Philosophica 14, 1974 (2), pp. 28–37.

RECHERCHE SUR LA RATIONALITE

A. PHALET

> "How dreadful ! " cried Lord Henry. "I can stand brute force, but brute reason is quite unbearable. There is something unfair about its use. It is hitting below the intellect."
> Oscar Wilde, The Picture of Dorian Gray.

La rationalité est un mode d'efficience de la pensée tel que, à défaut de pensée, il reste un mode d'éloquence. Sérieuse elle côtoie de préférence la cybernétique et la mathématique ensembliste. Même sous ses formes plus rustaudes il convient d'admettre son apport fondamental au caractère scientifique ou respectable d'un exposé. Elle se signale, en outre, par l'intransigeance de son vocabulaire, ce qui assure un effet plaisant sur le plan social : les adeptes du rationnel s'apprécient au son des mots ; en réalisant ainsi la solidarité des scientifiques et autres intéressés elle vise à l'unité de la science.

Il va sans dire qu'une des sources premières d'où puiser l'information sur le caractère propre de la rationalité est l'introspection : on analyse en son âme et conscience les composantes de sa propre lucidité. En présence de gens bien intentionnés la valeur scientifique d'une telle méthode est garantie par le principe démocratique que toute opinion a son apport propre et spécifique dans la grande aventure de la science contemporaine.

Mettons donc qu'une pensée soit rationnelle si la justification de son contenu peut être formalisée. Par justification nous entendons une preuve de la consistance et de l'unicité, à l'équivalence près, de cette pensée.

Précisons cette suggestion en admettant par exemple que cette pensée concerne une préférence ou un choix. Pourrions-nous dans ce cas-là supposer que notre pensée rationnelle soit une pensée qui se propose de déterminer un couple (d,v) tel que d est la distance la plus courte d'une valeur réalisée ou supposée d'être réalisée à la réalisation d'une valeur préférée v tel que notre préférence pour v dépend, entre autres, de la distance d ? Mais, dans ce cas il semble qu'on démontrera l'unicité en l'imposant; en effet, pourquoi la distance la plus courte ?

L'exigence d'une preuve d'unicité nous oblige donc d'insérer comme donnée de la justification un élément de nature générale d'où l'unicité puisse se déduire pour chaque cas particulier. On ne cesse donc point d'imposer l'unicité, mais on le fait à un niveau où l'action humaine apparaît dans son contour général. Par les têtes fortes de l'empirisme une telle donnée est mise en accusation comme une idée préconçue; nous préférons l'appeler élément constitutif d'un modèle général de l'action humaine; ainsi l'unicité se déduira de principes qui font partie de la description d'un tel modèle. Il nous faudra donc élaborer, ne fussent que les traits fondamentaux d'un mécanisme capable d'engendrer des énoncés consistants et "uniques". Ce mécanisme étant, par supposition, capable d'engendrer une multitude d'énoncés, l'énoncé unique est le résultat d'une sélection effectuée conforme aux exigences de consistance et d'unicité partant de certaines données. Nous sommes donc en présence d'un processus de solution de problèmes et notre mécanisme sera un automate capable d'un tel processus.

Ceci signifie que nous ne pouvons envisager une définition du rationnel que dans le cadre d'une pensée s'appliquant à la solution de problèmes. Cette délimitation est due à l'introduction de l'exigence d'unicité qui est plus forte et implique celle de la consistance.

Examinons les conditions qui permettent à un automate de donner une solution unique à un problème non trivial. Précisons qu'un problème est non trivial si la solution de ce problème est informative, c'est-à-dire qu'elle augmente la quantité d'information contenue dans les données du problème en question.

Supposons que tout problème non trivial puisse assumer une forme telle qu'il se résout par la transformation d'un texte infini en un texte fini équivalent au premier. Comme une donnée ne peut être que finie nous ne disposons que d'une partie de ce texte infini. La solution, c'est-à-dire le texte fini, étant équivalente au texte infini, est plus informative que cette partie finie du texte infini qui est la donnée du problème. Ensuite nous dirons qu'un automate fini A_2 est supérieur à un automate fini A_1, si A_2 est capable de résoudre un type de problèmes non triviaux qui exigeraient pour leur solution que A_1 soit composé d'une infinité d'éléments.

