire à un sujet hypnotisé: « Lorsque je compterai trois, vous serez sourd. Vous n'entendrez plus rien du tout, et vous resterez ainsi jusqu'à ce que je pose ma main sur votre épaule droite. » On peut alors produire toutes sortes de bruits et discours sans éveiller la moindre réaction chez le sujet qui, après l'hypnose, ne se rappellera rien. Hilgard a découvert qu'il pouvait entrer en contact avec « une partie » du sujet hypnotisé, qui entendait et se rappelait tout. Hilgard baptisa cette partie l'observateur caché.

Peut-être l'existence du droit et du gauche explique-t-elle ces phénomènes en partie ou en totalité - peut-être pas. Elle peut aussi expliquer les phénomènes de médium et les « canalisations » du New Age, voire éclairer les phénomènes que le psychologue Julian Jaynes a cru détecter chez nos ancêtres : ils entendaient littéralement des voix dans leur tête qui leur disaient quoi faire (avant que les gauches ne prennent le contrôle du corps?). D'un autre côté, elle ne peut certainement pas expliquer entièrement le dysfonctionnement de personnalité multiple, et nous devons peut-être conclure qu'aux divisions entre les deux hémisphères de la conscience pourraient répondre toutes sortes de divisions de conscience mal comprises, à l'intérieur de chaque cerveau. Comme d'habitude, plus on en découvre, moins on en sait.

Une chose paraît certaine, toutefois: postuler l'unité de la conscience ne nous mènera pas très loin ! Tout comme les éliminativistes prédisent que nous découvrirons que la conscience est une idée vide (voir chapitre 3), la notion même de « personne » et de « moi » pourrait alors se révéler une fiction. Peut-être n'y a-t-il pas de véritable contrôle exécutif dans le cerveau - peut-être les choses se produisent-elles comme ça, comme elles arrivent dans la vie ordinaire. Le philosophe Daniel Dennett et le psychologue Nicholas Humphrey comparent le moi' à un chef de gouvernement. Le chef de gouvernement a une fonction essentiellement symbolique, avec peu de pouvoir concret, mais il peut être bon pour la nation de vivre dans l'illusion qu'il y a à la barre quelqu'un de responsable. De même, il peut être utile à l'organisme humain de confabuler l'illusion que chacun de nous a une personne unique à la barre, qui dirige nos actes de façon rationnelle.

Mais supposez qu'il y ait vraiment plusieurs personnes. Si le postulat de l'unité de la conscience est faux, les implications éthiques potentielles en matière de punition et de récompense sont sérieuses, comme je l'ai dit plus haut. Cela induit certaines

conséquences en matière de thérapie, par ailleurs. Puisque nous considérons normalement l'unité stricte comme quelque chose de normal et de souhaitable, nous supposons naturellement que les sujets dissociés ont besoin d'être guéris, et nous célébrons la réintégration des personnalités dans les cas de personnalités multiples. Toutefois, il arrive que le sujet réintégré fasse état d'un chagrin profond et d'un sentiment de deuil, comme à la mort d'un proche. Cela dit, supposons que l'unité soit vraiment souhaitable et que toute organisation humaine soit vraiment un couple de frères siamois soudés au cerveau. Il y a une autre application pratique de la thérapie du téléporteur que j'ai décrite plus haut. Il suffit de scanner le gauche et le droit indépendamment, et de matérialiser deux sujets, dont l'un est l'élément de continuité du gauche tandis que l'autre est celui du droit. Bien sûr, le sujet original pourrait s'opposer à cette séparation, mais ce serait simplement exprimer l'opinion du gauche, et le droit a le droit d'être libéré du gauche, non?

Ce qui nous ramène à la fission de Kirk dans « L'ennemi intérieur». Peut-être le Bon Kirk est-il le gauche, et le Mauvais Kirk, le droit (il semble avoir quelques difficultés d'énonciation). Est-ce que vous ne passeriez pas un peu les bornes si vous étiez affranchi de votre servitude au gauche? Et ne combattriez vous pas toute fusion avec le gauche, et tout retour à la servitude, comme le fait le «Mauvais Kirk»? Apparemment, l'épisode était censé souligner les différences entre les fonctions cognitives supérieures de l'esprit, et ses fonctions inférieures et animales (voir Yvonne Fern, *Gene Roddenberry: The Last Conversation*), mais on a du mal à comprendre comment un accident de téléporteur pourrait produire une telle fusion. Un clivage gauche/ droit semble bien plus plausible.

Chapitre 6

Les distorsions temporelles

Picard: « Vous l'avez dit vous-même, ce n'est qu'une possibilité. »

Dr Crusher: « Mais vous avez été dans le futur - vous savez que cela va se produire. »

Picard: « Je préfère considérer que le futur n'est pas gravé dans la pierre. »

« Les meilleures choses ... » (NG)

Data: « On ne peut pas tricher avec le destin. »

Picard: « Peut-être pas, monsieur Data, mais on peut toujours essayer. »

« La flèche du temps 1 » (NG)

Riker : « Oh, je le sais bien.

On ne peut esquiver le futur. »

« Le temps au carré » (NG)

Dans La physique de Star Trek, Lawrence Krauss dénombre vingt-deux épisodes de la première série et de La nouvelle génération où l'on recourt au voyage dans le temps. Ajoutez à cela trois filins et quelques épisodes de la suite de la série, notamment « Le temps composé I & II » (PE9) et « La fin du futur I et II » (V). Il faut le reconnaître, dans Star Trek, on voyage beaucoup dans le temps. L'intérêt de Krauss pour le voyage dans le temps - naturellement- est de l'ordre de la science physique: il se demande si le voyage dans le temps est possible d'un point de vue physique. Mais de nombreux philosophes se sont demandé si le voyage dans le temps est possible d'un point de vue logique, autrement dit, si l'idée même renferme une contradiction. S'il est logiquement impossible, nous pouvons épargner aux physiciens bien des problèmes, puisque aussi bien rien ne saurait être logiquement impossible et physiquement possible.

Dans cette recherche, plusieurs paradoxes du voyage dans le temps ont été avancés. Mais, avant tout examen, mieux vaut saisir exactement ce qu'on entend par «

voyage dans le temps». Tout d'abord, il ne suffit pas de se croire dans une autre époque (comme dans le Conte de Noël de Dickens, où Ebenezer Scrooge voit des bribes de son passé et de son avenir potentiel). Cela ressemble vraiment trop à une hallucination. Scruter une boule de cristal (véritable!) ne sert pas non plus à grand-chose, parce que, de toute évidence, on reste dans le présent. La condition requise pour être considéré comme un voyageur dans le temps, c'est de se rendre dans une autre époque.

Mais le voyage dans le temps diffère encore de ce déplacement d'une époque à l'autre. Vous étiez là l'an dernier, vous êtes là aujourd'hui: cela seul fait de vous un voyageur dans le temps. Le voyage dans le temps, c'est quelque chose qui sort de l'ordinaire, qui implique forcément une sorte de déviation par rapport au déroulement ordinaire du temps.

Mais quelle déviation? Je ne pense pas vivre assez vieux pour voir le XXIV siècle, mais je peux me tromper. Supposez qu'on me place en état de vie suspendue et qu'on me réveille, tout comme Rip Van Winkle, dans quatre siècles jour pour jour. Pour moi, il se sera à peine écoulé une minute - et pourtant, ce n'est pas là un véritable voyage dans le temps. Et pourquoi? Parce que j'ai réellement été «là» tout le temps, à subir le passage du temps avec toute une succession d'autres gens; j'ai simplement eu le bonheur de pouvoir ignorer son écoulement. Pour y voir plus clair, imaginez qu'on m'ait implanté dans le crâne une pendule fiable. L'écoulement du temps sur cette horloge sera conforme à l'écoulement ordinaire du temps, tel qu'il se produit en dehors de moi (c'est-à-dire quatre siècles). J'aurai dormi trop longtemps, et non voyagé dans le temps (tout comme Scotty dans «Reliques» [NG]).

Cela vaut pour le cas inverse. Si je rêve ou éprouve - à tort - le sentiment qu'un temps considérable s'écoule, ce n'est pas davantage un voyage dans le temps. Picard fait ce genre d'expérience dans deux épisodes de La nouvelle génération. Dans «Lumière intérieure», une sonde mentale maniée par un extraterrestre le contraint à vivre l'expérience de quarante années en moins de vingt-cinq minutes de temps réel. Dans un épisode digne de Capra, «Tapisserie», il est sérieusement blessé. Tandis qu'il agonise sur la table d'opération de l'infirmierie de l'Entreprise-D, le tout-puissant Q lui apparaît, à lui seul. Picard lui avoue qu'il regrette les choix qu'il a faits dans sa jeunesse, et Q lui offre la chance d'en revivre quelques-uns. L'expérience s'étend sur nombre d'années (par fractions), sans que Picard quitte la table d'opération. Dans un cas comme dans l'autre, une pendule fiable implantée dans la crâne de Picard aurait enregistré le temps qui s'est réellement écoulé à ses yeux, et cette durée aurait été conforme au temps externe.

Voilà qui suggère une condition nécessaire au voyage dans le temps. L'écoulement du temps aux yeux du voyageur doit différer réellement, et non en apparence, de l'écoulement du temps externe. Le philosophe David Lewis parle d'une différence entre temps personnel et temps externe. Ainsi, je voyage dans le temps dès lors que le temps personnel de mon voyage diffère du temps externe qui s'est écoulé. Si mon voyage prend une heure de mon temps personnel, et que j'arrive quatre siècles plus tard en temps externe, alors c'est que j'ai voyagé dans le temps.

Une disproportion de cet ordre suffit-elle à caractériser le voyage dans le temps ? Question difficile. Pour ma part, j'en te. En voyageant tout simplement rapidement, on peut trouver disproportion temporelle - comme le sait le moindre débutant en sciences physiques. Je peux m'engager dans un voyage très long et très rapide, qui prendra dix ans, puis réaliser, à mon retour, que sur terre cinquante ans se sont écoulés. Même si cela permet d'accéder plus rapidement au futur, cela ne présente pas l'espèce de symétrie que révèlent les intrigues de Star Trek, puisque aussi bien je ne puis me rendre dans le passé par ce moyen. Le véritable voyage dans le temps semble impliquer la possibilité d'un aller-retour. Le fuseau horaire créé par l'inventeur Berlingoff Rasmussen dans « Une question de temps » (NG) est caractéristique à cet égard - grâce à lui, on peut circuler librement dans le temps, en avant et en arrière. Cela indique une condition relative à la nature du temps lui-même: il doit être une dimension, comme les dimensions spatiales. Beaucoup de philosophes, sinon la plupart, acceptent aujourd'hui cette condition, et adoptent le point de vue dit du quadridimensionnalisme. Si l'on en croit le quadridimensionnalisme, les autres époques sont tout simplement d'autres endroits. Dire « maintenant », c'est comme dire « ici ». De même qu'il existe d'autres endroits même si vous ne vous y trouvez pas à l'heure qu'il est, de même il existe d'autres époques, même si vous ne vous y trouvez pas à l'heure qu'il est. Ces autres endroits sont ailleurs, ces autres époques sont ailleurs.

Longueur, largeur, hauteur et temps

Au chapitre 4, j'ai rendu compte d'une conception de l'identité personnelle qui présuppose la quadridimensionnalité des personnes. Moi-maintenant ne suis qu'une tranche temporelle du ver quadridimensionnel de l'espace-temps qui a nom Richard Hanley, et vous-maintenant n'êtes qu'une tranche temporelle du ver quadridimensionnel de l'espace-temps qui vous constitue. À en croire le quadridimensionnalisme, nous devrions voir l'univers tout entier (passé, présent et avenir) sous cet angle, comme le père de tous les vers de l'espace-temps. Considérez l'analogie suivante: l'univers nous apparaît à nous autres, gens ordinaires, comme un film apparaît aux spectateurs d'une salle de cinéma. Notre expérience se limite, quel que soit l'instant choisi, à un extrait du film au « présent ». Nous pouvons reconstituer ce qui s'est passé avant de deux manières : en faisant appel à notre mémoire (histoire), et en tirant des conclusions à partir du présent. Nous pouvons aussi tenter de prédire l'avenir par ce moyen. Mais nous ne saurions vérifier indépendamment la justesse de ces souvenirs, ni de ces prédictions. Nous sommes pris au piège du présent. Dieu est censé être radicalement hors du temps (ou du moins, toutes les époques réunies sont pour Dieu ce que le présent est pour nous), donc Il a d'autres moyens de connaître passé et futur. Pour reprendre l'analogie avec le film, Dieu voit toutes les images à la fois. Qui voyage dans le temps se situe quelque part entre Lui et nous. Comme nous, il ne peut saisir qu'une époque à la fois, son présent, mais comme Dieu, il peut examiner directement ce que nous considérons comme les autres époques

- le passé et le futur. Pour étendre un peu l'analogie avec le film, sa machine temporelle est comme une télécommande qui lui permet de dérouler vers l'avant ou de dérouler vers l'arrière.

