

3 - POISSON

Esprit et machines

Simmons fit le tour de la table, servant à chacun des invités une sole meunière légèrement grillée et nageant dans le beurre. Snow réfléchissait au paradoxe soulevé par Wittgenstein. Comment une suite de 0 et de 1 sur une bande, ou aussi bien un schéma de *MARCHE/ARRET* parmi les neurones du cerveau, pouvaient-ils conduire à des pensées ? Comment était-il possible que de simples collections de symboles plus ou moins arbitraires, écrits sur une bande ou stockés dans le cerveau, puissent *signifier* des choses aussi diverses que le son d'un carillon, l'éclair de la foudre, ou ce paradoxe même de la relation entre symboles et pensée qui l'intriguait en ce moment ?

Il devait bien y avoir quelque chose de plus dans la pensée qu'un simple changement d'une suite de 0 et de 1 en une autre suite, indépendamment de la vitesse ou de la manière dont ces suites étaient transformées les unes dans les autres. Turing n'était certainement pas sérieux, n'est-ce pas, quand il prétendait qu'une machine ne faisant rien d'autre que ce genre de mélange de symboles était effectivement capable de *reproduire* les processus de pensée de l'esprit humain ?

Exprimant cette gêne, Snow se tourna vers Turing et dit : "Je pense parler au nom de la plupart d'entre nous si je dis que je trouve littéralement incroyable d'imaginer qu'une machine qui ne fait que déplacer des 0 et des 1 sur une bande puisse être capable d'une pensée humaine. Je pense que nous saisissons mieux l'essence de votre argument si vous vouliez bien nous expliquer *exactement* comment vous pensez que ces suites de symboles dans votre machine en viennent à signifier quelque chose."

Délaissant un moment le mets raffiné de son assiette, Haldane appuya la demande de Snow. "Je suis aussi troublé par la façon dont les manipulations de suites de symboles sur le ruban d'une machine à calculer peuvent conduire à des objets aussi sémantiquement riches que ce filet de sole dans mon assiette. Dites-moi, Turing, où dois-je regarder sur votre ruban pour trouver cette délicieuse bouchée de poisson, eh ?" demanda-t-il, tenant

un morceau de sole au bout de sa fourchette. "Répondez à cela, si vous voulez, et je vous accorderai que votre machine est capable d'avoir des pensées comme les miennes."

Voyant les difficultés qu'il devrait affronter pour tenter d'expliquer comment cette transfiguration de la syntaxe à la sémantique pourrait bien se passer, Turing regarda fixement la pluie qui battait les vitres de plus en plus fort. S'agitant nerveusement sur sa chaise, il était quelque peu dérouté par le paradoxe que soulevait Snow et par l'intensité avec laquelle Haldane avait présenté sa question. Comment pouvait-on *expliquer* scientifiquement une douleur à l'estomac ou une conviction bien ancrée ? Avec quel genre d'argument logique puis-je bien convaincre un matérialiste farouche comme Haldane ou Schrödinger que l'intelligence consiste seulement à suivre le bon ensemble de règles ? Quoi que je dise, Wittgenstein s'opposera à moi jusqu'à son dernier souffle. Pourquoi donc ai-je accepté de venir ce soir à cette réunion ? La situation est vraiment désespérée. Mais puisque j'y suis jusqu'au cou, autant aller de l'avant et espérer une issue favorable.

Buvant une longue gorgée d'eau, et s'éclaircissant la voix, Turing se laça dans une explication sur la manière dont les symboles sur le ruban pourrait donner lieu à une pensée véritable. D'abord il expliqua au groupe comment n'importe quel type d'idée, d'objet ou d'action qui peut s'exprimer dans un langage peut être codé en une suite de 0 et de 1 sur le ruban d'un calculateur.

"Prenons le filet de sole dont parlait Haldane à l'instant. Supposons que je veuille représenter le mot *SOLE* sur le ruban de ma machine. Une manière simple de procéder est de construire un schéma dans lequel une suite unique de 0 et de 1 correspond à chaque lettre de l'alphabet latin. Il y beaucoup de façons de pratiquer ce codage, mais laissez-moi vous en présenter une. Prenez un bloc de huit cases contenant chacune soit un 0 soit un 1. Il y a donc en tout $2 \times 2 \times 2 \times 2 \times 2 \times 2 \times 2 \times 2 = 2^8 = 256$ suites différentes de 0 et de 1 que peut montrer ce bloc de huit cases. Je peux donc associer à chacune de ces suites un symbole parmi 256, et 256 symboles suffisent largement à représenter toutes les lettres et symboles que l'on peut trouver dans un grand dictionnaire ou dans les langues européennes. Je peux décider par exemple que la suite 00000001 représente la lettre

