

| | |
|---------------------|---|
| Zeitschrift: | Revue de Théologie et de Philosophie |
| Herausgeber: | Revue de Théologie et de Philosophie |
| Band: | 45 (1995) |
| Heft: | 1 |
| Artikel: | Société romande de philosophie : séance du 7 mai 1994 : suivre une règle : Wittgenstein et les sciences cognitive |
| Autor: | Pinkas, Daniel |
| DOI: | https://doi.org/10.5169/seals-381527 |

Nutzungsbedingungen

Die ETH-Bibliothek ist die Anbieterin der digitalisierten Zeitschriften auf E-Periodica. Sie besitzt keine Urheberrechte an den Zeitschriften und ist nicht verantwortlich für deren Inhalte. Die Rechte liegen in der Regel bei den Herausgebern beziehungsweise den externen Rechteinhabern. Das Veröffentlichen von Bildern in Print- und Online-Publikationen sowie auf Social Media-Kanälen oder Webseiten ist nur mit vorheriger Genehmigung der Rechteinhaber erlaubt. [Mehr erfahren](#)

Conditions d'utilisation

L'ETH Library est le fournisseur des revues numérisées. Elle ne détient aucun droit d'auteur sur les revues et n'est pas responsable de leur contenu. En règle générale, les droits sont détenus par les éditeurs ou les détenteurs de droits externes. La reproduction d'images dans des publications imprimées ou en ligne ainsi que sur des canaux de médias sociaux ou des sites web n'est autorisée qu'avec l'accord préalable des détenteurs des droits. [En savoir plus](#)

Terms of use

The ETH Library is the provider of the digitised journals. It does not own any copyrights to the journals and is not responsible for their content. The rights usually lie with the publishers or the external rights holders. Publishing images in print and online publications, as well as on social media channels or websites, is only permitted with the prior consent of the rights holders. [Find out more](#)

Download PDF: 11.01.2026

ETH-Bibliothek Zürich, E-Periodica, <https://www.e-periodica.ch>

Société romande de philosophie

Séance du 7 mai 1994

SUIVRE UNE RÈGLE: WITTGENSTEIN ET LES SCIENCES COGNITIVES

DANIEL PINKAS

Resumé

Les sciences cognitives devraient-elles tenir compte des remarques de Wittgenstein sur la notion de règle ? J'aborde cette question par le biais d'un examen du paradoxe sceptique que Kripke attribue à Wittgenstein. Les principales solutions à ce paradoxe sont passées en revue et critiquées. Une attention particulière est accordée aux solutions de type télologique et computationnaliste que l'on peut tirer de postulats en vigueur dans les sciences cognitives. Après avoir contesté l'interprétation de Kripke, je m'efforce de caractériser la conception de la «force de la règle» que Wittgenstein propose. Cette conception a des conséquences antithéoriques.

Un des problèmes centraux dont traitent les remarques de Wittgenstein sur ce qu'est suivre une règle est celui que pose «l'asymétrie entre l'enseignement et l'apprentissage»¹: si l'apprentissage doit se faire nécessairement sur la base d'exemples et d'exercices limités dans le temps et quant au nombre, la compréhension à laquelle aboutit l'enseignement s'étend ou paraît s'étendre *au-delà* de tout exemple et de tout exercice². On pourrait s'attendre à ce qu'il existe une meilleure explication de la capacité acquise par l'élève lorsqu'il a compris ce que le maître attendait de lui et qu'il devient capable de prendre le relais que celle qui se réduit à la constatation, métaphorique et peu informative, que l'enseignement «donné» par l'un, a été «reçu» par l'autre. Ce problème est précisément celui que Chomsky a nommé «le problème de la pauvreté du stimulus» ou encore «le problème de Platon», dont la prise en considération est ce qui caractérise, selon lui, au niveau le plus général, la «nouvelle approche» en grammaire, à savoir: comment rendre compte «de la richesse, de

¹ S. CAVELL, *The Claim of Reason*, Oxford, Oxford University Press, 1979, p. 112. Sauf mention contraire, les traductions ont été effectuées par mes soins.