Avant d'examiner de plus près le quadridimensionnalisme, étudions son principal rival, le présentisme, considéré d'ordinaire comme la conception commune du temps. Le présentisme avance que seul le présent est réel. Il est exclu que le passé et le futur existent. Bien sûr, le passé a existé, jadis (sous forme de présent), mais n'existe plus. C'est passé, clos, terminé, sinon oublié. Et le futur existera, mais seulement lorsqu'il aura été du présent. (Attention à bien distinguer le quadridimensionnalisme du présentisme. Ainsi, le présentiste pense que le futur n'existe pas encore ; le quadridimensionnaliste, lui, convient qu'il n'existe pas maintenant. Soit une analogie spatiale : je suis à Melbourne, et New York existe certainement, mais pas ici.) Qui plus est, à la différence du quadridimensionnaliste, le présentiste postule une importante asymétrie entre le passé et le futur. Le passé est advenu comme il est advenu, et nous pouvons désormais en déduire certaines vérités. (Nous savons si Oswald a abattu Kennedy, oui ou non, point final.) Mais le futur est un assemblage de pures possibilités et n'est pas fixé, même s'il est vrai et bien vrai (à en croire la plupart des présentistes, en tout cas) qu'il prendra une certaine tournure et pas une autre. Les présentistes voient un peu le futur comme une fermeture Éclair. Le présent, c'est l'endroit où se trouve le fermoir, le passé, c'est la partie déjà refermée, et l'avenir, c'est ce qui baye encore. L'écoulement du temps est représenté par le mouvement du fermoir, qui passe d'un présent au suivant-transformant les possibilités futures en faits accomplis, pour ainsi dire.

Mais le présentisme n'accorde aucune place au voyage dans le temps. Si le futur n'existe pas avant d'être devenu le présent de par l'écoulement normal du temps, comment gagner «ce moment» en accéléré ? Il n'y a pas encore de « moment » où il serait possible de se rendre ! Et si le passé est passé, n'existe pas, et n'existera plus, il n'y a strictement aucun « moment ». Le présentisme, semble-t-il, prive le voyageur d'un ailleurs !

Je ne crois pas à la vérité du présentisme. Pour une seule raison: s'il est vrai, alors on ne peut strictement rien affirmer sur le futur. Il n'est pas vrai qu'Un tel gagnera la finale du 100 mètres hommes aux jeux Olympiques de Sidney en l'an 2000. Il n'est pas vrai que les jeux Olympiques auront lieu à Sidney en l'an 2000. Il n'est pas vrai que des jeux Olympiques auront lieu en l'an 2000. Mais il n'est pas vrai non plus qu'il n'y aura pas de jeux Olympiques en l'an 2000. Il n'est pas vrai qu'il y aura un an 2000, pas vrai non plus qu'il n'y aura pas d'an 2000. Il est vrai que les choses prendront une tournure et une seule, mais il n'est pas une seule tournure dont on puisse dire: c'est ainsi que cela se produira. (Il n'est même pas vrai qu'en l'an 2001 il sera vrai qu'il a existé une année 2000, même si aujourd'hui il est certain qu'il y a eu une année 1900.) De plus, s'il n'existe aucune vérité particulière sur le futur, alors personne ne peut savoir quoi que ce soit à son sujet, pas même Dieu. L'un dans l'autre, j'aime autant adopter quelques vérités particulières sur le futur, et suis enclin à rejeter le présentisme pour cette seule raison.

Mais je suppose que j'ai tort, et que le présentisme a raison. Pourtant, cette vérité n'est pas une vérité de nécessité. Le quadridimensionnalisme est logiquement possible. Et si le quadridimensionnalisme est possible, alors le voyage dans le temps pourrait bien l'être aussi. Tout au moins, l'existence du passé et du futur donne au voyageur un ail/heure où se rendre. Ainsi, la question qui se pose maintenant est celle-ci : le voyage dans le temps est-il possible dans un univers quadridimensionnel? S'il l'est, il faut résoudre quelques paradoxes bien déroutants.

Le paradoxe du grand-père

On ne peut pas changer l'histoire.

Spock à Kirk, « Spectre du fusil » (PS)

En voici un bien connu des philosophes et physiciens. Si le voyage temporel vers le passé est possible, alors quelqu'un - moi, par exemple - peut revenir dans le temps et infléchir les événements. Ainsi, je peux prendre un revolver et tuer quelqu'un. Je peux tuer un de mes ancêtres - disons mon grand père, en 1920. Le problème, c'est que mon père n'est né qu'en 1929. Donc, si je tuais mon grand-père en 1920, n'est-il pas vrai que je n'aurais jamais vu le jour? Mais il faut bien que j'existe, pour tuer mon grand-père! Dès lors que je ne puis exister et pourtant ne pas exister (contradiction), il doit m'être impossible de revenir dans le passé et de tuer mon grand-père en 1920. Mais si ce voyage est possible, alors je peux très certainement tuer mon grand-père en 1920. Donc le voyage dans le temps (en tout cas, dans le passé) est nécessairement impossible.

Star Trek tire force effets de ce filon. Dans l'ultime épisode de La nouvelle génération, « Les meilleures choses ... », la tentative de Picard de refermer une faille temporelle menace d'affecter le passé en supprimant l'événement premier, celui qui a enclenché l'évolution de la vie sur Terre. Picard doit réparer les dégâts (il y réussit, bien sûr), mais la conséquence est claire: s'il avait échoué, ni Picard ni le reste de la civilisation terrestre et de son histoire n'auraient existé. Une fois encore, cette possibilité contient une contradiction: comment Picard peut-il lui-même provoquer son inexistence? La solution semble être que toute l'intrigue est impossible. Rigolote, mais impossible.

Pourtant, si le postulat qui fonde cette intrigue est impossible (du moins en apparence - j'y reviendrai), cela ne vaut pas pour la plupart des épisodes de Star Trek où intervient le voyage dans le temps. Ainsi, on peut réfuter le paradoxe du grand-père sans nier pour autant toute possibilité de voyage dans le temps, mais c'est heurter les présupposés de presque tous les épisodes de Star Trek où intervient le voyage dans le temps. Tous se fondent sur l'idée de « chronoclasme » (j'emprunte ce terme merveilleux à l'histoire sur le voyage dans le temps intitulée « Chronoclasme », et dont l'auteur est l'écrivain de science-fiction John Wyndham). L'idée est qu'il y a dans l'histoire certains points cruciaux : apporter un quelconque changement à « ce qui s'est passé » produirait un désastreux « effet ricochet » sur l'avenir. Dans « La

citée au bord de l'éternité» (NG), l'Entreprise rencontre un petit coin d'espace cahoteux à proximité d'une planète qui ne figure pas sur les cartes. McCoy s'injecte accidentellement une dose élevée de « cordrazine », qui produit une manie paranoïaque. Bien décidé à s'échapper de l'Entreprise, il se téléporte sur la planète, et une mission se lance à sa poursuite. Sur la surface de cette planète, Kirk, Spock et les autres trouvent le « gardien de l'éternité », une machine dotée d'une fenêtre à travers laquelle on peut voir d'autres époques et d'autres lieux. Pendant que la mission regarde le spectacle, McCoy surgit soudain et bondit par l'entrée jusqu'en 1930. Mauvais coup. Sous les yeux horrifiés du groupe, l'histoire de la Terre change: les nazis gagnent la Seconde Guerre mondiale, le XXIV siècle n'est plus pacifique, il n'y a plus de Fédération unie des planètes, plus d'Entreprise (la mission appelle le vaisseau sans obtenir de réponse). Il s'est produit un chronoclasme, et il revient à Kirk et Spock, qui l'attribuent à McCoy, de réparer les dégâts en revenant à leur tour en 1930. Atterrissant à ce qui semble être New York, ils rencontrent Edith Keeler qui tient un abri pour les sans-logis. Lorsque Spock bricole un ordinateur de fortune pour lire le futur immédiat, ils découvrent qu'Edith Keeler est le personnage clef de ce chronoclasme : elle est censée mourir en 1930 dans un accident de voiture. Apparemment, McCoy va empêcher que cela n'arrive. Dans ce nouvel avenir, elle restera en vie et, en pacifiste ardente, persuadera les dirigeants des États-Unis de ne pas s'engager dans la Seconde Guerre mondiale, à moins que Kirk et Spock ne puissent arrêter McCoy.

L'idée de chronoclasme, si chère aux auteurs de Star Trek, présuppose qu'on puisse revenir dans le temps et agir de manière à changer le passé et ses suites par effet ricochet. Renoncer à l'hypothèse selon laquelle le passé peut être changé, c'est résoudre le paradoxe du grand-père. Faut-il alors le faire? Reconsidérons ce paradoxe. Ce qui devrait être évident, c'est que, puisque moi, le voyageur du temps, j'existe, je ne peux pas réussir à tuer mon grand-père avant que mon père ait été conçu. Cela ne signifie pas que je ne puisse pas revenir dans le passé, mais simplement que je ne puisse accomplir logiquement certaines choses dans le passé. Quoi au juste? On pourrait penser ici que les restrictions que m'impose la logique ne concernent que les aspects m'affectant directement. Mais ce n'est pas vrai. Mon père est né en 1929, et mon grand-père est mort en 1934. Donc, je pourrais remonter le temps jusqu'en 1930 et descendre Pépé quelques années avant la date, mais après la naissance de mon père, ce qui, somme toute, ne remettrait pas en cause mon existence. Mais il y a là un obstacle manifeste : mon grand-père est mort en 1934 ! (La résurrection est une hypothèse logique, mais écartons-la, et supposons que mon grand-père est mort une fois pour toutes). Puisqu'il est impossible que mon grand-père meure en 1930 et en 1934, je ne peux pas tuer mon grand-père en 1930. La logique aura raison de moi aussi bien en 1930 qu'en 1920.

Dans la conception du temps des quadridimensionnalistes comme dans celle des présentistes, il n'y a qu'un seul passé. Par exemple, pourquoi est-il vrai que Kennedy a été assassiné en 1963? Parce que c'est ce qui s'est passé. Pourquoi n'est-il pas vrai que Kennedy a survécu et conduit les États-Unis à la victoire au Viêt Nam? Parce que

ce n'est pas ce qui s'est passé. Ainsi, dans la vision quadridimensionnaliste que nous sommes en train d'étudier, il est logiquement possible de remonter dans le passé, mais il n'est logiquement pas possible de remonter dans le passé et de le changer. Dès lors, une histoire où l'on voit quelqu'un remonter dans le passé et le modifier exploite une intrigue logiquement impossible.

Mais, pour cette raison précise, ne devrions-nous pas reconnaître que le voyage dans le temps est logiquement impossible? Après tout, si vous voyagez dans le passé, vous l'affectez par votre seule présence. Si vous êtes vraiment dans le passé, vous affectez nécessairement le passé, et il vous affecte. Pour prendre le plus primaire des exemples, vous ne sauriez voir, entendre ou toucher quoi que ce soit si le monde alentour ne faisait pas effet sur vos sens, vous permettant d'absorber la lumière, de projeter une ombre, etc. Si prudent que vous soyez en remontant le passé, vous êtes dans le passé et vous l'affectez nécessairement. Comment faire autrement que le changer?

Cette objection repose sur une confusion : une chose est d'affecter le passé, une autre de le changer. Supposez que je saute dans ma machine à remonter le temps et que je revienne en 1920. La vision commune du voyage dans le temps, à laquelle je vous conjure de résister, est la suivante. Avant que je ne monte dans ma machine à remonter le temps, il y a eu un 1920 sans moi. Je ne suis né qu'en 1957, et c'est ma première apparition au monde. Après mon voyage dans la machine, il y a eu un autre 1920, modifié, avec moi dedans. Ce cliché semble manquer de cohérence : pour ma part, je ne vois guère ce que signifient les mots «avant» et «après». Soit une vision plus cohérente de la chose: il n'y a qu'un seul 1920, et je suis dedans. En d'autres termes, si je n'étais pas là en 1920, il est impossible qu'une machine à remonter le temps m'y transporte, et si j'étais là en 1920, il est impossible qu'elle ne m'y transporte pas! Si j'étais là en 1920, alors j'ai affecté le passé, comme tous ceux qui étaient là en 1920. Le chemin qu'a pris le futur a dépendu en partie de ce que j'ai fait en 1920.