minuscule *a*, la suite 00000010 la lettre *b* et ainsi de suite. De cette manière, on peut attribuer à chaque lettre de l'alphabet, aux chiffres de 0 à 9, aux signes de ponctuation et aux autres symboles typographiques comme § ou & leur code personnel sous la forme d'une suite de huit 0 et 1. En utilisant ce codage, je peux représenter le mot *SOLE* sur le ruban par un ensemble de quatre groupes de huit cases, un groupe pour chacune des lettres du mot. En utilisant les marques de ponctuations et les autres symboles de la langue écrite, je pourrai de la même manière coder sur le ruban n'importe quelle notion qui peut être communiquée sous forme écrite."

"Ceci est parfaitement clair," dit Schrödinger, "mais cela revient simplement à remplacer un ensemble de symboles par un autre."

"C'est juste," acquiesça Turing. "Une fois que nous avons codé une pensée ou une situation particulière sous la forme d'une suite de 0 et de 1, le programme de la machine convertit cette suite en une autre suite. Cette nouvelle suite peut alors être *décodée* en une série d'affirmations en langage ordinaire, pouvant par exemple exprimer la texture et le goût du morceau de sole au bout de la fourchette de Haldane, ou même les pensées qui traversent son esprit à l'idée du plaisir qu'il recevra quand il mordra dedans."

"Etes-vous en train de dire que les règles par lesquelles le programme modifie les symboles sur le ruban accomplissent la même chose que le cerveau humain quand il bascule ses neurones d'*ARRET* à *MARCHE* au cours d'une pensée ?" s'enquit Snow, les sourcils levés en signe d'incrédulité.

"Essentiellement, oui. Bien sûr, nous ne savons pas encore vraiment quelles règles emploie le cerveau. En fait, nous ne savons pas vraiment *comment* le cerveau stocke et utilise ces règles."

"Je me fais de la pensée une autre idée qu'une transformation purement passive de symboles suivant une règle prédéfinie," intervint brutalement Haldane. "Mais, éventuellement, si le programme a, d'une certaine façon, la capacité *d'apprendre* en examinant le résultat de ses actions et en les modifiant en conséquence, alors ce serait, peut-être, suffisant pour que nous soyons tentés d'appeler un tel programme *intelligent*."

Reprenant la parole, Turing continua à expliquer comment, à son avis, le cerveau fabriquait des pensées. "Je suis convaincu que ce que fait une machine en modifiant des symboles sur une bande est l'exact analogue de ce que fait le cerveau quand il modifie les ensembles de neurones qui déchargent et engendre ainsi ce que nous appelons des *pensées*."

"Pur sophisme," s'écria Wittgenstein. "Où voyez-vous donc la *signification* du mot *SOLE* dans toute cette symbolique ? Comment pouvez-vous prétendre que des suites de 0 et de 1 assemblés selon un code complètement arbitraire puissent en quoi que ce soit *signifier* ce filet de poisson accroché à l'extrémité de la fourchette de Haldane ? L'acte de nommer ce morceau de protéine que nous appelons un "filet de sole" ne peut survenir que dans le contexte d'un langage développé, un langage dans lequel existent déjà des règles pour identifier des objets, les nommer et effectuer des opérations à leur sujet. Les critères mis en jeu ne sont pas la logique des machines, des rubans ou des codages, mais la pratique effective d'une *communauté* de langage. Vous ne pouvez pas transmettre ce genre de signification à une suite morte de symboles en construisant simplement un ensemble de règles pour transformer ces suites en de nouvelles suites."

"Voulez-vous dire que la signification émerge seulement d'une sorte de consensus social qui nous permet de communiquer les uns avec les autres par le langage naturel ?" demanda Snow d'un ton surpris.

"Exactement. La signification ne peut venir que de la participation à un jeu de langage. Des machines à calculer ne joueront jamais au genre de jeu auquel nous jouons en ce moment. Turing se trompe complètement en imaginant que ce qu'une machine considère comme un filet de sole a une quelconque ressemblance avec ce que *nous* considérons comme un filet de sole. Nous avons un avis commun sur ce morceau de poisson parce que nous partageons le même genre de vie. Si la machine jouait un jour à un jeu de langage, ce dont je doute fort, ce ne serait sûrement pas à l'un des jeux auxquels jouent les humains. En dernier ressort, la signification procède de la pratique sociale, et non de la logique."