² Cf. L. WITTGENSTEIN, *Philosophische Untersuchungen*, § 209.

la complexité et de la spécificité des connaissances partagées, compte tenu du fait que les données disponibles sont limitées»³

Le problème de l'asymétrie entre enseignement et apprentissage se pose évidemment de manière particulièrement aiguë à propos de l'acquisition d'une première langue par l'enfant, étant donné que la tâche à accomplir paraît, dans ce cas, éminemment disproportionnée aux moyens dont celui-ci dispose. Ce que l'enfant apprend est, en un certain sens, nécessairement quelque chose de fini mais aussi, en un autre sens, quelque chose d'illimité (la capacité de comprendre et de «générer» un nombre indéfini de phrases de la langue). Dès lors, la forme générale de la solution proposée au «problème de Platon» tel qu'il se présente dans le domaine de l'acquisition du langage est que le locuteur a acquis ou avait déjà, une appréhension implicite de divers principes constructionnels ou règles récursives du langage dont la connaissance explique la capacité problématique de créativité linguistique.

Une question que l'on peut se poser est de savoir si ce genre de solution au «problème de Platon» ne coïncide pas avec l'une ou l'autre des pseudo-solutions que Wittgenstein a considérées comme intrinsèquement incapables, en raison de leur vacuité et de leur caractère parasitaire par rapport à une simple description de faits incontestés, d'expliquer ou de justifier notre certitude et notre confiance que nos pratiques (linguistiques, arithmétiques, sociales) n'iront pas à la dérive sur des points centraux et élémentaires. En cas de réponse affirmative, il est vraisemblable que les notions de règle, de compétence et de connaissance tacite employées dans le «mentalisme sophistique» que l'on associe aux travaux de Chomsky et de Fodor devraient être, comme le pense Kripke, «sérieusement réexaminées»⁴.

Quant à l'incidence des considérations de Wittgenstein au sujet des règles sur d'autres domaines des sciences cognitives, l'hétérogénéité des théories que ce terme recouvre suffit à rendre tout verdict général dangereux et, à l'heure qu'il est, prématuré. Jusqu'à une date relativement récente, la plupart des théoriciens de la signification, des linguistes et des psychologues se désintéressaient de ce que Wittgenstein pouvait éventuellement avoir à dire sur les présuppositions conceptuelles de leurs disciplines respectives, sous prétexte que ses réflexions avaient été suscitées par des conceptions (behavioristes, Gestaltistes) tenues aujourd'hui pour dépassées. La situation, notamment depuis la publication du *Wittgenstein on Rules and Private Language* de Kripke, semble avoir évolué⁵. Faut-il en conclure que les tenants de

³ N. CHOMSKY, *Knowledge of Language*, New York, Praeger Publishers, 1986, p. 7.

⁴ S. Kripke, *Wittgenstein on Rules and Private Language*, Cambridge, Mass., Harvard University Press, 1982, note 22, p. 30-31.

⁵ Plusieurs indices, en tout cas, le suggèrent. Chomsky, par exemple, a reconnu l'importance des «problèmes wittgensteiniens», tout en soutenant que le paradoxe sceptique relatif aux règles que Kripke attribue à Wittgenstein admet une «solution directe» (cf. *op. cit.*, p. 223). Par ailleurs, les chercheurs en Intelligence Artificielle se

l'«approche scientifique» ont désormais assimilé les arguments de Wittgenstein, mais n'ont pas été convaincus (puisque'ils persistent dans leurs efforts pour résoudre empiriquement des problèmes dont l'existence même aurait vraisemblablement été considérée par Wittgenstein comme le fruit de confusions conceptuelles)? Ou bien l'étude des *Recherches philosophiques* recèle-t-elle, comme le suggère Crispin Wright, des «bombes philosophiques» qui n'ont pas encore été déclenchées, simplement en raison du caractère insuffisamment profond de nos excavations?⁶