Ayant fait nôtre l'idée selon laquelle il n'existe qu'un seul passé, nous pouvons rendre tout son sens à « La cité au bord de l'éternité ». Au dénouement, le chronoclasme est «détourné». En fait, il aurait pu ne pas se produire du tout, puisqu'il ne s'est pas produit. Parce que les nazis ont perdu la Deuxième Guerre mondiale, rien de ce qu'aurait pu faire McCoy n'aurait empêché leur défaite. Mais, en même temps, il n'est pas vrai que McCoy n'aurait pu agir après son retour en 1930. La façon dont les choses se sont passées dépendaient de la présence des voyageurs du temps. Il n'y a eu qu'un 1930, et Kirk, Spock, McCoy en étaient. Kirk et Spock n'ont pas simplement empêché McCoy de sauver Edith Keeler, ils ont été les instruments de ce qui s'est effectivement passé. Au dénouement de l'épisode, Keeler a déjà traversé la rue sans le moindre dommage lorsqu'elle voit Kirk et Spock se battre avec McCoy: elle fait volte face, court vers eux ... et est percutée par une voiture. Ce sont les voyageurs du temps qui l'ont tuée! Néanmoins, il faut user de quelques libertés avec le reste de l'histoire. En somme, l'histoire de la Terre ne change pas - pourquoi la mission a-t-elle donc observé les «changements» sur la planète du Gardien ? Parce que le Gardien (on

se croirait chez Capra !) leur a montré ce qui se serait passé si (contrairement aux faits) Edith Keeler était restée en vie. Étant donné que Kirk et Spock étaient là en 1930, il fallait bien que quelque chose les incite à suivre McCoy à travers la fenêtre : c'est la ruse du Gardien - et leur croyance erronée dans le chronoclasme - qui a monté cette supercherie. S'ils avaient gardé tous leurs esprits, ils ne se seraient jamais laissé avoir par cette ruse.

Si le chronoclasme s'était vraiment produit, et que, par conséquent, ni la Fédération ni l'Entreprise n'avaient existé, comment aurait-on d'emblée pu expliquer leur présence, en uniformes de la Fédération, sur la planète du Gardien ? En général, les histoires de voyage dans le temps requièrent fondamentalement que les voyageurs s'abusent sur leurs propres pouvoirs.

Je devrais signaler que les auteurs de Star Trek, comme souvent, fournissent une explication après coup de leur erreur - ici, le fait qu'un changement apparemment radical de l'histoire ne produise aucun effet sur la mission elle-même. J'ai invoqué une ruse, mais on trouve une explication différente dans « Passé composé 1 » (PE9). Les officiers confirmés de Profondeur Espace 9 sont en orbite autour de la terre, à bord du Défiant. Le commandeur Benjamin Sisko, le Dr Bashir et Jadzia Dax sont téléportés, puis ramenés par accident dans le San Francisco de 2024, époque quelque peu mouvementée. Sisko sait son histoire : il y avait de profondes inégalités sociales, et des foules de sans-logis avaient été rassemblés dans des ghettos murés de la ville, connus sous le nom de « sanctuaires », et dirigés par les troupes gouvernementales. En 2024, les sans-logis se sont emparés d'un sanctuaire, et ont fait des otages. Un des sans logis, Gabriel Bell, a contribué à préserver la vie des otages et contraint les autorités à mettre en place des réformes sociales, engageant les États-Unis sur la voie d'un gouvernement plus éclairé. Parce qu'ils sont sans argent et sans papiers d'identité, Sisko et Bashir se retrouvent bientôt dans un sanctuaire, et sont mêlés à une bagarre avec quelques brutes. Un homme se porte à leur secours, et se fait tuer. Bon sang, mais c'est bien sûr, c'était Gabriel Bell ! Chronoclasme ! Sisko comprend qu'il doit prendre la place de Bell et réussir tout ce qui a fait la gloire de Bell.

De retour à bord du Défiant, placé en orbite autour de la Terre du XXIV siècle, Kira Nerys et Miles O'Brien se sont rendu compte que l'équipe expédiée a été envoyée dans le passé. Mais la Flotte interstellaire de commandement, qui se soucie de toute modification de la ligne temporelle, leur interdit toute expédition de secours. Lorsque Kira et O'Brien perdent soudainement le contact avec les appareils de la Flotte interstellaire - y compris avec les senseurs - comme jadis la mission de « La cité au bord de l'éternité », ils en concluent rapidement que la chronologie a été altérée; d'une manière ou d'une autre, l'histoire a changé, mieux vaut tenter de la faire revenir dans le droit chemin. Kira pose alors une question évidente: Pourquoi n'avons-nous pas été affectés par ces changements ? O'Brien avance ce que j'appellerai l'hypothèse de la protection par particules de chronoton. Apparemment, le mécanisme « horloger » du Défiant a généré des « particules de chronoton », et le passage d'un élément microscopique a cristallisé ces particules, faisant du téléporteur une machine à remonter le temps. L'hypothèse d'O'Brien, c'est que cet

élément a aussi « créé une bulle subspatiale autour du vaisseau - l'isolant ainsi des variations temporelles. »

C'est précisément l'explication du genre *deus ex machina* qu'on devrait éviter si l'on veut que les histoires de voyages dans le temps aient un sens. L'hypothèse de la protection par particules de chronoton suggère non seulement que les chronoclasmes sont possibles et qu'il existe plusieurs chronologies, mais aussi qu'on peut relever d'une chronologie sans pour autant dépendre d'elle. Après tout, la « bulle subspatiale » n'empêche nullement les informations ordinaires en provenance du « nouveau » monde de parvenir au vaisseau (c'est ainsi que Kira et O'Brien découvrent que la situation a changé). Par conséquent, les changements intervenus dans le « nouveau » monde ont un effet sur ce qui se passe à l'intérieur de la bulle. La bulle est, pour ainsi dire, semi-perméable aux causes, selon un principe qui m'échappe. Je ne suis pas preneur, et je vous déconseille de gober ce truc! Voici mon interprétation alternative: pour porter secours à Sisko, Bashir et Jadzia Dax, O'Brien et compagnie doivent les retrouver en scrutant le temps. Le seul *deus ex machina* là-dedans, c'est une coïncidence: la perte totale de contact avec la Flotte interstellaire coïncide avec une lecture défectueuse des senseurs. L'équipage du *Défiant* suppose, à tort, que le passé a été modifié, ignore les ordres de la Flotte interstellaire, tente une recherche, et conclut avec succès son opération de secours.

Un autre aspect de la « modification du passé » (si l'on admet momentanément cette possibilité) n'est jamais résolu de manière satisfaisante dans *Star Trek*. L'attitude de la Flotte interstellaire à l'égard de la « contamination chronologique » - connue aussi sous le nom de politique de déplacement temporel de la flotte interstellaire - est clairement en rapport avec la règle d'or, une sorte d'analogie, sur le mode N'interférez pas avec l'histoire. Dans « *De temps à autre* » (V), le capitaine Janeway invoque cette politique. Elle appartient à une équipe partie examiner les restes d'une planète. Les membres de l'expédition découvrent qu'une énorme explosion a détruit la civilisation tout entière quelques heures seulement avant leur arrivée. De plus, le souffle de l'explosion a produit des failles temporelles locales, et l'une d'elles envoie Janeway et le lieutenant Tom Paris un jour en arrière. Ils se retrouvent ainsi dans une désagréable situation : ils savent que tous les habitants du lieu sont condamnés, et Paris incline à le leur faire savoir, dans un effort pour leur sauver la vie. Janeway n'est pas d'accord, invoquant tout particulièrement la « non-interférence dans l'évolution naturelle » stipulée par la règle d'or. Paris se demande avec une hésitation qui n'est pas injustifiée si une annihilation totale représente un événement « naturel » dans l'évolution - mais cela n'entre pas dans notre propos. L'extension exacte de cette politique de déplacement temporel n'est pas très claire (du moins à mes yeux). Supposez qu'une interférence avec l'histoire permette un avenir meilleur - un chronobonus. Après tout, aucune histoire, aucune chronologie, aucun avenir n'a de privilèges inhérents, et il n'est donc pas intrinsèquement mauvais de changer les choses. Peut-être cette politique vise-t-elle à empêcher les équipages de la Flotte d'empirer involontairement la situation, mais si les choses sont (furent, seront) déjà au pire, ne vaut-il pas la peine de tenter le coup, maintenant, avant et

après?

Qu'aurais-je pu faire?

Je n'en ai pas fini avec le paradoxe du grand-père. Reste un problème: j'ai dit que je ne saurais tuer mon grand-père en 1920 ou en 1930, mais, en un sens tout à fait viable, je le peux. Je peux certainement m'entraîner pour devenir un tireur d'élite, entretenir mon arme, rester à l'affût de Pépé un jour sans vent, et appuyer sur la détente à bout portant. Si tout le monde peut tuer tout le monde, je peux aussi tuer mon grand-père ! Il faudrait un miracle pour le sauver de la balle mortelle. Mais je ne veux pas invoquer de miracles pour cautionner le voyage dans le temps, et suis donc d'accord pour dire que je peux tuer mon grand-père. Mais là encore, je suis confronté à la contradiction, puisque j'affirme à la fois que je peux tuer mon grand-père et que je ne peux pas le tuer.

Heureusement, il y a une issue de secours : les mots « peut » et « pourrais » sont ambigus. Vous m'accorderez sans doute que Michael Jordan peut marquer un panier à trois points dans les dernières secondes et remporter le match. Si quelqu'un peut le faire, c'est bien Jordan. C'est pourquoi les Bulls lui passent la balle aux limites de la zone. Voilà le sens général de «pouvoir» qui relève des capacités de chacun. Mais Jordan rate parfois, et il pourrait dire alors : «J'ai essayé de marquer, et je n'ai pas pu.» Voilà le sens particulier de «pouvoir», qui renvoie simplement au succès ou à l'échec particuliers. Jordan ne se contredirait pas s'il disait: « Je pouvais le faire, mais je n'ai pas pu.» De même, en ce qui concerne le voyage dans le temps, je suis d'accord pour dire que je pourrais tuer mon grand-père dans le sens général de «pouvoir», mais non dans le sens particulier. Il n'y a pas ici de contradiction.

Un seul monde, ou plusieurs?

Vous aurez peut-être entendu parler de l'interprétation de la mécanique quantique dite des « mondes pluriels». En voici l'idée, sommairement résumée. Au niveau quantique, l'avenir est authentiquement ouvert (comme le suggère le présentisme), mais il n'est pas voué à emprunter telle voie particulière. Disons plutôt que chaque embranchement possible de l'avenir devient un embranchement réel. Chaque fois qu'un «choix» quantique est fait, il produit des univers parallèles, mais différents. Comme le dit Data dans «Parallèles» (NG), au niveau quantique, « tout ce qui peut arriver arrive ». Si ces choix quantiques autorisent des alternatives authentiques au niveau macroscopique, peut-être avons-nous tort de rejeter l'idée selon laquelle il y a plus d'un passé. Dans ce cas, il y aurait nombre de passés.

C'est une idée fascinante, mais qui n'importe guère pour notre discussion présente. Tout d'abord, l'aspect le plus important du quadridimensionnalisme, pour notre propos, n'est pas l'affirmation qu'il y a une dimension temporelle unique, mais plutôt que le temps ressemble fondamentalement à l'espace. C'est la spatialisation du temps qui importe. J'entends soutenir que le voyage dans le temps est logiquement

possible en recourant à la notion la plus plausible du temps (à mon idée, mais je ne serais pas étonné d'apprendre qu'il y a plus d'une dimension temporelle et cela ne changerait rien à l'impossibilité du chronoclasme). Supposons qu'il y ait de fait « des mondes pluriels ». Néanmoins, le monde qui nous inclut, vous, moi et ce livre, est un monde unique, soumis à une chronologie unique. Dans ce monde, le passé ne s'est écoulé que d'une seule manière. C'est pourquoi le meurtre du grand-père n'est pas un événement de ce monde - il est impossible, puisque j'existe. Et vous remarquerez un autre fait - il est très difficile de donner un sens à l'identité personnelle à travers le temps dans le cadre des « mondes pluriels ». Par exemple, comment justifier l'existence, dans tous ces différents mondes, d'individus tous nommés « Richard Hanley » ? Ces individus ne sont pas identiques les uns aux autres, ils sont plutôt équivalents. Même dans un autre monde, on ne saurait dire que ce soit moi qui agisse. Le Richard Hanley de là-bas est un tout autre type ! Cette différence d'identité est reconnue implicitement dans Star Trek lorsque des chronologies alternatives se croisent d'une façon ou d'une autre (voir « Miroir, mon beau miroir » [PS], et ses suites dans « Traversée », « À travers le miroir » et « Le miroir brisé »). On ne remplace pas de l'intérieur son équivalent de l'univers en miroir : on est miraculeusement transplanté à sa place (comme dans « Miroir, mon beau miroir »), ou on se retrouve là avec lui (comme dans les suites).