Considérons alors une série de formes générales de problèmes de complexité croissante. A chacune de ces formes nous ferons correspondre un automate qui répondra par une réponse unique au problème posé et tel que chaque automate A_{n+1} sera supérieur à l'automate A_n qui le précède. L'unicité de la solution signifie que le problème est soluble pour l'automate en question; en effet nous supposons que le problème n'a qu'une seule solution; une multiplicité de solutions non équivalentes impliquerait dès lors l'impuissance de l'automate à résoudre le problème posé.

Mettons que D soit un domaine qui fait l'objet de recherches en vue d'une connaissance exacte et totale de D. Cette connaissance équivaut à l'identification d'une fonction f_D qui représente D comme un système. Pour faciliter la compréhension de ce qui suit nous mentionnons le cas trivial où D est un ensemble fini, par exemple $D = \{a,b,c\}$. Il n'y a guère de différence essentielle entre D et f_D, ce dernier énumérant la liste des éléments de D.

La forme de problèmes de complexité du premier ordre est celle où D est infini et f_D une fonction récursive, notamment la fonction caractéristique de D : a est un élément de D si et seulement si $f_D(a) = 1$ et si $f_D(x) \neq 1$, $f_D(x) = 0$. Dans ce cas cette fonction désigne en fait un automate fini de l'espèce la plus simple : partant à chaque fois d'un état initial E_o il termine invariablement son action dans l'état final E_f.

La notion de connaissance associée à cet automate imperturbable est assez triviale. En effet, savoir que a est un élément de D et qu'il en est de même pour b ne permet pas de conclure que $a = b$ ou que $a \neq b$.

Supposons que les éléments de D constituent des configurations, par exemple, des séquences. Alors, si (c,a) est une telle séquence, tandis que (c,b) ne l'est point, nous pouvons dire que $a \neq b$. L'identification de a comme étant un objet différent de l'objet b dépend donc de c. Ceci peut s'exprimer de la façon suivante : $f_{Dc}(a) = 1$ tandis que $f_{Dc}(b) = 0$. Mais rien n'empêche qu'il y ait un d tel que (d,b) soit une séquence constituée par les éléments de D, tandis que (d,f) ne le soit point. Dans ce cas $f_{Dd}(b) = 1$ et $f_{Dd}(f) = 0$. Les fonctions f_{Dc}, f_{Dd}, ... sont donc des fonctions caractéristiques de différentes sous-classes de D. Les éléments c, d, ... diffèrent par supposition. A chaque supposition répond une fonction f_{Dx}, c'est-à-dire une composante de l'automate désigné par (f_D; f_{Dc} f_{Dd}, ...). En conséquence, notre automate pourra trancher pour tout a et b la question si $a = b$ ou $a \neq b$, s'il est muni d'une infinité de composantes, c'est-à-dire une composante pour chaque f_{Dx} où D est le domaine des valeurs pour la variable x.

Partons de l'automate (f_D;f_{Dd}) et supposons que, en effet, $f_{Dd}(a) = 1$ et $f_{Dd}(b) = 0$. Donc $a \neq b$. Mais que $a \neq b$ est une nouvelle donnée qui pourrait aider notre automate à faire de nouvelles distinctions s'il était doué de mémoire.

Dans le cas qui nous préoccupe un automate qui s'instruit est un automate qui, par exemple, réalise la transition de $(f_D;f_{Dd})$ à $(f_D; f_{Dd}, f_{Da}, f_{Db})$ ou à $(f_D; f_{Dd}, f_{Da})$ si d = b etc. Cet automate qui par sa mémoire augmente sa capacité de discernement est donc un automate qui passe d'un état à un autre et qui cesse de ce fait d'être une machine imperturbable : la réponse de l'automate à une même question pourra varier avec l'état de l'automate.