Wittgenstein s'excitait de plus en plus en exposant ses idées, et devant cette agitation Schrödinger se pencha et posa la main sur le bras de son concitoyen pour l'apaiser un peu. "Un instant, Wittgenstein," lui dit-il, "Turing n'a peut-être pas tout à fait tort ici. Même si la signification que nous attachons à cette délicieuse bouchée de sole vient effectivement de notre expérience humaine et de notre participation à une communauté de vie, il n'est pas clair à mes yeux qu'il soit impossible que cette signification soit codée dans les circuits de neurones de notre cerveau comme Turing affirme qu'elle pourrait être codée sur le ruban d'une machine à calculer. Malgré tout, je ne vois pas au terme de ce raisonnement comment le comportement intelligent associé à un apprentissage intervient dans le schéma proposé par Turing."

"Oui," intervint Snow. "D'après la description de Turing, il ne semble pas que sa machine puisse faire autre chose que de déplacer des symboles sur la bande, en suivant les instructions *prédéfinies* de son programme. Mais ce n'est pas du tout le comportement des humains. Nous sommes toujours prêts à changer d'avis, à nous adapter à des circonstances nouvelles, et à réagir différemment devant des situations en apparence identiques. En général, notre comportement est étrange et imprévisible. Si une machine à calculer ne peut faire cela, je ne vois pas comment elle pourrait un jour montrer quelque chose d'analogue à l'intelligence humaine."

Turing répondit qu'il était d'accord avec Snow et Schrödinger. "Une machine à calculer ne montrera une certaine intelligence que si elle peut modifier son programme en réaction à de nouvelles informations. Pour cela, il faut que la machine lisant les données présentées sur le ruban dispose de règles pour modifier les règles de fonctionnement qu'elle utilise à cet instant là, des *métarègles*, si vous voulez. De cette manière, le programme peut apprendre et s'adapter — exactement comme le font les humains — à un environnement qui change, et aux circonstances qu'il "perçoit" à travers les données d'entrée sur le ruban."

Schrödinger intervint tranquillement : "Pour cela, vous devez accorder à la machine les mêmes données sensorielles que celles dont nous disposons, et lui permettre de "grandir", en quelque sorte, dans le même environnement qu'un humain, n'est-ce pas ?"

Turing réfléchit un moment, et répondit : "Permettez-moi de reformuler quelque peu ma thèse sur la possibilité logique d'une machine pensante. Elle s'appuie en grande partie sur la possibilité d'imiter pratiquement tout organe humain par un mécanisme. Le microphone imite l'oreille, et la caméra l'œil. Les questions dont nous discutons concernent essentiellement le système nerveux. Nous pouvons certainement construire des modèles électriques relativement précis copiant la structure des nerfs, bien que cela n'ait guère d'utilité pratique. Cela ressemblerait à vouloir à tout prix construire des voitures avec des jambes, plutôt que d'utiliser des roues."

"Voulez-vous vraiment dire que, pour construire une machine pensante, nous devrions prendre un homme et nous efforcer de remplacer chaque organe par une machinerie ?" attaqua Haldane. "Ce serait sans nul doute un travail colossal, et même si c'était possible, la créature n'aurait aucun contact avec la nourriture, le sexe, le sport, et bien des choses qui présentent un intérêt considérable pour les êtres humains."

"Bien que ce soit là sans doute la manière assurée de produire une machine pensante, elle me semble beaucoup trop lente et impraticable," répondit Turing. "Ma proposition est au contraire d'essayer de construire un "cerveau" qui n'aurait pas de corps, et en ne lui fournissant guère que la vue, l'ouïe et la parole. Bien entendu, il nous faudra alors trouver les branches de la pensée dans lesquelles la machine pourrait montrer ses capacités. A ce sujet, je pense que des domaines comme le jeu d'échecs, la cryptographie ou les mathématiques sont de bons candidats parce qu'ils nécessitent peu de contacts avec le monde extérieur."

"Supposons pour le moment que vous puissiez donner à votre machine ces entrées sensorielles," dit Haldane. "Comment pourrait-elle les utiliser en pratique pour modifier son programme, et apprendre ainsi quelque chose, même dans un domaine aussi restreint que les échecs ou les mathématiques ?"

"Eh bien, " réfléchit Turing, "nous commencerions sans doute avec une machine dotée de capacités minimales, à peine capable de mener des calculs élaborés ou de réagir de manière disciplinée à des ordres. Ensuite, en agissant sur elle d'une manière qui imite

l'éducation, nous pouvons espérer modifier la machine jusqu'à pouvoir attendre des réactions précises à certaines commandes."

"Prenons un domaine très réduit comme les mathématiques," dit Wittgenstein. "Comment la machine pourrait-elle apprendre les mathématiques ?"