Supposons que ma machine à remonter le temps soit une « machine à faire des bonds d'un univers à l'autre », qui me ramène à un univers alternatif des années vingt. Ne pourrais-je alors tuer mon grand-père - c'est-à-dire son équivalent - dans ce monde, puisque mon existence n'est pas logiquement menacée par sa mort ? Apparemment, puisque mon grand-père n'a pas de descendant dans l'univers que je visite. Mais encore une fois, il ne s'agit pas de changer le passé. Je n'ai pas modifié le passé dans mon propre monde, et (toujours en suivant notre raisonnement) puisque j'étais dans l'autre monde pour empêcher mon grand-père de procréer, je n'ai rien changé là-bas non plus. Il n'y a pas de contradiction là non plus, puisqu'il n'y a pas de monde où il soit vrai que j'existe sans exister, ou que je tue mon grand-père sans le tuer.

«L'Entreprise d'hier»

Je suis certain que vous considérez encore l'idée de changer le passé comme susceptible d'être maintenue : elle est trop enracinée dans notre imaginaire de science-fiction. Admettons qu'il y ait des chronologies alternatives - sans chercher à les expliquer - et regardons les choses du point de vue non pas de ceux qui voyagent dans le temps ou sautent d'un univers à l'autre, mais de tous les autres. Prenons l'épisode de la troisième saison de La nouvelle génération, intitulé « L'Entreprise d'hier », qui a généré une petite industrie d'interprétation chez les fans. Picard et compagnie se penchent sur une nouvelle anomalie, mais celle-ci s'avère une faille temporelle, et un nouveau vaisseau - qui semble appartenir à la Fédération - apparaît au seuil de la faille. Soudain, tout change à bord de l'Entreprise-D, Le Klingon Worf

n'est plus un membre de l'équipage, le lieutenant Tasha Yar - tuée sur Vagra II par une créature du nom d'Armus, dans l'épisode de la première saison intitulé « La peau du mal » - est de retour, et tout le monde se rappelle les vingt-deux dernières années comme étant dominées par une guerre longue et sanglante avec l'Empire klingon, que la Fédération est en voie de perdre.

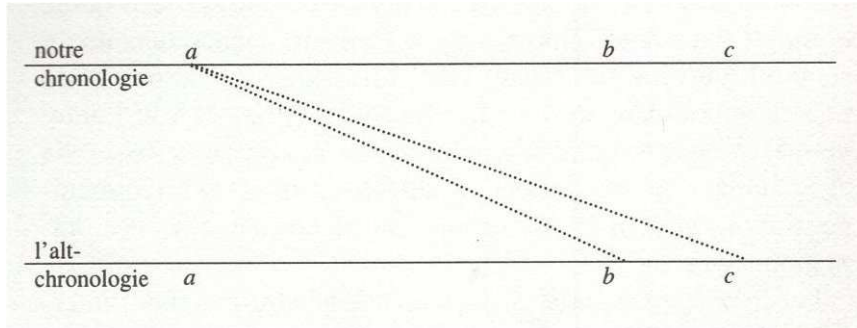
Le vaisseau qui a émergé de la faille temporelle est l'Entreprise-C, qui a fait un bond de vingt-deux ans dans l'avenir. Guinan, à bord de l'Entreprise-D, a l'intuition que quelque chose cloche - par exemple, elle sait que le lieutenant Yar « n'a pas sa place ici » et le dit à Picard. Ils finissent par comprendre ce qui s'est produit. En quittant le passé au beau milieu de la bataille, l'Entreprise-C n'a pu protéger un avant-poste klingon d'un assaut romulien, et chronoclasme ! La bataille n'a pas mené à une alliance entre l'Empire klingon reconnaissant et la Fédération : au contraire, l'incident a provoqué une guerre à grande échelle. Il faut y remédier, et l'équipage de l'Entreprise-C est prêt à le faire : ils retourneront à leur propre époque pour défendre l'avant-poste klingon, même si cela signifie une mort certaine. Tasha Yar, à qui Guinan a dit que sa mort (sur Vagra II) n'avait aucun sens, se porte volontaire pour les accompagner. Entre-temps, les Klingons débarquent et commencent à tirer sur les deux vaisseaux. L'Entreprise-D se place dans la ligne de tir - et est apparemment détruite - pour protéger l'Entreprise-C qui disparaît dans la faille temporelle. Soudain, nous voyons les membres de l'équipage à nouveau dans leurs rôles familiers à bord de l'Entreprise-D, comme si rien ne s'était produit.

Cet épisode ne saurait être tenu pour une simple histoire de voyage prenant place dans une chronologie unique, parce que ces événements chronoclastiques se produisent véritablement. L'Entreprise-D, altérée, relève du chronoclasme, et les gens qui se trouvent à son bord contribuent au « retour à la normale » des événements. S'ils provoquent des causes, ils doivent exister. Mais cela suscite immédiatement des contradictions : par exemple, il est vrai que Tasha Yar a été tuée sur Vagra II et n'a pas été tuée sur Vagra II. J'en conclus que « l'Entreprise d'hier » ne prend pas place dans une chronologie unique, mais renvoie plutôt au bond d'un univers à l'autre.

Voici ce qui se produit véritablement. Il y a deux chronologies - deux mondes ou univers. L'un est l'univers qui nous est devenu familier à travers les épisodes précédents de La nouvelle génération. Bien qu'aucune chronologie ne jouisse de privilèges inhérents, je supposerai qu'il s'agit de notre monde et de notre temps (c'est à coup sûr ce que Star Trek nous incite à penser). Notre monde contient l'Entreprise-D et son équipage; le monde alternatif contient leurs équivalents. Je renverrai aux habitants du monde alternatif comme alt-Entreprise-D, alt-Entreprise-C, alt-Picard, alt-Yar, etc. « L'Entreprise d'hier » nous offre un aperçu de l'ait-chronologie. Au début de l'épisode, puis à sa fin, nous voyons les choses du point de vue des - de notre chronologie, mais pendant le reste de l'épisode - adoptons celui des gens de l'alt-chronologie.

Voici ce qui se produit dans l'ait-chronologie. L'alt-Entreprise-C et son équipage échouent à sauver l'avant-poste gon, provoquant une guerre de vingt-deux ans entre l'Empire gon et la Fédération, que celle-ci est en voie de perdre. Une faille temporelle

relie les deux chronologies. (Voyez le diagramme ci-dessous, qui montre le cours parallèle des deux chronologies, avec vingt-deux ans entre le temps a et le temps b.) Elle fait intersection entre notre chronologie, au temps et au lieu où se passe l'affaire de l'avant-poste klingon (temps a) ainsi qu'à l'ait-chronologie, au temps et au lieu de l'alt-Entreprise-D (entre b etc). L'Entreprise-C traverse la faille et arrive dans le temps b.



À bord de l'alt-Entreprise-D, l'alt-Guinan -qui doit avoir une forme de perception extrasensorielle touchant les sauts d'univers - convainc tout le monde qu'un chronoclasme s'est produit. C'est une erreur, bien sûr. Mais cela suggère à l'équipage de l'Entreprise-Cet à alt-Yar de regagner notre chronologie, et à l'équipage de l'Entreprise-D de se sacrifier, hommes et vaisseau, pour permettre ce retour (au temps c). L'alt-Entreprise-D est véritablement détruite, et l'ait-Picard et l'équipage avec.

Voici ce qui se produit dans notre chronologie. L'Entreprise-C sauve l'avant-poste klingon, et alt-Yar joue un rôle crucial dans sa défense. Sans ses connaissances tactiques avancées - de vingt-deux ans - l'Entreprise-C (condamnée de toute façon) n'aurait jamais duré assez pour sauver l'avant-poste. Le vaisseau aurait connu un sort semblable à celui de l'alt-Entreprise-C. Cet aperçu de l'ait-chronologie nous montre ce qui se serait produit dans notre chronologie s'il n'y avait pas eu de faille temporelle. De plus, l'action des gens de l'ait-chronologie a épargné pareil sort à leurs équivalents dans notre chronologie.

Cette histoire est parfaitement cohérente, et fournirait une excellente interprétation de « l'Entreprise d'hier» si nous pouvions justifier quelques détails. Au début de l'épisode, l'Entreprise-D croise une faille temporelle et voit un vaisseau émerger. Ou il s'agit là d'une faille temporelle entièrement différente de celle traversée par l'Entreprise-C (et c'est un vaisseau différent), ou la faille temporelle relie par ailleurs notre chronologie à l'ait-chronologie, au point b: l'Entreprise-C quitte a, surgit dans notre chronologie à b, repart séance tenante, et saute un univers, retrouvant l'ait-chronologie b. Ce qui laisse une extrémité dans le flou: lorsque l'Entreprise-D voit l'autre vaisseau surgir de la faille temporelle, nous voyons la passerelle et l'équipage se transformer en alt-Entreprise-D et en ait-équipage. Nous voyons Picard devenir ait-Picard - une licence dramatique, peut-être?

Un épisode plus tardif de La nouvelle génération était l'interprétation de « l'Entreprise d'hier» fondée sur le saut d'univers. Dans « Rédemption I & II», on voit une femme officier romulienne nommée Sela, qui ressemble de façon étonnante à Tasha · Yar. Elle affirme être sa fille. Mais elle a vingt-trois ans, et Picard calcule que

Tasha aurait été enfant à l'époque où Sela est née. Comment expliquer cela? L'histoire que raconte Sela à Tasha, c'est que, lorsque l'Entreprise-C fut détruite, une partie du personnel de la Fédération fut pris en otage par les Romuliens, dont Tasha Yar. En retour pour la sécurité des autres captifs, Tasha accepta de devenir la compagne d'un général romulien, et donna naissance à Sela. Cette dernière avait quatre ans lorsque Yar essaya de s'enfuir et fut exécutée.

Sela a d'autres nouvelles pour Picard. Sa mère lui avait dit que c'était lui qui l'avait renvoyée dans le temps. Picard rechigne à croire son histoire, ce qui est compréhensible parce qu'il ne se rappelle pas avoir agi de la sorte. Et de fait, ce n'était pas lui, mais alt-Picard. L'histoire de Sela sonne juste, précisément parce que, dans notre chronologie, il n'y a que l'équipage de l'Entreprise-D qui était au courant des événements dépeints dans «L'Entreprise d'hier». La mère de Sela, bien sûr, n'était pas Yar mais alt-Yar. Yar n'aurait pas pu être à la fois exécutée par les Romuliens et tuée sur Vagra II par Armus. L'existence d'alt-Yar dans notre chronologie tranche la question: la chronologie alternative existe bel et bien.

Crise d'identité

J'ai affirmé que nous ne pouvions pas avoir recours à la perspective des « mondes pluriels » pour donner sens au postulat selon lequel les voyages dans le temps de Star Trek modifient le passé, en partie à cause des problèmes que cela pose pour l'identité. Toutefois, il y a des problèmes d'identité dans la perspective quadridimensionnaliste, si nous admettons le voyage dans le passé. Par exemple, dans « Le temps au carré » (NG), Picard se rencontre lui-même ! Ou plutôt, le « présent » Picard rencontre son moi du futur, un moi qui arrive par navette spatiale (depuis un futur de six heures). S'il est possible de remonter son propre passé et de se heurter à son moi passé, c'est un peu ce qui arrive à Picard. Bien sûr tout l'équipage de l'Entreprise-D, notamment Picard, a du mal à accepter les preuves de l'intrus, même si Troi leur assure qu'il s'agit bien de Picard. Voyez la réaction de Worf lorsqu'il découvre que la navette où l'« autre Picard » a débarqué est bien la même (NCC-17010, USS Entreprise-D, Navette 05) que celle qui la joute dans le port à navettes. « Comment est-ce possible ! », s'exclame-t-il. Deux navettes 05 ? Deux Picard ?