Représentons cet automate à mémoire qui s'instruit par $(f_D;MI)$. Nous avons donc distingué trois types d'automate : f_D, $(f_D;f_{Dx})$ et enfin $(f_D;MI)$. f_D est un automate imperturbable sans mémoire et $(f_D;MI)$ un automate à mémoire qui s'instruit, mais, jusqu'à présent, nous avons omis de caractériser le second type, notamment $(f_D;f_{Dx})$. Considérons $(f_D;f_{Dd})$. Il n'est pas exclu que $(f_D;f_{Dd})$ puisse reconnaître dans une infinité de cas que $x \neq y$, par example, $a \neq b$, $e \neq f$, etc. En premier lieu $(f_D;f_{Dd})$ devra déterminer que a et b sont des éléments de D, ensuite retenir qu'ils le sont pour pouvoir, enfin, déterminer que $a \neq b$. Le fonctionnement de $(f_D;f_{Dd})$ implique donc la présence d'une mémoire. En effet, il s'agit de comparer les résultats de certaines actions comme $f_D(a)$, $f_D(b)$ etc. Mais la machine ne s'instruit pas dans le vrai sens du terme; le contenu de sa mémoire s'efface dès qu'un nouveau cas, par exemple la comparaison de e et f se présente. Cette performance de $(f_D;f_{Dd})$ transcend la capacité d'un automate du type f_D.

Le type f_D étant sans mémoire ne peut comparer différentes entrées. Il n'y aura donc qu'une entrée, par example a; dès lors l'élément b devra figurer dans l'algorithme ainsi que d. Mais comme il y a une infinité d'éléments qui, comme b, diffèrent d'au moins un élément de D, il y aura nécessairement une infinité d'algorithmes présents dans la machine f_D, qui sera donc elle-même infinie. L'automate $(f_D;f_{Dx})$ est donc supérieur à l'automate f_D. En est-il de même pour $(f_D;MI)$ par rapport à $(f_D;f_{Dx})$?

Seul un automate infini du type $(f_D;f_{Dx})$ aura une capacité équivalent à celle de $(f_D;MI)$. Mais ce deriner est-il lui-même fini ? Il est clair que $(f_D;MI)$ sera fini à chaque moment t_n où n est un nombre naturel quelconque. Mais en principe $(f_D;MI)$ grandira sans cesse : bien qu'il reste fini il est impossible d'imposer une limite (finie) à son extension sans le rendre équivalent à un automate fini du type $(f_D;f_{Dx})$, c'est-à-dire sans le condamner à un niveau inférieur. Cette infinité potentielle se reflète dans son impuissance d'assurer pour chaque question posée du type "$x \neq y$?" une réponse unique dans un laps de temps fini.

En outre ne sachant comment répondre à la question " u $\neq$ v ? " au moment t_n, il n'y a au moment t_n non seulement pas de réponse unique, mais même pas de réponse du tout.

On remarquera qu'il est, au moins en principe, facile de remédier à cette situation en déterminant une machine probabiliste (f_D;MI,P). Dans ce cas nous obtiendrons tout au plus une probabilité égale (notamment 0,5) pour les réponses "u $\neq$ v" et "u = v".

Bien que ces deux réponses se valent lorsqu'elles ont la même probabilité 0,5, elles ne sont pas équivalentes et on ne peut garantir une réponse unique dans un espace de temps fini. D'autre part il ne suffit point que la réponse "u $\neq$ v" ait une probabilité plus grande que sa négation pour qu'elle puisse être considérée comme réponse unique. Précisément le fait que la probabilité de sa négation "u = v" n'est pas nulle signifie que "u $\neq$ v" n'est pas une réponse unique. Mais il y a un certain progrès : nous pourrions justifier notre préférence pour une des deux réponses par sa plus grande probabilité, c'est-à-dire, les réponses deviennent comparables.

Augmentons la capacité de notre automate en supposant que toute action de la machine, que nous appellerons désormais (f_D,MI,P,B), vise à résoudre un sous-problème du problème initial "u $\neq$ v ? ". Toute action de l'automate vise donc à atteindre un but. Seulement rien n'empêche que ce but se situe à l'infini, puisque D a une infinité d'éléments.