Il est certain, en tout cas, que lorsque l'on voit un cognitiviste de la première heure comme George Miller avouer candidement qu'«en dépit de notre recours de plus en plus fréquent à des règles en tant qu'explications de la pensée et du comportement, [il] ne connaît aucun compte rendu clair de ce que sont les règles et de comment elles fonctionnent»⁷, on ne peut manquer de se dire que la confrontation avec la pensée de Wittgenstein risque bien de constituer, pour ces disciplines, une épreuve déterminante.

1. *Le paradoxe de Kripke-Wittgenstein*

Ce que Kripke a nommé le «paradoxe de Wittgenstein» constitue une radicalisation et une extension du «problème de Platon»: même en supposant un enseignement aussi explicite et pointilleux, et un objet d'enseignement aussi élémentaire qu'on voudra, la question de la conformité *réelle* de l'application singulière d'une règle par rapport à «ce que la règle dit de faire» paraît toujours pouvoir se poser (théoriquement sinon en pratique). En effet, différentes interprétations de ce que signifie «faire la même chose» dans un cas particulier sont toujours en principe compatibles avec différentes formulations de la règle. Aucune formulation d'une règle n'exclut la possibilité d'une mauvaise compréhension de celle-ci; et essayer d'exclure cette possibilité en s'appuyant sur une règle pour l'interprétation de la règle nous entraîne dans une régression à l'infini, car ce qui est vrai de la règle à interpréter l'est aussi de la règle qui interprète⁸. Comme l'écrit Wittgenstein dans sa *Grammaire philosophique*:

sont attelés à la résolution du «Problème du Cadre» («frame problem») et les constructeurs de théories du contenu des représentations internes à celle du «Problème de la disjonction», qui sont des problèmes présentant un «air de famille» frappant avec le paradoxe sceptique de Kripke-Wittgenstein.

⁶ C. WRIGHT, «Critical Notice of C. McGinn's "Wittgenstein on Meaning"», *Mind*, vol. XCIII, 1985, p. 298.

⁷ Cité par M.D. ROOT, «Language, Rules and Complex Behavior» in *Minnesota Studies in the Philosophy of Science*, VII, 1975, p. 322.

⁸ Cf. L. WITTGENSTEIN, *op. cit.*, § 86.

Et pourtant une interprétation est bien quelque chose qui est donné dans le signe, c'est *cette* interprétation par opposition à une autre (qui est différente). Et si on disait: toute proposition demande une interprétation, cela signifierait: sans additif, une proposition ne peut être comprise. Naturellement, il arrive que *j'interprète* des signes [...] mais pas toutes les fois que je comprends un signe⁹.

Dans son exposé du «paradoxe», Kripke s'appuie sur le début du paragraphe 201 des *Recherches Philosophiques*: «ceci était notre paradoxe: aucune manière de faire ne peut être déterminée par une règle car toute manière de faire peut être mise en accord avec la règle». Comme Wittgenstein, il développe le problème, qui est un problème général concernant toute signification et toute formation de concept, à l'aide d'un exemple mathématique. Il part de ce qu'on pourrait appeler la représentation standard de ce en quoi consiste le fait de savoir additionner: il existe une fonction (l'addition) définie pour tout couple d'entiers positifs et dénotée par le signe «+» ; savoir additionner, c'est avoir compris, après un nombre fini d'applications, la règle qui détermine la réponse à un nombre indéfini d'additions nouvelles. La compréhension qui s'est mise en place à un moment du passé sur la base des cas rencontrés règle un nombre indéfini de cas encore non rencontrés.