Il est certain que Picard ne saurait être à deux endroits à la fois, et cela signifie qu'il est impossible qu'il remonte le temps et se rencontre lui-même. Le voyage dans le temps (du moins dans le passé) doit être impossible. Mais en fait, il n'est pas impossible pour que Picard voyage dans le temps. Rappelez-vous que, suivant le quadridimensionnalisme, Picard est un objet quadridimensionnel consistant en un agrégat de parties temporelles, dont chacune est une tranche de personne ou tranche temporelle. Sans le voyage dans le temps, les diverses tranches de personne ne se rencontrent jamais : chacune se maintient dans un intervalle de temps donné. Mais avec le voyage dans le temps, il est possible que deux tranches ou plus d'une personne soient en divers lieux au même moment, voire se rencontrent. Cela ne signifie pas que la même tranche de personne soit dans plusieurs lieux en même temps - ce serait

impossible - mais plutôt que la même personne (Picard) est en partie dans deux différents lieux en même temps, parce que ses parties temporelles occupent des endroits distincts. (De même, lorsqu'elle pose un pied de chaque côté de la frontière, une personne peut habiter deux États en même temps, parce que ses parties spatiales occupent chacune un État.) Une partie temporelle est le présent Picard (soit Picard 1), et l'autre est le futur Picard (soit Picard 2). Picard 1 est plus âgé de six heures que Picard 2 dans le temps personnel de Picard. Grâce au voyage dans le passé, tous deux peuvent exister dans le même temps externe, mais à des lieux différents (même lorsqu'ils sont dans la même pièce, ils sont à des endroits différents de la pièce).

Malheureusement, il n'est pas possible de rendre compte de tout ce qui se produit dans « Le temps au carré » - du moins pas si on postule qu'il existe une seule chronologie, que Picard 2 existe et qu'il voyage dans le temps. Un problème qui se pose, c'est que le passé est modifié pour éviter un désastre (chronobonus !). Les cadrans de la navette (qui informent l'équipage qu'elle date de six heures dans l'avenir) indiquent que l'Entreprise-D a été détruite par « un vortex d'énergie intelligente », Picard 1 finit par comprendre que Picard 2 a commis une erreur en essayant d'éviter le vortex; il change de programme, vole droit sur le vortex et sauve la situation. On pourrait essayer de contourner le problème en disant que c'est l'avenir qui est modifié, mais rappelez-vous que dans le quadridimensionnalisme il n'y a pas de différence réelle entre le passé et l'avenir, et que c'est tout aussi impossible. (Regardez de nouveau la citation de Riker en exergue.) Voici une autre façon d'y réfléchir. Rappelez-vous que Picard 1 et Picard 2 sont des parties temporelles différentes de la même personne. Tout ce que fait l'un d'eux est aussi fait par le seul Picard, et la logique exige que Picard ne peut à la fois faire une chose et ne pas la faire. Il est impossible qu'il évite le vortex tout en fonçant droit dessus.

Cette histoire pose un autre problème. Pour se tirer d'affaire et diriger l'Entreprise-D dans le vortex, Picard 1 découvre qu'il est nécessaire de tuer Picard 2 ! Bien que ce ne soit pas impossible en soi (alors que, pour les mêmes raisons qu'avec grand père, il est impossible que Picard 2 tue Picard 1). Vous voyez bien, j'espère, que c'est imprévoyant. En tuant Picard 2, une partie temporelle plus tardive de la même personne, Picard s'assure qu'il n'y aura plus d'éléments de continuité personnels de lui-même (si l'on laisse de côté les possibilités de fission évoquées dans le chapitre 5). Il commet donc un suicide retardé. Toutefois, l'histoire continue, avec Picard; un autre désastre plus personnel est donc évité (autre chronobonus). De plus, rien de ce que connaît Picard 2 n'arrive à Picard 1 ni à ses éléments de continuité. Pire encore, une fois que l'Entreprise-D a volé dans le vortex, obtenant le chronobonus, le cadavre de Picard et la tranche contemporaine de la Navette 05 disparaissent tout simplement! Vont-ils ailleurs (ou aillheure), ou cessent-ils d'exister? Quelle est la cause naturelle de leur disparition? (Un semblable «gommage» apparaît dans « Hier est demain» (PS), où c'est l'Entreprise qui se fait «gommer» depuis les cieux du Nebraska.)

Vers la fin du « Temps au carré», Riker suggère que, peut-être, rien de ce qu'ils ont éprouvé n'était réel, et, si l'histoire doit être cohérente, je pense que nous

devons agréer ce jugement. Parmi les pouvoirs extraordinaires du vortex, il y a sa capacité de produire une hallucination de masse; peut-être le vortex souhaitait-il épargner l'Entreprise-D et a-t-il concocté un plan pour persuader Picard d'abandonner ses instincts habituels et d'aller droit devant.

Boucles de causalité

Nous avons l'habitude de considérer que la cause et l'effet sont nécessairement asymétriques en matière de temps - à savoir qu'une cause doit toujours précéder son effet (ou du moins qu'un effet ne doit jamais précéder une cause). Un exemple de causalité normale, c'est lorsque quelqu'un attache mes lacets ensemble, provoquant par la suite ma chute et une bosse sur ma tête. Une causalité inversée, où un effet précède sa cause, est certainement contre-intuitive, mais, si le voyage dans le passé est possible, la causalité inversée l'est aussi. De fait, elle paraît inévitable, étant donné le voyage dans le temps. Supposons que, tandis que je suis assis dans la machine à l'an 2000, et que je m'apprête à remonter à l'an 1920, quelqu'un noue mes lacets. J'arrive en 1920, tente de sortir de la machine et tombe tête la première sur le sol. Un événement de 2000 cause donc un événement de 1920, et c'est là un cas de causalité inversée. De même dans « La cité au bord de l'éternité », le comportement maniaque de McCoy est dû à une injection accidentelle de « cordrazine » qui a lieu au XXIV siècle, quoiqu'une partie du comportement de McCoy ait lieu en 1930.

La causalité inversée n'est pas moins plausible que le voyage dans le passé lui-même, et nous ne devrions pas rejeter la possibilité du second en invoquant la première. Mais cette possibilité engendre un second paradoxe, puisqu'elle semble permettre des boucles de causalité. La plus simple de ces boucles implique deux événements, A et B, où A cause B et B cause A. Il y a diverses occasions dans Star Trek où ces boucles simples se produisent. Par exemple, dans « De temps à autre », on trouve la séquence apparente d'événements suivante. Après que Janeway et Paris, à la surface de la planète dévastée, ont été happées dans la fracture temporelle qui leur fait remonter le temps d'un jour, leur disparition incite les autres membres d'équipage à monter une équipe de sauvetage. C'est cela qui provoque l'explosion qui détruit la civilisation planétaire ; et cette explosion crée aussi la fracture temporelle qui piège Janeway et Paris et les renvoie dans le passé. Cette séquence serait un excellent exemple d'une boucle de causalité si les auteurs s'en étaient tenus là, mais toute la séquence est « gommée » à la fin de l'épisode, lorsque l'explosion est empêchée et que l'équipage remporte un chronobonus. On trouve une véritable boucle de causalité dans un autre épisode du Voyageur intitulé « Parallax » (voir la section à venir sur les « boucles en boucle »), et quelques autres dans l'histoire de voyage dans le temps la plus cohérente de tout le canon de Star Trek: « La flèche du temps 1 & II ».

En bref: au XIVe siècle, les scientifiques découvrent certains objets vieux de cinq cents ans, dont un revolver et une montre appartenant à Samuel Clemens (alias

Mark Twain), dans une caverne située un kilomètre en dessous de San Francisco. Le plus surprenant de tous ces objets - et il surprend surtout Data - c'est la tête décapitée de Data! Il y a des perturbations temporelles dans cette caverne qui semblent avoir leur source dans le système de Devidia. L'Entreprise-D se rend à Devidia II, où Data se fait téléporter avec le groupe de mission. Ils découvrent une faille temporelle, mais celle-ci est trop dangereuse pour que des humains la pénètrent, et Data est donc envoyé, et arrive dans le San Francisco de 1893. Parmi les habitants se trouvent Samuel Clemens et Guinan, la barmaid extraterrestre, dotée d'une grande longévité, à bord de l'Entreprise-D, Clemens commence à suivre Data, par soupçon et par curiosité. Plusieurs des membres d'équipage traversent à leur tour la faille temporelle, dont Picard, à la poursuite de deux Devidiens, d'affreux voyageurs de temps extraterrestres qui se nourrissent des habitants de la cité tels des vampires, à cela près qu'ils en sucent «l'énergie neuronale». Tout le monde se retrouve dans la caverne où Data se bat avec l'un des extraterrestres, et il se produit une explosion qui décapite Data, plongeant son corps dans la faille temporelle et le renvoyant au XXIV siècle. Picard et Guinan restent dans la caverne, mais Clemens saute dans la faille avant qu'elle ne se referme. De l'autre côté, Clemens et le corps de Data sont ramenés à bord de l'Entreprise-D où LaForge rattache la tête de Data (celle qui gisait depuis cinq cents ans). À son réveil, Data y trouve un message tapé par Picard en 1893, en code binaire (assez fastidieux, mais passons). Picard a appris d'un des extraterrestres que, si l'Entreprise-D détruit l'opération des extraterrestres sur Devidia II avec ses torpilles à photon, il causera - via la faille temporelle qui relie Devidia II à San Francisco de- 1893 - la destruction de la Terre (chronoclasme !) et cet avertissement sauve la mise. Finalement, l'unique façon qu'a Picard de rejoindre l'avenir, c'est d'utiliser une machine à voyager dans le temps conçue pour une personne, qu'il vole aux Devidiens. Mais Clemens l'a rapportée avec lui au XXIVe siècle! Il l'utilise pour retourner en 1893, et Picard la reprend pour regagner le XXIV siècle.

L'histoire de la montre de Clemens relève d'une boucle de causalité. Lorsque celui-ci rejoint l'Entreprise-D au XXIV siècle, il voit sa montre sur le bureau de LaForge, la prend et la remet dans sa poche. Lorsqu'il regagne la caverne en 1893, il pose sa montre à terre pour aider Guinan, qui est blessée. Elle est ramenée, et Clemens reprend sa montre, puis se ravise et la laisse, sachant qu'elle y sera trouvée cinq cents ans plus tard. (J'aurai un commentaire à faire sur cette décision.) Le fait qu'il ait laissé la montre dans la caverne est (en partie) ce qui fait qu'on l'y trouve cinq cents ans plus tard. Sa découverte est (en partie) ce qui fait qu'elle est ramassée par Clemens au XXIV siècle. Et lorsqu'il reprend sa montre, c'est (en partie) ce qui fait qu'il la laisse derrière lui cinq cents ans plus tard, puisque, s'il ne l'avait pas reprise, elle n'aurait pas été dans sa poche en 1893. C'est cela, une boucle de causalité.

Une autre boucle de causalité concerne Picard et Guinan. À bord de l'Entreprise-D, qui fait route vers Devidia II, Guinan assure Picard que, contrairement à la politique de la Flotte, il doit se joindre au groupe de mission. Picard

est étonné, et la conversation se poursuit ainsi :

Guinan : « Tu te rappelles la première fois où on s'est rencontrés ? »

Picard : « Bien sûr. »

Guinan : « Ce n'est pas si sûr. Parce que, si tu ne participes pas à cette mission, nous ne nous rencontrerons jamais. »

Dans le temps externe, Picard et Guinan se rencontrent pour la première fois en 1893. Cette rencontre est (en partie) ce qui fait que Guinan persuade Picard de se joindre au groupe de mission. Et c'est (en partie) ce qui fait qu'ils se rencontrent en 1893.

D'après moi, ces boucles de causalité n'ont rien de très énigmatique. Mais une nouvelle énigme surgit si nous admettons que le voyage dans le temps est possible, car alors certaines boucles de causalité assez étranges sont possibles. Elles manifestent une fermeture explicative ; c'est pourquoi je les nommerai des boucles fermées. Dans « Demain est hier » (PS), l'Entreprise, pour échapper à l'attraction d'un trou noir, est rejetée juste à temps dans les cieux du Nebraska, dans les années soixante. Grièvement endommagé, le vaisseau est détecté par un radar et pourchassé par un jet de l'armée de l'air. Comme il ne veut pas détruire l'avion, Kirk ordonne d'utiliser un faisceau tracteur, mais celui-ci écrase le fragile avion et ils sont obligés de téléporter le pilote, le capitaine John Christopher, à bord de l'Entreprise. Tout d'abord, Kirk pense qu'il vaut mieux empêcher Christopher de revenir sur terre: leurs archives montrent qu'il n'a apporté aucune contribution importante à l'histoire, et l'on craint qu'il ne trouble la chronologie avec sa connaissance du futur. Mais lorsqu'ils découvrent que Christopher a un fils qui contribuera au programme terrestre d'exploration de l'espace, on décide finalement qu'il faut le ramener. Spock dit à Christopher que son fils aura pour nom Sean Jeffrey Christopher. En supposant que Christopher et sa femme n'eussent pas déjà choisi ce nom pour leur premier fils, il y a ici une boucle de causalité : le fait de devoir baptiser l'enfant Sean Jeffrey provoque l'inscription de ce nom dans les archives, ce qui permet à Spock d'apprendre le nom, ce qui l'amène à le révéler à Christopher, ce qui fait que Christopher le donne au bébé. Ce qui est étonnant avec cette boucle-là, c'est que personne n'invente le nom « Sean Jeffrey ». D'où vient-il donc ? Bien que nous puissions expliquer entièrement chaque événement de la boucle en renvoyant à d'autres événements de ladite boucle, elle-même ne s'explique pas.