Comme notre automate s'instruit (de ses succès) il se transforme vu qu'il accumule de l'information. Lorsque chaque action de l'automate vise à atteindre un but, l'information acquise est utile à l'exception de l'information concernant la probabilité des réponses. Cette information utile est composée d'affirmations de la forme "x = y" ou "x = y" qui furent des réponses uniques. Or nous devons concevoir un automate qui s'instruit non seulement de ses succès, mais surtout de ses échecs.

Il faut se rendre à l'évidence qu'on peut toujours concevoir un domaine D qui pose des problèmes insolubles pour un automate qui ne dispose que d'un temps fini pour les résoudre. Rappelons à ce sujet que selon le théorème de Gödel toute théorie non triviale est incomplète.

Nous supposons donc que notre automate résout son problème (initial) en un laps de temps fini ou abandonne la partie après un nombre fini de tentatives de solution. Quelles sont les caractéristiques d'un tel automate (f_D;MI,P,B,A) ? Il est clair qu'une suppression arbitraire des recherches par la machine est inadmissible : elle n'abandonnera un problème qu'après avoir fait le tour complet de ses moyens d'action. Essayons de préciser le contenu de cette affirmation relativement raisonnable mais absolument vague.

Notre problème central, pour l'instant, est la transformation d'information inutilisable concernant la probabilité des réponses du type "$x \neq y$", "$x = y$" en information utile et "vraie" ou nuisible et "fausse" — nous nous bornerons à ne considérer que ces deux possibilités. Supposons que cette transformation soit réalisée par une fonction F, c'est-à-dire que ou bien $F(r,p(r)) = 1$, ou bien $F(r,p(r)) = 0$; le symbole $p(r)$ représente la probabilité de la réponse r. Mais l'introduction de F entraîne l'introduction de la possibilité de contradiction : la fonction F peut se tromper. Raisonnons maintenant par "anthropomorphisme". Considérons le domaine infini D comme une machine M_D aux entrées r, r', ..., sorties 1,0 et fonction de transition F_D. Vu le mode d'action de notre automate (f_D;MI,P,B,A) une sortie de M_D dépend en général de plusieurs entrées et sorties antérieures. Ceci signifie que M_D est une machine à "mémoire" et que par conséquent les sorties ne dépendent non seulement d'une entrée mais aussi de l'état de la machine; les états e, e', ... représentent la part de détermination d'une sortie par des sorties ou entrées antérieures et opèrent donc comme mémoire.

La tâche de l'automate (f_D;MI,P,B,A) consiste en une identification de M_D, c'est-à-dire en une description (finie) de la fonction de transfert F_D de M_D. Mais cela ne suffit pas. En effet, comme il nous faut atteindre le résultat : "$F(r,p(r)) = 1$ au moment t_n si et seulement si e est l'état de M_D au moment t_n et $F_D(r,e) = 1$", on doit en outre procéder à l'identification de l'état "actuel" dit "initial" de M_D; car il se pourrait que $F_D(r,e') = 0$ ou que $F_D(r,e')$ ne soit pas défini. Exprimons notre conjecture concernant l'état initial de M_D en donnant une certaine valeur à un paramètre a : $F(r,p(r),a)$. Si les probabilités $p(r)$ — acceptons sans plus que cette probabilité possède une interprétation adéquate dans ce système — nous permettent d'aborder d'une façon efficace l'identification de M_D, il suffira dans les cas les plus simples de ne faire varier que les valeurs du paramètre a dans l'expression $F(r,p(r),a)$. Mais, si $F_D(r,e) = 1$ et $F_D(r,e') = 0$, alors il y a une discontinuité, une perte de mêmoire ou d'information qui contraint (f_D;MI,P,B,A) à la contradiction, puisque dès lors $F(r,p(r),a) = 1 = 0$, et oblige de ce fait notre automate (f_D;MI,P,B,A) à revoir ses acceptations concernant les caratéristiques fondamentales du domaine D et donc du système M_D.