Puisque je n'ai effectué qu'un nombre fini d'additions, il en existe nécessairement que je n'ai encore jamais effectuées (Kripke choisit arbitrairement l'exemple «68 + 57»). J'effectue donc cette addition et j'obtiens le résultat correct («125»). Je soumets ce résultat à un sceptique qui suggère qu'étant donné la manière dont j'ai employé le signe «+» par le passé, j'aurais tout aussi bien pu répondre «5». Qu'est-ce qui me rend si sûr, demande-t-il, que le signe «+» tel que je l'ai compris et employé par le passé dénote la fonction qui donne la valeur 125 lorsqu'on l'applique aux arguments 68 et 57? Par hypothèse, je n'ai pu m'indiquer à moi-même que 125 était le résultat de cette addition particulière. Peut-être ai-je employé le signe «+» dans le passé avec l'intention de dénoter non pas la fonction «plus», mais la fonction «quus» (la «quaddition») symbolisée par \oplus , et définie ainsi:

$$\begin{aligned} x \oplus y &= x + y, \text{ si } x, y \text{ plus petit que } 57 \\ &= 5, \text{ dans tous les autres cas.} \end{aligned}$$

Le sceptique concédera que son hypothèse est parfaitement bizarre, totalement dénuée de plausibilité, etc.; mais il nous met au défi de citer un fait quelconque se rapportant à mon comportement, mes états mentaux ou mes intentions passés, qui la réfute et justifie la réponse «125» pour «68 + 57». Tous les faits de cette espèce ne sont-ils pas compatibles avec l'hypothèse que j'ai toujours voulu dire la «quaddition» lorsque j'employais le signe «+» et

⁹ L. WITTGENSTEIN, *Grammaire philosophique*, tr. fr. M.-A. Lescouret, Paris, Gallimard, 1980, p. 55-56.

que, lorsque j'ai répondu «125», j'ai brusquement et inexplicablement modifié mon usage antérieur ? Le problème est que «quoi que je fasse, cela est, sous quelque interprétation, en accord avec la règle»¹⁰.

La stratégie argumentative générale que Kripke trouve – ou croit trouver – chez Wittgenstein est la suivante: la conclusion, c'est qu'il n'existe pas de *faits* (pas de ‘fact of the matter’) me concernant, qui distingue le cas où, en employant le signe «+», je veux dire une certaine fonction bien définie, du cas où je ne veux rien dire du tout. L'argument en faveur de cette thèse radicale procède par élimination: tous les types de faits potentiellement pertinents pour la détermination de ce en quoi consiste l'application correcte d'une règle sont passés en revue et écartés en faisant état de l'absence des caractéristiques qui seraient nécessaires à cet effet. Il s'agit, pour l'essentiel, de faits relatifs à l'usage passé de la règle, de faits relatifs aux expériences mentales qualitatives accompagnant l'application de la règle, et de faits relatifs aux dispositions qu'a le sujet à appliquer la règle de telle ou telle façon.

Kripke relève qu'une réaction fréquente au paradoxe est de tenir son surgissement pour une conséquence d'une conception inépte de ce en quoi consiste la compréhension que j'ai acquise lorsque j'ai appris à additionner¹¹. Celle-ci n'est pas adéquatement représentée par une liste finie d'exemples suivie d'une clause d'extrapolation. En fait, on a envie de dire que j'ai appris à additionner lorsque j'ai intériorisé *une* règle ou *un* algorithme pour l'addition qui sont clairement incompatibles avec l'hypothèse du sceptique (par exemple, la procédure du dénombrement: pour additionner x et y , prenez un tas de billes; comptez x billes et réservez-les; comptez y billes et faites de même; reunissez les deux tas et comptez le nombre total de billes; le résultat est $x + y$). Cependant, observe Kripke,