"Dans le cas des mathématiques, cela nécessiterait d'informer la machine de l'existence d'objets comme les points et les lignes, et des opérations logiques permettant de construire de nouveaux ensembles. Mais je ne peux pas vous décrire complètement en détail la façon de procéder, pour la simple raison que je ne la connais pas, pas encore du moins! C'est le principal axe de nos recherches en ce moment. Mais je suis pleinement convaincu qu'il n'y a pas de barrière logique ou technique qui s'oppose à ce projet. Ce qui manque pour le moment, c'est la volonté de la faire, et bien entendu les ressources financières."

Haldane remarqua alors : "Vous parliez il y a un moment des travaux de McCulloch et Pitts sur la modélisation des circuits neuronaux du cerveau par des formules mathématiques qui pourraient, en principe, être réalisées par des composants électroniques modernes comme des relais, des lampes ou des valves. Il est possible qu'un ajustement de l'intensité des connexions entre les neurones artificiels d'un tel circuit donne à la machine la capacité d'apprendre et de s'adapter."

"C'est exactement ce que je pense," répliqua Turing. "Laissez-moi vous montrer à quoi pourrait ressembler l'un de ces réseaux de neurones artificiels."

Reprenant son carnet, Turing esquaissa un diagramme montrant comment pourrait être agencé un réseau de neurones idéalisés, tels que McCulloch et Pitts les avaient imaginés.



FIGURE

"Je considère le schéma *MARCHE/ARRET* des neurones d'entrée comme l'analogue des symboles d'entrée sur le ruban du calculateur, et le schéma de décharge des neurones de sortie comme l'équivalent des symboles de sortie sur le ruban. Je dis

alors que les connexions entre neurones transforment le schéma de la couche d'entrée en schéma de sortie, et qu'elles accomplissent donc la même fonction que le programme du calculateur. Par conséquent un réseau de neurones et une machine de Turing sont parfaitement équivalents. Ce que l'un peut faire, l'autre le peut aussi."

"Cela signifie-t-il que McCulloch et Pitts ont démontré qu'il n'y avait pas de différence entre un réseau de neurones mathématiques et le calculateur abstrait dont vous parliez plus tôt ?" demanda Snow.

"Oui, c'est précisément ce qu'ils ont démontré," répondit Turing, bégayant dans son enthousiasme à décrire son travail. "Le réseau de neurones et la machine effectuent exactement les mêmes opérations. Dans le jargon mathématique, nous disons qu'ils sont isomorphes. Tout ce que peut faire un réseau de neurones peut être accompli par une de mes machines à calculer, et vice-versa. Ainsi, pour prendre un exemple, l'état mental de Haldane quand il pense au morceau de sole au bout de sa fourchette peut être considéré comme une simple étape du programme d'un calculateur. Puisque nous pouvons construire aussi bien des réseaux de neurones que des calculateurs en utilisant des dispositifs électroniques, nous pouvons les considérer comme fournissant une théorie électromécanique des états mentaux. En un sens, la biologie égale l'électronique."

"Vous êtes bien silencieux, Schrödinger," observa Snow. "En tant que physicien, que pensez-vous de l'idée de construire un cerveau artificiel avec les lampes, des fils et des valves ?"

"Un point m'est obscur pour le moment." répondit Schrödinger, "Turing pense-t-il que nous puissions construire une machine électronique qui *imite* certaines fonctions du cerveau humain ? Ou veut-il dire qu'il est possible de copier fidèlement, de *reproduire* un cerveau humain sous forme électronique ? Je me demande, Turing, si vous pouvez éclairer ce point ?"

"Je vais essayer. Je ne pense pas que cela fasse de toute façon une grande différence pour ce qui est de montrer un comportement intelligent à la manière humaine. A moins bien sûr que vous pensiez qu'il y a quelque chose de particulier dans la constitution

matérielle du cerveau humain qui explique ses capacités cognitives, et que ce quelque chose de particulier ne puisse être capturé dans des circuits électroniques."

"Eh bien, je ne pense pas qu'il y ait besoin d'un voyant extralucide pour deviner votre position là-dessus, Turing," dit Snow en souriant. "Mais dites-la nous quand même."

"Bien entendu," répondit Turing. "Mes sentiments sont clairs et définis sur le sujet. Je pense que la composition matérielle de notre cerveau n'a absolument rien de spécial, du moins en ce qui concerne la pensée."

"Vous pourriez dire que l'essence n'est pas essentielle, eh ?" plaisanta Haldane.