Mettons que D soit un système ouvert; dans ce cas la fonction $F(r,p(r),a)$ perd son efficacité et il faudra la remplacer par une autre F'. Notre automate (f_D;MI,P,B,A) devra donc organiser ses propres moyens d'action. Une telle machine est un automate à auto-organisation.

Nous sommes partis de la considération d'un domaine D en nous imposant la tâche de spécifier une fonction f_D représentant une connaissance exacte et totale de D. Or nous voici en présence d'un automate "sujet", notamment la machine à auto-organisation (f_D;MI,P,B,A), et d'un automate M_D "objet". Ce qu'on peut en fait considérer comme un dédoublement de f_D est dû originairement à l'introduction des séquences (c,a), (d,b) ... qui expriment une dépendance entre les éléments c et a, d et b, ... en vue de l'identification de a,b,... . Nous introduisons donc une (ou des) relation(s) qui sont des moyens de connaissance. Dès cet instant nous inaugurons la conception d'un automate connaissant; en effet, un nombre infini de telles relations nous oblige de faire la transition à un automate qui s'instruit : nous ne pouvons concevoir la résistance d'un domaine aux efforts de connaissance que d'après les formes que prennent ces efforts pour parer à cette résistance. La connaissance d'un domaine exige donc la connaissance de la forme de ces efforts et, partant de là, une élaboration des formes de résistance, ce qui vient d'aboutir plus haut à la conception de M_D.

Après la réalisation de ce dédoublement menant d'une part à (f_D;MI,P,B,A) et d'autre part à M_D, le premier doit pouvoir poser le problème de la correspondance — dans la théorie des systèmes ce problème s'énonce au niveau formel comme le problème de la détermination de l'état initial d'un automate — entre son propre modèle (F, F',...) de D et le modèle de D au niveau d'une machine apte à représenter le comportement d'une classe d'éléments. Ce dédoublement repose donc sur la possibilité de représenter une relation, d'une part, du point de vue de l'extension, comme une classe, et d'autre part au niveau de l'intention où elle s'exprime au moyen de la notion de but, notamment l'identification de a, b, ...

Essayons de reprendre le thème de l'unicité avant qu'il ne s'égare dans le fin fond de la métaphysique. L'obstacle tenace s'opposant à une caractérisation générale des moyens d'action réalisant une solution unique à un problème non trivial comme nous l'avons posé, est l'infini potentiel, qui reflète dans l'automate l'infini actuel du domaine D. L'introduction de la possibilité pour l'automate d'arrêter ses efforts dans une direction pour les porter dans une autre ou de mettre fin à ses recherches équivaut, comme nous l'avons vu, à la conception de cet automate comme machine à auto-organisation. L'infini potentiel ne peut être vaincu que par la possibilité d'un abandon justifié.

Or le problème général de la résolution du problème de l'infini potentiel nous semble être inabordable au niveau des fonctions F, F',

...,c'est-à-dire au niveau des types de systèmes. Ceci signifie qu'il est impossible de concevoir d'une manière générale une forme de justification pour un abandon (par l'automate d'un problème) décidé après un nombre fini d'opérations finies et que cette impossibilité est due au fait que nous ne disposons pas des moyens conceptuels pour aborder le problème d'une telle justification sur le plan des types de systèmes.