Mais peut-être l'apparition inexplicée d'un nom ne vous perturbe-t-elle pas. Mon second exemple vient de Star Trek IV: Retour à la maison. La terre du XXIV siècle est sur le point d'être détruite par un objet cylindrique massif, et Spock devine pourquoi. Une lointaine intelligence a communiqué avec des baleines à bosse, mais elles ont disparu à force d'être chassées, et le cylindre a été envoyé pour les rechercher. (Ce qui n'explique pas pourquoi il devrait tout détruire.) Kirk et son équipage remontent au San Francisco du XX- siècle pour trouver un couple de baleines à bosse et les ramener au XX- siècle. Mais pour cela ils doivent modifier leur

vaisseau- un Rapace klingon - pour le rendre à même de charrier un tas de plancton, et la technologie du XXe siècle ne suffira pas. Scotty et McCoy (déguisés adéquatement) visitent une usine de plexiglas, et Scotty fournit au gérant la formule d'un aluminium transparent très léger, contre une partie de la production. Lorsque McCoy exprime ses doutes sur cette apparente perturbation chronologique, Scotty demande, en toute raison: « Comment savez-vous qu'il ne l'a pas inventée? »

Admettons la vérité probable: les archives historiques disent que le gérant d'usine est l'inventeur de l'aluminium transparent. Une fois de plus, nous avons là une boucle de causalité, et ce qui est étrange, c'est que personne n'a vraiment inventé l'aluminium transparent. Scotty tenait la formule de ses connaissances en matière de technique du XXIV siècle ; il l'a passée au gérant ; et c'est pourquoi cette matière relève en fin de compte de la technologie du XXIV siècle. Bien que chaque événement de la boucle soit expliqué, la boucle elle-même reste inexplicquée, encore qu'elle donne naissance à une technologie importante. Toutefois, les philosophes ont l'habitude de penser que certaines choses importantes n'ont pas d'explication apparente. Il se pourrait bien qu'il n'y ait pas d'explication apparente pour l'existence de toutes les choses. Peut-être l'univers a-t-il commencé sur un Big Bang, mais celui-ci reste lui-même inexplicqué. Les explications doivent bien finir quelque part ! (Même s'il y avait une chaîne d'explications infinie, il peut n'y avoir aucune explication du fait qu'il y ait une chaîne infinie d'explications.) L'inattendu est une chose, l'impossibilité une tout autre chose, et, s'il existe des boucles de causalité fermées, il faut nous attendre à de l'inattendu.

Boucles en boucle

Les boucles que j'ai décrites jusqu'ici s'intègrent exactement dans le tableau métaphysique du quadridimensionnalisme, et ce parce que chaque événement ne se produit qu'une fois. (Dans « La flèche du temps », Samuel Clemens reprend sa montre dans le laboratoire de LaForge et ne la met dans sa poche qu'une seule et unique fois.) Maintenant, examinez la conversation suivante, tirée du « Temps au carré » :

Worf: « Il y a la théorie de Mœbius : une torsion dans la conception de l'espace, qui fait du temps une boucle dont on ne peut s'échapper. »

LaForge : « À ce point, tout ce qui advient adviendra de nouveau. L'Entreprise sera détruite, l'autre Picard nous sera renvoyé, et nous referons tout depuis le début. Cela m'apparaît comme l'enfer. »

Bien que l'équipage évite, pour cette fois, la boucle en question, il est clair qu'ils sont pris dans une autre dans « Cause et effet », où il revivent la même séquence d'événements quatre fois de suite. Les trois premières s'achèvent sur une collision de l'Entreprise-D avec un autre vaisseau, et sur sa destruction, mais la quatrième fois ils ont compris la leçon et échappent à la fois à la boucle et à la destruction. Dans La physique de Star Trek, Lawrence Krauss mentionne cet épisode avec une approbation

apparente, puisqu'elle implique une « courbure temporelle fermée » ; mais, sans être physicien, je dirai qu'on ne peut s'en tenir là. C'est vrai, l'Entreprise-D peut se rendre à une destination et revenir au même point spatial et temporel, mais cela relèverait du voyage dans le temps ; ce faisant, le vaisseau vieillirait, et, au retour, nous aurions une tranche temporelle différente du vaisseau et de ses habitants. À chaque fois, l'équipage aurait fait une série d'expériences différentes. De plus, si le vaisseau regagne la même région en matière d'espace et de temps, il y a au moins deux tranches temporelles du vaisseau présentes à ce moment. Ce n'est pas qu'il y ait une tranche temporelle, puis la fois suivante, deux, etc. Lorsque le vaisseau rencontre la boucle, il devrait y avoir trois autres tranches de l'Entreprise-D dans la boucle, et pas seulement le Bozeman, le vaisseau avec lequel l'Entreprise-D entre en collision. Mais ce scénario révisé est de toute façon impossible, puisque l'Entreprise-D est détruite au cours de la première occurrence. C'est pourquoi il n'y a pas de tranches temporelles ultérieures capables de revenir au même point de l'espace-temps.

Voici les révisions instantanées du nouveau script que je vous soumetts. Mettons que l'Entreprise-D échappe de peu à la collision. Le sursis ainsi obtenu revient à nouveau, mais depuis une perspective différente. Puis une troisième fois, et l'équipage voit deux autres vaisseaux manquant de peu la collision, etc. Lentement mais sûrement, l'équipage comprend qu'il est pris dans une boucle temporelle et a raté de peu la collision avec lui-même!

Mais j'oublie peut-être quelque chose. Les prémices de « Cause et effet », c'est qu'à chaque fois, dans la boucle, la même tranche temporelle de l'Entreprise-D revit les mêmes événements. Inlassablement, les événements de la boucle sont rejoués, le temps « se remonte » comme un ressort, et la même série d'événements se rejoue à l'infini dans son intégralité. Ces boucles n'impliquent pas de causalité inversée, mais un temps inversé. Mais si l'inversion temporelle est possible, ce qui se produit dans l'épisode ne l'est pas. Worf découvre que, pendant qu'ils étaient piégés dans la boucle, il s'est écoulé dix-sept jours pour le reste de l'univers: ce n'était donc pas le temps externe qui était inversé, mais le temps personnel de l'équipage. Jusqu'ici, tout va bien. Le problème, c'est que l'équipage ne cesse d'éprouver une sensation de déjà-vu, ce qui l'amène peu à peu à prendre conscience de sa situation. C'est impossible, étant donné l'inversion temporelle: cette remémoration implique l'inverse - que les seconde, troisième et quatrième fois passées dans la boucle sont ultérieures, en temps local, à la première. On ne saurait avoir l'inversion temporelle et en même temps le déjà vu ! Et bien sûr, le fait d'échapper à la destruction est déjà un autre chronobonus, donc l'épisode est doublement impossible.

Enfin voici l'« effet de Mannheim » observé dans « Nous aurons toujours Paris ». Picard, Riker et Data sont devant le turbo lift vide, et s'apprêtent à gagner la passerelle. Data dit : « Au fait, capitaine. Tes effets des distorsions temporelles se font à présent sentir dans le système d'Ilakom. » Le trio pénètre dans le turbolift, Picard dit: « Passerelle », et le turbolift semble amorcer l'ascension. Lorsque la porte s'ouvre, le trio - appelons le le trio A - se voit lui-même (trio B) à l'extérieur de la porte. Le trio A le contemple sans en croire ses yeux, et voici la conversation qui

s'ensuit:

Data B : « Au fait, capitaine, les effets des distorsions temporelles se font à présent sentir dans le système d'Ilakom. »

Picard A : « C'est nous, avant de monter dans le turbolift ! »

Picard B : « Cela se produit de nouveau. »

Data B : « Je ne me sens pas désorienté. »

Data A : « Moi non plus. »

Puis nous voyons les portes du turbolift se renfermer sur le trio A, et le trio B poursuit sa conversation :

Riker B : « Qu'est-ce que c'était ? »

Data B : « Je crois que ce que nous pourrions baptiser l'effet de Mannheim s'accroît. »

Riker B : « Nous revoici au point de départ. Si nous sommes nous. »

Data B : « Oh, nous sommes nous, monsieur. Mais eux aussi sont nous. Si bien que nous sommes tous nous. À différents points, le long du même continuum temporel. »

Le trio B attend anxieusement que les portes du turbolift s'ouvrent sur... une cabine vide. Ils entrent, et Picard B dit : « Passerelle. »

Voici une des rares occasions où les scénaristes ont vraiment sapé le travail - à tel point que je me demande si la séquence tout entière n'est pas une vaste blague. Il est clair que cette série d'événements est considérée comme une boucle de causalité se produisant sur une chronologie unique, comme le soulignent les propos de Picard et de Data. Par la suite, le second dit au premier : « Les distorsions temporelles se sont produites le long d'un même continuum, comme la reprise d'un point temporel spécifique. » Je crains que cela ne soit une erreur. Le problème, c'est que les scénaristes n'ont pas collé aux événements qui se produisaient antérieurement. Chaque membre du trio A devrait en être à une étape temporelle ultérieure par rapport à chaque membre du trio B. Mais lorsque le trio Best sorti du turbolift, et que Data B dit : « Au fait, capitaine, etc. », le turbolift est-il vide ? La réponse - impossible - est oui et non ! Picard B dit-il : « Cela se produit de nouveau ? » Oui et non ! Data B dit-il : « Je ne me sens pas désorienté ? » Oui et non ! Le trio B pénètre-t-il dans le turbolift juste après que Data a dit : « Au fait, capitaine ... » Oui et non ! etc. Pire encore, lorsque le trio B poursuit sa conversation, c'est comme s'il était, lui, une étape ultérieure du trio A puisque c'est Riker B qui dit : « Nous voici revenus au point de départ. » (Comparé à cela, le fait que le turbolift est allé quelque part sans aller nulle part est un détail négligeable.) A mesure que les boucles se succèdent, l'effet de Mannheim se démultiplie jusqu'à l'absurde !

Les philosophes ont joué avec une autre boucle en boucle. Elle n'apparaît pas dans Star Trek, mais j'en suggérerai une, avec votre permission. Il y a une boucle de

causalité tout à fait cohérente dans « Parallax » (V) où on voit se produire la séquence événementielle qui suit. Janeway reçoit un message brouillé par un canal d'urgence. Elle met le cap sur une singularité où l'équipage détecte la présence d'un vaisseau. Janeway envoie un message au vaisseau à travers un canal d'urgence. Et c'est le même message que celui qu'elle avait reçu en premier lieu - autrement dit, elle s'envoie à elle-même un message en remontant le temps ! Jouons un peu avec cette boucle en la systématisant. Une boucle semblable se serait fermée si Janeway avait réclamé la construction d'un engin recevant des messages et s'envoyant à lui-même des messages en remontant le temps. Quand on veut lancer un message, l'engin est programmé pour n'envoyer un message que s'il le reçoit (cela active un commutateur). Cet engin est forcément possible, même s'il est possible d'envoyer des messages en remontant le temps. Mais supposons que Janeway ordonne une modification dans sa conception : au lieu d'être activé par la réception d'un message, le commutateur est désactivé. Si le premier engin est réalisable, celui-ci l'est forcément aussi. Dans la mesure où il ne peut empêcher la réception du message s'il l'envoie, alors le second engin est conçu pour envoyer le message si et seulement s'il ne l'envoie pas ! Voilà où il devient forcément impossible. Quelque chose s'est détraqué, et certains philosophes ont suggéré de résoudre ce problème en niant la possibilité du voyage dans le passé - par des messages ou par quelque autre moyen.