"C'est certainement ne façon de présenter les choses," répondit Turing avec une ombre de sourire. "Ce qui est essentiel, par contre, c'est ce que les composants du cerveau, les neurones, font effectivement, et la manière dont ils sont reliés les uns aux autres. Ce sont ces aspects structurels et fonctionnels qui donnent au cerveau sa capacité cognitive. Si nous pouvons construire des neurones électroniques et les relier comme ils sont connectés dans le cerveau humain, je suis convaincu que ce dispositif électronique contiendrait des règles gouvernant la pensée et l'action, règles équivalentes à celles qui sont présentes dans le cerveau humain. Cette machine posséderait les mêmes fonctionnalités que le cerveau humain. L'intelligence de la machine vient de la complexité de l'ensemble des règles qui constituent le programme, et non des étapes individuelles de ce programme, qui peuvent être extrêmement élémentaires comme nous l'avons vu dans l'addition de 1 et de 2."

Snow vit que par ces assertions audacieuses, Turing identifiait le mécanisme du calcul avec la continuité entre apprentissage, traitement des informations et activité de connaissance, continuité affirmée par les psychologues du comportement, les behavioristes.

Schrödinger précisa : "Comme je vois les choses, vous remplacez le réseau causal des relations stimulus-réponse des behavioristes par la notion de réseau des règles mécaniques. Vous affirmez que l'on peut ainsi capturer la grammaire logique de

l'intention, du choix et de l'apprentissage dans le cadre d'un ensemble de règles mécaniques, règles qui peuvent s'inscrire dans le programme d'un calculateur."

Wittgenstein repoussa sa chaise et se dressa d'un bond, incapable de se contrôler plus longtemps. Allant et venant à grandes enjambées à travers la pièce, les yeux perdus dans un royaume au-delà du temps et de l'espace qu'il était seul à voir, il contesta les dires de Turing.

"Dans quelles circonstances peut-on dire que quelqu'un suit une règle ? Si vous poussez des boutons marqués 20, 25 et \times sur un calculateur, cela ne veut *pas* dire que vous avez calculé 20×25 . La question "Comment êtes-vous arrivés à la bonne réponse ?" est une question au sujet des règles utilisées. Apporter la bonne réponse ne permet pas de dire que quelqu'un, ou quelque chose, a calculé."

"Mais Turing dit que la seule chose importante est le comportement, et non la manière dont le comportement a été acquis," remarqua Schrödinger.

"Et moi je dis que si le fait de calculer nous paraît mécanique," continua Wittgenstein en lançant un regard furieux à celui qui avait la témérité de l'interrompre, "c'est que l'homme qui effectue le calcul se comporte comme une machine. Ce n'est pas parce qu'on peut mécaniser une règle que celle-ci est "mécanique". Toute règle peut être considérée comme la description d'un mécanisme. Les machines de Turing ne sont rien d'autre que des hommes qui calculent. Mais la question de savoir si une machine pense ne peut recevoir aucune réponse, tout simplement parce que c'est une question absurde, logiquement absurde. Autant demander de quelle couleur est 3 !"

Ensuite, comme un ballon qui se dégonfle doucement, Wittgenstein parut à bout de ressources. Avec un air de distraction peinée sur le visage, il retourna à sa chaise et s'y laissa lourdement retomber en regardant fixement la table, paraissant se désintéresser complètement de la conversation.

Pendant ce temps, Haldane s'était tourné vers Turing et lui dit : "Les arguments de Wittgenstein semblent indiquer que la pensée est bien autre chose que le fait de suivre un ensemble de règles. Je me demande comment on pourra un jour savoir si l'un de vos "cerveaux électroniques" pense réellement, ou s'il se contente juste de produire des

résultats qui découlent d'un ensemble de règles, et qui donnent *l'impression* qu'il pense comme vous et moi. Peut-on imaginer une épreuve objective qui permettrait de distinguer ces deux possibilités ?"

"Comment peut-on savoir si un autre être humain pense réellement ?" répliqua Turing d'un air maussade. "Nous n'avons jamais accès à la vie intérieure de personne. Tout ce que nous pouvons faire, c'est porter un jugement basé sur le *comportement* d'une personne. Je vous dis ou je vous fais quelque chose, et vous réagissez d'une certaine manière. Je réagis ensuite à votre réponse, et ainsi de suite. Après une suite de telles interactions, je décide que vous êtes un être pensant et non un morceau de matière inanimée comme ce pichet d'eau ou le couteau à côté de mon assiette. C'est de cette manière que nous considérons que les autres êtres humains pensent comme nous."

"Comment pourriez-vous vérifier empiriquement cette forme de pensée ?" demanda Snow.

"Je pense que je proposerai l'épreuve suivante. Installons dans la pièce d'à côté un calculateur programmé pour penser comme un être humain, et un humain près de lui. Donnons leur à chacun une liaison du genre télétype avec une machine à écrire placée ici. Je vous demanderai alors de vous installer à la machine à écrire et de poursuivre une conversation par écrit, soit avec l'homme, soit avec le calculateur, mais sans que vous sachiez auquel des deux vous avez affaire."