Posons le problème au niveau du paramètre a. Alors la résolution de ce problème impliquerait la possibilité de déterminer une fonction K(x,a) qui atteindrait son extremum, disons son maximum, pour une certaine valeur de a ou pour aucune valeur de a, lorsqu'on considère une certaine classe de valeurs pour x. Dans le second cas un changement de structure s'impose, car l'extrémum indique que l'action de l'automate est optimale, ce qui signifie qu'à un certain moment la fonction F(r,p(r),a) ne se trompe plus. Si celà ne sera jamais le cas pour aucun a, il faut la remplacer par une autre fonction. Ou bien il n'y a qu'un nombre fini de fonctions F, F', ... et alors le problème de savoir si l'automate doit abandonner ou continuer la recherche d'une solution est résolu s'il est résolu au niveau du paramètre a. Mais ce cas implique qu'il serait possible d'acquérir une connaissance complète des structures de la connaissance, ce qui est exclu. Ou bien il y a une infinité de fonctions F, F', ... C onsidérons dès lors une fonction K'(x',F) où F est une variable dont {F, F', ...} est le domaine des valeurs. Posons le problème de l'extrémum de K'(x',F) partant de F et d'une classe de valeurs pour x'. Seulement la recherche d'un extrémum est basée sur la considération d'accroissements $\triangle$ de l'argument de la fonction ce qui implique la nécessité pour le domaine de définition de la fonction considérée, d'être ordonné. Or comment ordonner le domaine {F, F', ...}, c'est-à-dire, puisque ces fonctions correspondent à et peuvent même être considérées comme l'expression de types de systèmes (abstraits), comment ordonner des structures ? Ainsi la résolution du problème de l'infini potentiel dépend de la possibilité d'ordonner un ensemble de structures (sans faire appel à l'axiome du choix); pour l'instant ceci nous semble être plutôt une impossibilité; ainsi, nous espérons qu'on comprendra mieux notre conclusion anticipée que nous ne disposons point des moyens conceptuels capables de forcer une solution du problème.

Mais alors, que ce passe-t-il en réalité ? On peut supposer que chacun choisit ou accepte un domaine fini de fonctions F, F', ... plus ou moins bien déterminé et ordonne ce domaine d'une manière plus ou moins définitive. L'explicitation et la justification de cet ordre s'appellera "philosophie". Partant d'ici nous rencontrons les grands problèmes sérieux : qu'est-ce qu'une philosophie rationnelle ? qu'est-ce que la rationalité en philosophie ? , et passons outre.

Notre problème se réduit donc, en fin de compte, au problème de la transition de cette "philosophie" à une théorie générale des systèmes dont le problème fondamental serait la détermination de la notion d'ordre pour des ensembles de structures.

Revenons à notre problème initial. Quel est au niveau d'une caractérisation générale de l'action humaine l'élément d'où l'unicité se déduit et sur lequel repose, par conséquent, la justification au sens formel d'une affirmation, c'est-à-dire d'une pensée ou d'une action ? Comme nous l'avons vu, la condition nécessaire en vue d'une identification de cet élément est la transition menant de la "philosophie" (de la connaissance) à une théorie formelle des systèmes jouant le rôle d'une mathématique dont le domaine d'individus est un domaine de structures. Cet élément, s'il existe, garantira la suppression du problème de l'infini potentiel, l'obstacle fondamental à la réalisation de solutions uniques pour problèmes non triviaux, auquel tend la conception même de l'auto-organisation. Nous proposons donc comme forme générale d'une preuve de (quasi-) unicité d'une affirmation P le modèle suivant :

> si la conjonction de non-P et M_A est inconsistante, alors P est (quasi) unique et donc (quasi-) rationnel, où M_A est la description d'un automate à auto-organisation engendrant ou réalisant P.

Il nous semble que cette suggestion est applicable et donne un résultat positif dans le cas particulier que nous avons proposé au début concernant la rationalité d'une préférence ou d'un choix.

Un examen approfondi du critère que nous venons de proposer pour le rationnel ne correspondrait plus au caractère introductif de cet exposé.

On pourrait, certes, proposer des critères moins formels — cette transformation de la "philosophie" en théorie générale des systèmes n'est pas obligatoirement une solution unique, qui en donnerait la preuve formelle, ni même une solution certainement réalisable — qui posent une philosophie élaborée, voire une idéologie comme point de départ. Alors une pensée ou un acte sera rationnel par rapport à cette philosophie ou idéologie, seulement dans ce cas là l'unicité n'implique plus nécessairement l'efficacité. Et, en effet, proposer une philosophie ou idéologie comme la seule possible est un acte témoignant d'un penchant insolite pour l'arbitraire ou l'irréel.

Enfin, en conclusion, remarquons que l'exigence d'unicité, qui est le thème central de cet exposé, nécessite la conception d'une machine qui pense d'une manière efficace, ce qui forme le trait distinctif de la rationalité; par consequent l'unicité est le caractère distinctif fondamental de la rationalité.

Rijksuniversiteit Gent