Cette réaction paraît trop hâtive. Bien que nous puissions décrire pareil engin auto-invalidant en supposant que le voyage dans le passé soit viable, cela ne signifie pas qu'il soit possible d'en construire un. Soit une analogie. Je peux décrire un moulin dans lequel l'eau chute du haut d'une cascade vers une roue placée en bas. Je peux aussi décrire un moulin où, par la seule force de gravité, l'eau monte de la roue au sommet de la cascade. Je peux enfin décrire un moulin où les deux se produisent, alors qu'il n'y a qu'une roue et qu'une cascade. Je viens de le faire. Il existe même une description visuelle de ce moulin : La cascade, de M. C. Escher. Et pourtant, comme vous en conviendrez sûrement avec moi, nous ne pouvons construire ce moulin. Cela ne veut pas dire qu'une des composantes du moulin est impossible, mais seulement que la combinaison des composantes est impossible. De même, le fait que l'engin auto-invalidant soit impossible ne démontre pas que l'une des composantes (celle qui envoie des messages en remontant dans le temps) soit impossible.

La connaissance du futur

Je l'ai dit plus haut, la mauvaise compréhension des voyageurs face à ce qui leur arrive sert les intrigues où l'on voyage dans le temps. Mais ces histoires requièrent-elles une ignorance pure et simple ? Reprenons « La flèche du temps ». Lorsque la tête de Data est rattachée à son corps, le message vieux de cinq cents ans tapé par Picard est découvert. Mais nous savons que cette tête peut fonctionner indépendamment du corps de Data (cela se produit dans « Désastre » [NG]). La Forge aurait pu aussi bien la réactiver immédiatement, et le message aurait alors été récupéré avant que Picard ne parte en mission. Supposons maintenant qu'une partie seulement du message soit

retrouvée avant ce départ- celle qui identifie le message comme un avertissement de Picard. Picard sait donc qu'il s'agit d'un avertissement, mais il en ignore le contenu (et sait, par ailleurs, que le message finira par être complété).

Dans notre version revue et corrigée, Picard est dans une position assez délicate. Voyez son raisonnement dans la caverne, lorsque l'extraterrestre lui révèle le danger qui menace la terre si l'Entreprise-D tire avec la torpille à photons, comme c'est probable. Dans l'histoire originale, c'est ce savoir qui incite Picard à envoyer le message vers l'avenir. Dans notre version, Picard sait déjà que le message a été reçu dans l'avenir, et il n'a pas besoin de se donner à nouveau tout ce tracass pour l'envoyer, non? Mais s'il ne l'envoie pas, qui l'enverra? Du coup, il faut bien qu'il l'envoie. Imaginez que tout cela traverse la tête de Picard; peut-être ne peut-il que réaliser qu'il doit envoyer le message, et se résigner à le faire. S'il est sage, il adoptera une attitude fataliste et mènera la tâche à terme. Sinon, il se figera peut-être dans l'inaction.

Le problème, avec ce genre de connaissance, c'est que, selon les philosophes, elle menace la liberté de chacun. Si vous savez déjà ce que vous ferez, vous n'avez aucune possibilité d'y réfléchir rationnellement, et de mener une délibération authentique et efficace. Picard pense qu'il sait ce qu'il fera: il ne peut pas choisir librement de le faire. Il y a un risque véritable que le voyage dans le temps vous prive de votre liberté en compromettant la délibération efficace.

Les scénaristes de Star Trek, reconnaissons-le, en sont plus ou moins conscients. Dans « Le temps au carré », par exemple, lorsque le journal de bord de la navette « future » est découvert, Picard (le présent Picard, ou Picard 1) exprime son appréhension à l'idée d'entendre un journal qui n'a pas encore été enregistré. Le Dr Pulaski suggère que Picard 1 a peur d'être paralysé par le doute au moment d'agir - un doute lié au fait qu'il sait d'avance que l'Entreprise-D sera détruite. Riker a un peu le même problème dans « La flèche du temps », lorsqu'il essaie de soutirer des informations - sur Picard, piégé dans le passé - à Guinan, cinq cents ans après les événements de 1893.

Riker : « Vous étiez dans cette caverne. Vous savez ce qui s'y est produit. Qu'est-ce que je dois faire? »

Guinan: « Si je vous disais ce qui s'est passé dans cette caverne, cela pourrait affecter la décision que vous prendriez maintenant. Je ne peux pas le faire. Je ne vais pas le faire. »

Riker: « C'est si vous ne me le dites pas que cela peut affecter ma décision. Avez-vous pensé à cela? ... Je ne peux pas rester assis à espérer que tout ira bien. Il faut que je fasse quelque chose ! »

Il semble qu'une maigre connaissance du futur soit une chose dangereuse, et on se trouve bien mieux - bien plus libre - en ignorant ce qui se passera et ce qu'on fera. C'est du moins ce qu'avancent certains philosophes - encore que je doute que la liberté, ou la délibération efficace soient nécessairement compromises dans ces

situations.

La délibération efficace face à la connaissance

Data: « Il est clair que ma vie s'achèvera vers la fin du XIX^e siècle. »)

Riker: « Pas si nous pouvons y parer. »

Data: « Personne ne peut y parer, monsieur. À une date future, je serai ramené à la Terre du XIX^e siècle où je mourrai. Cela s'est produit. Cela se produira ! »

« La flèche du temps 1 »

Au chapitre 3, j'ai affirmé que l'action libre est produite par une volonté bien rodée, et qu'une volonté, en résumé, est une faculté composée d'un générateur d'opinion, d'un générateur de désir et d'un délibérateur. Une volonté bien rodée fait appel à ces trois sous-facultés pour produire une action, à condition qu'elles-mêmes soient bien rodées. Maintenant, dans un sens mineur, une délibération est efficace si le délibérateur est mobilisé - s'il joue un rôle causal dans la production d'une action à l'étude (en d'autres termes, si l'opération du délibérateur fait une différence dans la production de l'action). Je suppose qu'il paraît évident qu'une action produite par un libre arbitre est efficace dans ce sens mineur.

Disons, pour notre propos, que le sens majeur d'« efficace », c'est que la délibération est efficace si elle introduit une différence notoire dans ce qui se produit, et toutes les actions librement produites ne sont pas efficaces en ce sens. Si les actions sont inefficaces, alors la délibération qui les a produites est également inefficace. Voici un exemple de délibération inefficace adapté de Dan Dennett. Supposons que j'aie imprudemment sauté du sommet d'un immeuble élevé sans parachute, ou sans autre moyen d'arrêter ma chute. Je peux délibérer autant que je veux pendant ma chute, mais aucune de ces délibérations, aucune action qui en résulterait ne changeront quoi que ce soit à ce qui doit arriver. Je mourrai, quelle que soit ma décision: autant renoncer à délibérer et me résigner à mon destin.

Lorsqu'on sait à l'avance que nos actions et délibérations seront inefficaces, il se peut que délibération et libre arbitre soient compromis. Mais il me semble qu'on peut parfaitement savoir à l'avance ce qu'on fera tout en délibérant authentiquement, si l'on a de bonnes raisons de penser que cette délibération sera efficace. Par exemple, supposons que je me connaisse à fond. Je sais que je remets jusqu'à la dernière minute les tâches désagréables. Mais je sais aussi que je suis trop consciencieux pour ne pas les accomplir dans les temps. Par exemple, celle qui vient immédiatement à mon esprit d'universitaire, c'est la pile de copies à noter pour mercredi prochain. Je sais donc que je noterai ces copies mardi soir. Notez bien que je sais ce que je ferai, et quand je le ferai. Mais n'est-il pas étrange de dire que mardi soir je ne délibérerai pas pour savoir si je corrigerai ces copies, et n'agirai pas librement lorsque je m'assiérai pour les noter ? On ne saurait affirmer que j'agis librement en notant les copies parce que j'ai décidé à l'avance de noter les copies le mardi soir. Pourquoi ? Parce que la notation de ces copies exige sans aucun doute un

effort de volonté de ma part, et cet effort se produira le mardi soir, et non des jours à l'avance. Pour que les copies soient notées par moi, il doit se produire une délibération efficace le mardi soir.

Si l'on applique ce raisonnement au cas de Picard, revu et corrigé, dans la caverne, il sait qu'il enverra le message, mais cela ne l'empêche pas d'agir librement en l'envoyant. Bien sûr, ce sera une tâche désagréable, mais pas parce qu'elle ne rime à rien. Lorsque Picard se contraint, par un effort de volonté, à envoyer le message, il sait que sa délibération est effective puisqu'elle cause la réception du message au XXIV siècle. L'attitude que Picard se doit d'adopter entre-temps est philosophique au sens courant du terme - une acceptation résignée de son sort. De fait, cette attitude, loin de menacer sa liberté, lui est probablement essentielle lorsqu'il s'agit de connaître à l'avance ce qu'il fera.

Nous apprenons beaucoup de Data à cet égard (voyez l'exergue). Lorsque l'Entreprise-D est en route vers Devidia II, Riker tente de persuader Data de ne pas se joindre à la mission : il pense pouvoir empêcher par là la décapitation de Data, mais celui-ci en sait plus et refuse de permettre à la prescience de paralyser ses processus de délibération. De fait, il semble considérer que sa prescience lui donne une raison de se joindre à la mission. Clemens, lui aussi, adopte une attitude philosophique lorsqu'il décide de laisser sa montre dans la caverne. Il raisonne ainsi: la montre sera trouvée ici au XXIV siècle, autant l'y laisser. Et c'est bien ce qu'il fait.

Les meilleures choses ...

J'ai un faible pour ce dernier épisode de La nouvelle génération, où Q fait que Picard tourne en rond dans le temps, essaie d'empêcher la destruction rétroactive de toute forme de vie sur Terre, qu'il a déclenchée précisément en essayant de la prévenir. Cette histoire contient tous les ingrédients passés en revue. Le paradoxe du grand-père menace Picard lorsqu'il semble risquer de susciter sa propre non-existence. On trouve un paradoxe identique lorsque trois Entreprise-D (avec un Picard à bord de chacune) se croisent dans une faille temporelle. Plus une belle petite boucle de causalité lorsque Picard découvre que son moi futur souffre d'une maladie neurologique, diagnostiquée à l'origine par le Dr Crusher. Cela l'amène à chercher une aide médicale dans le passé, où le Dr Crusher découvre le défaut neurologique qui provoquera par la suite la maladie (voir le dialogue en exergue du chapitre). De plus, si nous prenons vraiment cette histoire au sérieux, il y a une boucle fermée, puisque Picard, en ramenant la vie sur Terre, suscite sa propre existence, ce qui n'explique pas son existence. Il semble parfois incapable d'agir parce qu'il sait ce qui se produira ou ce qui s'est produit; et il se produit même une inversion temporelle locale (qui permet à Geordi LaForge de retrouver la vue).

Malheureusement, cette histoire ne peut pas être jugée sur les apparences: en fait, elle n'a rien à voir avec une histoire de voyage dans le temps. Rappelez-vous les conditions que nous avons posées à l'origine : pour voyager dans le temps, il faut se rendre ailleurs ou aillheure. Mais si Picard se déplace dans le temps - il se déplace

entre le présent, l'avenir vingt-cinq ans après, et le passé sept ans avant -, il ne le fait pas de façon concrète. Pour les autres qui l'entourent, il semble seulement disparaître un moment ou deux. Peut-être ce qui se produit n'est-il pas un voyage dans le temps, mais une communication interne entre diverses tranches temporelles de Picard. Même si c'est ce qui se produit (avec un coup de pouce de Q), on ne saurait faire sens de tout ce qui se passe.