"Quel genre de question pourrai-je poser ?" demanda Snow.

"Vous pourrez poser toutes les questions que vous voulez, déclarer tout ce que vous voulez, et de manière générale poursuivre une conversation comme nous l'avons fait ce soir."

Ah, pensa Schrödinger, Turing est en train de proposer une expérience de pensée pour illustrer ces notions. Cela lui rappela sa propre, et désormais célèbre, expérience de pensée sur le chat dans la boîte fermée, qui illustre si bien les paradoxes entourant l'acte de la mesure en mécanique quantique.

"Supposons donc que vous puissiez interagir de cette manière avec la personne — ou la machine — qui est dans la pièce voisine," continuait Turing. "Supposons que

vous y passiez plusieurs heures, à plusieurs reprises. Si à la fin de l'expérience vous n'êtes pas capable de distinguer l'homme de la machine d'une façon fiable, alors je prétends que la machine est intelligente, ou que vous, l'homme, ne l'êtes pas. Par conséquent, si vous êtes prêt à admettre que les hommes sont intelligents, je ne vois pas comment vous pourriez refuser les mêmes droits à la machine. Après tout, c'est exactement la procédure que j'emploie en ce moment pour décider que vous êtes intelligent : en observant vos réactions à ce que je dis et à ce que je fais en diverses circonstances, je conclus que vous êtes un être pensant comme moi. Ce n'est pas parce que vous avez une moustache, deux yeux, ou toute autre raison liée à votre apparence physique. C'est uniquement parce que vous agissez et réagissez d'une certaine manière que je considère comme la réaction normale des hommes intelligents dans ces circonstances."

"Vous estimez donc que cette sorte de *Jeu d'imitation* est une bonne épreuve objective pour décider si un ordinateur peut penser comme un humain. C'est bien là votre opinion ?" demanda Snow.

"Tout à fait," répondit Turing.

Schrödinger intervint : "En concentrant votre épreuve d'intelligence sur le comportement externe de la machine ou de la personne, plutôt que sur son fonctionnement interne, vous vous placez au cœur de la tradition behavioriste de la psychologie." Il poursuivit en rappelant que, s'ils avaient des opinions variées sur le rôle des "états mentaux" parmi les causes de la pensée, les behavioristes des années 20 comme J.D. Watson rejetaient fermement l'idée que les qualités internes du cerveau puissent jouer un rôle quelconque dans une théorie *scientifique* de la pensée ou du comportement.

"Selon les behavioristes, seules les actions visibles de l'extérieur peuvent former la base d'une théorie légitimement scientifique du comportement," ajouta Haldane. "L'épreuve d'intelligence de Turing me semble un simple transfert de ce paradigme de l'homme à la machine."

Tout comme les arguments des behavioristes eux-mêmes, l'épreuve de Turing fut immédiatement attaquée par toutes les personnes autour de la table.

Haldane ouvrit l'assaut contre le Jeu d'imitation en arguant que "Il me semble que l'on en apprend généralement plus sur la nature d'un autre être en le combattant qu'en lui obéissant. Mais les machines ne peuvent ressentir de plaisir, être amadouées par la flatterie, et de façon générale ne montrent aucun signe de réaction émotionnelle consciente. Il me semble à moi que la seule façon de savoir si une machine pense est *d'être* une machine."

"C'est un point de vue particulièrement solipsiste qui rendrait impossible la communication d'idées," répliqua Turing. "Si vous aviez raison, ce que je ne crois pas pour le moment, alors certes mon Jeu d'imitation ne serait pas une épreuve suffisante d'intelligence, puisque nous ne serions jamais sûr que quiconque d'autre pense à moins *d'être* ce quelqu'un d'autre. Pourtant, je suis prêt à reconnaître que vous pensez, et je soupçonne que vous êtes également prêt à admettre la même chose de moi. Je suis donc au regret de vous dire que je ne trouve pas votre argument convaincant. Le solipsisme n'est la réponse à rien."

Snow revint alors dans la conversation en disant : "Peut-être, Turing. D'après ce que vous avez dit jusqu'à maintenant du fonctionnement d'un calculateur, il me semble que la machine ne fait rien d'autre que ce qu'on lui a ordonné de faire. Elle dispose d'un jeu d'instructions qui constitue son programme, et ces instructions sont aveuglément suivies, étape par étape jusqu'à l'arrêt de la machine. Je ne vois donc pas comment la machine pourrait faire montre d'imprévisibilité, de libre arbitre, ou d'incohérence, ou des nombreux comportements que nous voyons dans la conduite humaine."

Turing répondit immédiatement : "C'est exactement l'objection que soulevait, il y a un siècle de cela, la fille de Lord Byron, Lady Ada Lovelace, quand elle travaillait avec Charles Babbage sur sa "machine analytique". Je vous ferai la réponse que Babbage a dû lui faire. Il est rarement facile de prévoir toutes les conséquences d'un fait donné. Il est en particulier très difficile de savoir quelles quantités seront calculées au cours d'un calcul mené en suivant un certain jeu de règles. Même si les règles sont simples, prises séparément, le fait de traverser des milliers ou des millions d'étapes en respectant ces règles peut facilement engendrer des quantités complètement inattendues. Une machine

qui peut apprendre, et qui est capable de modifier les instructions de son programme en fonction des nouvelles données qui se présentent, fournit une excellente illustration d'un tel mécanisme. Donc oui, sans doute la machine n'est-elle capable de faire que ce qu'on lui a demandé. Mais nous sommes nous-mêmes incapables de prévoir les conséquences de ces instructions."

Schrödinger remarqua alors : "Wittgenstein a présenté d'excellentes raisons à l'appui du caractère informel du comportement humain. Il semble qu'il soit tout simplement impossible de fournir des règles de conduite qui couvrent toutes les éventualités. Bref, il y a beaucoup plus dans la vie que le suivi d'un ensemble de règles. Je ne vois donc pas comment une machine, quelle qu'elle soit, et même si elle peut modifier ses règles, pourra jamais reproduire le comportement humain si certains aspects de ce comportement ne répondent à *aucune* règle."

Turing fut un peu étonné d'une telle objection, surtout venant d'un scientifique de la stature de Schrödinger, car il lui semblait que cela revenait à nier l'existence d'une structure ou de schémas de comportement. Répondant à Schrödinger, il dit : "La seule façon de découvrir les lois du comportement, c'est de les chercher. Mais nous ne pouvons jamais être sûrs d'avoir cherché avec assez de soin et de persévérance. Il n'est pas impossible que les comportements que nous jugeons en dehors de toute règle soient déterminés par des règles que nous n'avons pas été assez intelligents ou diligents pour découvrir."

Schrödinger contra cette riposte en rappelant les travaux de Gödel en logique mathématique, ceux que Turing avait évoqués au début du repas. "Vous nous avez dit vous-même qu'il existe des affirmations au sujet des nombres qui ne peuvent pas être démontrés ou infirmés en suivant un ensemble de règles logiques. Pourtant, nous les hommes, nous voyons que ces affirmations sont nécessairement vraies, mais nous ne pouvons pas les prouver. Cela ne signifie-t-il pas qu'il y a des choses qu'un esprit humain peut connaître et qu'une machine ne pourra jamais connaître ?"

"Les résultats de Gödel n'ont rien à voir ici," objecta Turing. "Son Théorème d'Incomplétude suppose que le système logique que l'on utilise pour prouver ou réfuter

des propositions sur les nombres est cohérent et non-contradictoire. C'est à dire qu'il n'est pas possible de prouver et d'infirmer simultanément la même proposition en utilisant les règles du système, et que nous ne commettons jamais d'erreur logique en appliquant les règles déductives. Si l'une de ces conditions n'est pas réalisée, le théorème de Gödel ne s'applique pas. Mais les hommes commettent *vraiment* des erreurs et se comportent *vraiment* de façon incohérente. Aussi une machine qui reproduirait le comportement humain devrait faire de même. Aussi je ne vois pas que le théorème de Gödel s'applique ici."

S'écartant un instant de la discussion qui devenait houleuse, Turing prit le pichet d'eau pendant que les autres discutaient des thèses qui avaient été échangées avec la plus grande ardeur. Snow relança la controverse en tentant de résumer la situation.

"Il me semble qu'au cœur des arguments en débat se trouve la question de savoir si un jeu de règles est la seule origine du comportement humain, en particulier du comportement cognitif. La conviction de Turing repose sur l'hypothèse qu'un ensemble de règles, assez vaste ou agissant assez longtemps, peut conduire à un comportement qui semble illogique, spontané, émotionnel, irrationnel ou créatif à un observateur extérieur. L'argument de Wittgenstein est qu'aucune accumulation de règles, fussent-elles longues et complexes, ne peut rendre pleinement compte de la vie cognitive de l'homme. Au mieux, une telle machine suivant des règles ne pourrait qu'imiter ou simuler une petite partie de l'expérience humaine. La seule manière de dupliquer un humain est d'en *être* un."

"L'idée d'une machine pensante est trop horrible à imaginer," s'exclama Haldane en tendant son verre pour que Snow puisse lui verser encore un peu de Montrachet. "Certains de nos collègues à l'esprit théologique, ici à Cambridge, vous diraient que la pensée est une fonction de l'âme immortelle de l'homme. Les machines n'ont pas d'âme, et par conséquent elles ne peuvent pas penser, jamais ! Qu'avez-vous à répondre à cela, Turing ?"