Tout d'abord, prenons le supposé chronoclasme. Si l'épisode postulait une faille entre le temps et « l'anti-temps », il s'en produirait bel et bien un. (C'est clairement visible dans les cieux, lorsque Q ramène Picard sur Terre au premier matin.) Comment les choses peuvent-elles redevenir normales sans contradiction explicite ? En outre, dans chacune des trois périodes différentes entre lesquelles Picard est censé se déplacer, l'Entreprise-D est détruite dans la (aille temporelle. Il est clair (comme dans le paradoxe du grand-père) que seule la tranche temporelle plus âgée-de-vingt-cinq-ans de l'Entreprise-D peut être détruite, et que ce serait vraiment la fin de Picard et compagnie. Et que penser de l'incohérence inhérente à l'histoire ? La faille entre temps et anti-temps remonte le temps, augmentant de taille à mesure qu'elle le fait (c'est pourquoi la Terre des origines est menacée). Mais vingt-cinq ans plus tard, Picard, le Dr Crusher, LaForge et Data cherchent la faille et ne la trouvent pas ; ils reviennent un peu plus tard, et elle est là. Entre-temps, ils ont contribué à l'élargir - en la sondant avec une sonde à tachyons inversée - mais forcément, la faille ne sera pas présente après son apparition, si elle remonte le temps ! Sans compter l'autre incohérence interne : à un moment donné, Q montre à Picard en quoi ses actions ont empêché la vie de se former sur la Terre, à l'origine. Pourquoi tout n'a-t-il pas déjà changé - puisque aussi bien le chronoclasme a disparu ? Supposons que nous invoquions une hypothétique protection par particules de chronoton, qui maintienne en l'état les éléments autour de Picard à trois périodes différentes ? Mais c'est impossible, parce que lorsque Picard demande à Q, à la fin de l'épisode, si l'humanité a été sauvée, Q dit : « Vous êtes toujours là, non ? »

Toutes choses considérées, je dois approuver la suggestion que fait à Picard le Dr Crusher : « Je veux que vous admettiez cette possibilité : rien de tout ceci n'est réel. » Ce qui s'est produit n'est qu'un vaste fantasme inventé par Q pour mettre à l'épreuve Picard, et il n'y a rien de surprenant à ce que Picard note dans son journal de bord qu'il est le seul à s'en souvenir ! Toutefois, même Data renie tout ce qu'il sait du temps à la fin de cet épisode. Voici un extrait de la conversation finale :

Dr Crusher : « Je repensais à ce que nous a révélé le capitaine de l'avenir ... pourquoi a-t-il choisi de nous dire tout ce qui doit advenir ? »

Geordi : « Ça contredit tout ce qu'on nous a inculqué sur la pollution de la chronologie, non ? »

Data : « Je crois, toutefois, que cette situation est unique. Puisque l'anomalie ne s'est pas produite, il y a déjà eu des changements dans la façon dont la chronologie se déroule. L'avenir dont nous faisons l'expérience est indubitablement différent de celui dont le capitaine fait l'expérience. »

Riker: « C'est peut-être la raison pour laquelle il en a parlé. Savoir ce qui adviendra nous permet de changer les choses maintenant, si bien que certaines choses n'arriveront jamais.»

De tous les propos sur le temps que j'ai présentés, celui de Riker s'apparente le plus au non-sens. Peut-être a-t-il un sens dans certaines perspectives, peut-être même est-il véridique, mais j'en doute. Bien sûr, je n'ai pas passé en revue toutes les perspectives philosophiques sur le temps. Dans certaines, le temps est essentiellement un élément psychologique, dans d'autres, cette notion n'existe pas! Mais nous n'avons pas le temps d'étudier tout cela ici. Sur cette remarque, admettons que « les meilleures choses ... » ont une fin.

Conclusion

En ces dernières années du second millénaire, la science se voit attaquée sur tous les fronts. Les conservateurs religieux accusent les scientifiques de prêter une foi aveugle au dogme darwinien. Les penseurs du New Age tiennent le matérialisme naturaliste pour une superstition qu'il faut mettre au rebut si nous voulons progresser; les féministes remettent en question la tradition intellectuelle occidentale tout entière, science incluse, comme une tradition phallocrate; les non-blancs la tiennent pour eurocentriste ; les postmodernes la rejetteraient dans son intégralité si le concept de rejet n'était pas lui-même entaché d'un préjugé intellectuel occidental. Même les gens ordinaires expriment un certain mépris envers la science, tenue pour largement responsable des dégâts que nous observons aujourd'hui. Plus les critiques conservateurs observent la science comme la Fédération, au XXIV^e siècle, observe l'Empire klingon - comme un allié avec qui le pacte est incertain, et dont les tentatives d'expansion requièrent leur vigilance-, plus les radicaux l'observent comme ils observeraient les Borgs: à force de s'y impliquer, ils finiraient par s'y assimiler.

Dans la mesure où règne une thèse antiscientifique cohérente, commune à ces groupes, elle se divise en deux parties. Tout d'abord, tous considèrent que la science est nécessairement bornée dans ses ambitions. Puis ils la soupçonnent fortement d'avoir plus ou moins bluffé son monde en se donnant pour illimitée, en laissant croire qu'elle détient la clef de toute question essentielle.

J'approuve entièrement la théorie selon laquelle la science est nécessairement bornée dans ses ambitions, mais je n'irai pas plus loin. La science, en sa nature et sa visée, pose des questions profondes et délicates qu'examinent bien des esprits subtils: ainsi, la question de savoir s'il suffit d'appliquer avec succès une théorie scientifique pour y adhérer. Mais bien des critiques populaires, aux avis tranchés, puisent dans une ignorance totale de la science et de la philosophie, ou, pire encore, dans une familiarité de surface qui fonde leur hostilité. Ainsi, certains affirment que les théories scientifiques devraient s'en tenir par principe à ce qui peut faire l'objet d'expériences répétées, et que la science devrait nécessairement garder le silence sur, par exemple, l'histoire biblique (silence tout à leur avantage, soit dit en passant).

Mon opinion personnelle est diamétralement opposée. Je vois une certaine ironie dans ce déchaînement de critiques et d'avertissements en matière de science, car je partage la vision de Star Trek : si nous pouvons investiguer les questions scientifiques en nous gardant de toute superstition, nous avons sans doute une chance

de progresser non seulement dans notre technologie mais dans notre humanité. (Et Star Trek n'est pas sans prodiguer, à son tour, quelques avertissements. Un thème souvent évoqué, c'est que les sociétés très avancées ont perdu quelque chose de précieux que nous autres, humains relativement primitifs, avons en abondance - une innocence enfantine, une curiosité illimitée ou des émotions intenses. Un autre thème, c'est que les scientifiques tendent fortement à poursuivre leurs investigations, quel qu'en soit le prix.) Lorsqu'on observe combien s'est répandue la croyance populaire dans les kidnappings ou les visites perpétrés par les extraterrestres, dans les phénomènes psychiques, etc., et combien les médias nous désinforment à ce sujet, on pourrait désespérer d'un quelconque progrès. L'entreprise scientifique est humaine, trop humaine, et elle a ses limites ; mais, en refusant de l'appliquer lorsqu'elle peut nous transmettre des informations importantes, on sombre dans une incohérence aussi dangereuse que fondamentale.

Les critiques affirment souvent que le progrès scientifique et technologique est inutile ou nocif s'il ne s'accompagne pas d'un progrès philosophique. Qui nierait ce fait? Mais bien des critiques parlent de régression philosophique sous forme d'un retour aux superstitions traditionnelles ou à un relativisme obtus. Même si la philosophie ne faisait que corriger cette forme de pensée, il serait utile de la pratiquer. Or elle peut beaucoup plus. Une bonne partie de cet ouvrage traite de la façon dont une réflexion sur les possibilités à venir de la science peut étayer la philosophie actuelle : ma vision de l'avenir inclut l'espoir que la science puise de plus en plus dans la philosophie. Ce n'est pas là ce que suggère Star Trek, à mon grand regret. Les personnages principaux manifestent souvent une ignorance considérable de la philosophie, qui contraste avec leurs acquis dans d'autres domaines. Tous les membres, ou peu s'en faut, de la Fédération savent l'art de manipuler ou d'installer le moindre élément technologique à bord d'un vaisseau spatial, et chaque fois qu'on entame une discussion de haut vol sur la physique théorique, tout l'équipage peut suivre. Les spectateurs scientifiques, eux, ergotent sur les détails et réagissent violemment au jargon scientifique - encore faut-il rappeler que Star Trek a longuement élaboré sa représentation d'une technologie et d'une science futures. Cela dit, le niveau général de l'équipage en matière de philosophie est celui du citoyen lambda occidental en cette fin de XX- siècle.

Je n'entends pas condamner les scénaristes - loin de moi cette idée, car Star Trek a au moins le mérite de poser des questions philosophiques. Qui plus est, pour séduire un large public, il faut manifestement présenter des personnages auxquels il puisse s'identifier. Mais, en tenant compte de ces contraintes, il est possible de représenter un avenir plus réaliste en matière de philosophie. En accélérant le progrès scientifique et technologique - voyez les manipulations génétiques, la possibilité de maintenir les gens en vie même si cela ne vaut plus tellement le coup, les greffes de cerveau, le clonage, les progrès incessants en matière d'informatique, etc.-, on demande à l'individu ordinaire de remettre en question les opinions métaphysiques qui lui sont chères. Et parfois la raison imposera même d'y renoncer. On dit souvent que nous menons une énorme expérience physiologique sur nous-mêmes

lorsque nous soumettons notre organisme physique, qui relève essentiellement de l'âge de pierre, à un régime moderne occidental, sans savoir exactement ce que sera le résultat. De même, nous menons une énorme expérience psychologique sur nous-mêmes lorsque nous soumettons nos perspectives relativement primitives à la pléthore d'informations modernes. Ce que j'espère, c'est qu'à l'avenir nous adapterons plus facilement nos opinions que nos organes, et nous avons, Dieu merci, toutes les raisons de le croire. Tout l'avantage d'un comportement largement influencé par nos opinions et désirs, c'est que nous sommes en principe des êtres flexibles, capables de réagir relativement vite aux changements de notre environnement. Si nous réussissons à surmonter une inertie provoquée par des siècles de superstition, nous aurons encore un avenir digne d'intérêt - une ère de paix sur la terre aux hommes et aux femmes de bonne volonté, voire un avenir tel qu'il est dépeint dans Star Trek. Après tout, nous pourrions faire pire.

Table des matières

Introduction.	
L'entreprise philosophique	002
Première partie: Une nouvelle vie, de nouvelles civilisations	
1. Suspects numéro un	006
2. Des données insuffisantes	032
3. Procréation	053
Deuxième partie : Questions de survie	
4. Être ou ne pas être téléporté	089
5. L'épanouissement personnel	117
6. Les distorsions temporelles	140
Conclusion	167
Remerciements	170

Remerciements

Mes remerciements vont à Robyn Shuster, le premier à voir dans Star Trek un outil pédagogique en matière de philosophie ; aux départements de philosophie respectifs de l'université de Maryland, de College Park et de l'université centrale de Michigan, qui m'ont autorisé à prodiguer des cours basés sur Star Trek ; à Gary Fuller, qui a lu la première version de cet ouvrage ; à Val Kelly, qui a lu une partie de la seconde version ; à Sam Gates et Steven Glaister, qui ont commenté le chapitre 6 et ont offert des suggestions générales ; à Steve Matthews, dont les avis se sont révélés précieux ; à Ian Mond et David Hull, qui m'ont prêté des cassettes vidéo ; à Callao Ledsham, qui m'a encouragé et assisté dans mes recherches ; à l'équipe de Blockbuster Video (Mount Waverley) et de Movieland (Ashwood) pour leur assistance et leur courtoisie ; aux membres, enseignants et étudiants de l'université de Maryland, de College Park, de l'université centrale de Michigan, et de l'université de Monash, qui m'ont soutenu moralement ; à ma famille et à mes amis, dont la présence fut constante ; à tous ceux, qui continuent à m'enseigner la philosophie, notamment Fred 4~. John Bigelow. Michael Devitt, Richard Holton, Jerrold Martin. Michael Smote et Aubrey Townsend.

Cet ouvrage n'aurait jamais vu le jour sans le soutien de Susan Rabiner, ma correspondante à Basic Books, qui accepta ma suggestion et m'aida à en faire une idée. L'idée se fit manuscrit, mais je n'ai fait du manuscrit un livre qu'avec les efforts incessants, l'intuition pénétrante et le soutien vigoureux de Sara Lippincott. S'il subsiste quelques défauts, j'en suis le seul responsable. Le titre, La métaphysique de Star Trek, joue en partie sur les mots : le terme « métaphysique » tient son sens d'un commentaire d'Aristote, comme ce qui vient littéralement après la physique. En jouant sur cet héritage, j'adresse ma gratitude à Lawrence Krauss, auteur de La physique de Star Trek.

Enfin, je remercie ma femme, Joanne Sampson, parce qu'elle m'a prodigué un soutien sans faille, a flatté mon caprice pour Star Trek avec un dévouement extrême, et a donné naissance à notre fille alors que cet ouvrage était encore en gestation.

Qu'est-ce qui caractérise fondamentalement l'humanité et pourquoi peut-on désirer être humain? Du Dr Spock, qui préfère les Vulcariens aux hommes, trop largement soumis à leurs émotions, à Data l'androïde, qui les mime désespérément dans l'espoir de leur ressembler, les personnages de la série Star Trek nous conduisent à méditer sur les grandes questions philosophiques liées à l'humanité: la raison opposée aux passions, la transcendance, le temps, l'intelligence artificielle, la connaissance, le langage, la morale ...

Si Star Trek nous conduit ainsi au cœur de la réflexion métaphysique, c'est parce que, affirme Richard Hanley, « en observant ses diverses péripéties, nous sommes contraints d'accueillir des idées que nous aurions refoulées par préjugé ».

Pour s'initier en s'amusant à la philosophie.

Richard Hanley est professeur de philosophie à l'université Monash en Australie.