"Ne pensez-vous pas que cet argument écorne sérieusement l'omnipotence du Tout-Puissant ?" contra Turing. "Vous semblez défendre l'idée que les hommes sont de

quelque manière subtile supérieurs au reste de la création. Si tel est le cas, alors bien sûr nous devons tous nous résigner à ce fait, et j'abandonnerai sans regret ma vision d'une machine qui penserait comme un homme. Mais, pour autant que je sache, il n'existe aucune preuve d'une telle supériorité des humains. Par conséquent, tant que quelqu'un ne m'aura pas présenté un argument imparable à ce sujet, je suis obligé de considérer que ce genre d'objection relève d'une pensée incertaine qui prend ses désirs pour des réalités."

Pour réfuter le Jeu d'imitation, Haldane avança alors ce qui parut à chacun une possibilité hautement spéculative et peu sérieuse. "Supposez," dit-il, "que votre interrogateur dispose d'un canal de communication extrasensoriel avec l'homme ou la machine de la pièce d'à côté. Il pourrait les distinguer sans utiliser le télétype. Cela ne rendrait-il pas votre épreuve sans valeur ?"

"Grand Dieux," s'exclama Wittgenstein, "vous êtes en train d'introduire quelque chose d'encore plus hypothétique que la machine cognitive de Turing. Si nous laissons la discussion partir dans une stratosphère mystique, pourquoi ne pas imaginer une intervention divine disant à l'interrogateur qui est de l'autre côté ? Vous allez vraiment trop loin avec cette perception extrasensorielle."

Turing resta étonnamment silencieux un instant avant de répondre à l'objection de Haldane. "J'imagine qu'une communication de cette sorte invaliderait mon test d'intelligence. Tout ce que je puis dire, c'est que si vous admettez cette forme de perception, n'importe quoi peut arriver. En ce cas, le Jeu d'imitation ne permettrait plus de dire si la machine pense ou non comme un homme, puisqu'on pourrait toujours les distinguer. Mais tant que cette forme de communication ne sera pas scientifiquement établie, je continuerai à défendre le Jeu d'imitation comme la façon correcte de procéder."

Pendant cet échange, Snow réfléchissait en silence à l'argument de Wittgenstein sur l'origine sociale du langage. Il vit soudain comment cela s'accordait avec sa vieille idée que les mots sont toujours plus simples que la réalité qu'ils représentent. Si ce n'était pas le cas, se disait-il, la discussion et l'action collective seraient impossibles. Par conséquent, si Wittgenstein avait raison de dire que les mots émergent en quelque sorte de la société, il était raisonnable que cette société s'accorde sur une façon d'exprimer la

réalité brute qui soit plus simple que cette réalité elle-même. Sinon, le langage ne pourrait jamais être une forme abrégée de communication. Heureux de ce brillant aperçu de la relation entre le langage et le monde, Snow fut brutalement tiré de sa concentration par l'apparition de Simmons sur le pas de la porte, qui demandait s'il pouvait servir le plat de résistance.

"Simmons semble prêt à passer à la suite. C'est peut-être le bon moment d'interrompre un moment cette discussion et de nous accorder une brève pause avant de passer au plat de résistance," proposa Snow au groupe.

"Tout à fait," répondit Schrödinger. "Turing nous a encore donné de quoi réfléchir avec son Jeu d'imitation, puisqu'il semble directement relié au problème que Wittgenstein a posé au début du dîner sur la relation entre pensée et langage. Peut-être pourrions nous nous aventurer un peu plus loin sur ce sujet et sur son rapport avec le problème des machines pensantes. A mes yeux, en tout cas, si nous employons le test que Turing suggérerait pour déterminer si une machine pense ou non, alors je ne vois pas comment nous pouvons éviter de nous poser la question du rôle du langage dans la pensée."

Haldane ajouta : "Si je comprends bien, le Jeu d'imitation de Turing est exclusivement basé sur un échange linguistique avec celui ou celle qui est de l'autre côté. Cela suggère certainement que toute machine pensante doit disposer de capacités de langage équivalentes à celles d'un homme pour être jugée intelligente. Pourtant Wittgenstein nous affirme que ces capacités ne peuvent venir que d'une vie commune, ce qui semble exclure les machines. Il me semble qu'il y a là une contradiction directe avec la notion de machine pensant comme un homme. J'aimerais pour ma part que la discussion se focalise sur ce point, pour que nous puissions voir s'il y a une issue à ce dilemme."