en dépit de la plausibilité initiale de cette objection, la réponse du sceptique n'est que trop évidente. Certes, si mon usage du mot ‘compter’, dans le passé, se référail à l'action de compter [...] alors ‘plus’ doit avoir dénoté l'addition. Mais à l'instar de ‘plus’, je n'ai appliqué ‘compter’ qu'à un nombre fini de cas par le passé. Le sceptique peut donc mettre en doute mon interprétation présente de mon usage passé de ‘compter’, comme il l'avait fait pour ‘plus’. En particulier, il peut soutenir que par ‘compter’ j'ai précédemment voulu dire ‘quompter’, où quompter un tas est la même chose que le compter au sens ordinaire, à moins que le tas ait été formé en réunissant deux tas dont l'un comporte 57 billes ou davantage, auquel cas il faut répondre automatiquement ‘5’ [...]. Le point est parfaitement général: si ‘plus’ est expliqué en termes de ‘compter’, une interprétation non standard de ce dernier terme donnera une interprétation non standard du premier¹².

Le paradoxe sceptique de Kripke-Wittgenstein ou, comme dit Putnam, de «Kripkenstein», met en évidence une sorte de *faiblesse* de la règle. Il provoque

¹⁰ L. WITTGENSTEIN, *Philosophische Untersuchungen*, § 198.

¹¹ S. Kripke, *op. cit.*, p. 15-16.

¹² *Ibid.*, p. 17.

un certain vertige intellectuel induit par l'impression que l'uniformité des pratiques devrait reposer sur quelque réalité plus tangible que la «congruence des subjectivités»¹³ à laquelle paraît se réduire la participation à une même «forme de vie».

Dans cette «faiblesse de la règle», on pourrait encore voir, avec Wright, le reflet d'une caractéristique à la fois remarquable et troublante des concepts psychologiques auxquels Wittgenstein s'intéresse prioritairement dans les *Recherches philosophiques* et dans d'autres écrits tardifs sur la philosophie de la psychologie¹⁴. Wright observe que des concepts comme ceux de signification, compréhension, intention, attente, etc. semblent vaciller entre deux paradigmes. D'un côté, on trouve les épisodes ou processus conscients qui ont une «épistémologie à la première personne»: une douleur, un chatouillement, l'expérience perceptive d'une tache de couleur, etc., toutes choses qui ont un début et une fin datables, une «durée authentique», et dont l'occurrence puise à peine ou pas du tout dans les ressources conceptuelles du sujet. De l'autre côté, on trouve ce qu'on appelle des «qualités de caractère» (patience, courage, vanité, etc.) à propos desquelles un compte rendu comportemental et dispositionnel paraît tout à fait adéquat: il ne semble pas y avoir de meilleure façon d'établir la possession ou l'absence de ces qualités, chez autrui comme chez soi-même, que d'examiner ce que l'on est enclin à faire et à dire.

La conception cartésienne du mental, explique Wright, tend à tout tirer vers le pôle des épisodes conscients; un Ryle, au contraire, et par surcompensation, tend à ratisser le plus largement possible à partir des qualités dispositionnelles de caractère. Et la difficulté soulevée par les concepts qui intéressent Wittgenstein est qu'ils semblent participer à la fois des deux: ils ne sont pas dénués d'une forme d'épistémologie à la première personne car, à moins de trop concéder, dès le départ, au sceptique, il est difficile de nier que le sujet possède normalement une forme d'autorité épistémique à l'égard de ses propres intentions sémantiques. Mais, par ailleurs, aucune occurrence introspectivement accessible ne semble posséder les caractéristiques qui nous autoriseraient à dire que l'occurrence *est* telle ou telle intention sémantique. Cependant, si l'on essaie de concevoir ces concepts selon le paradigme des qualités dispositionnelles de caractère, on se heurte à une autre difficulté: un compte rendu dispositionnaliste laisse entier le cœur du problème, à savoir celui de la normativité, sans éclairer non plus véritablement le phénomène de connaissance non inférentielle, à la première personne, des significations, règles, ou intentions, présentes ou passées¹⁵.

¹³ J. McDowell, «Non-cognitivism and Rule-Following» in S.H. HOLTZMANN, C.M. LEICH (eds), *Wittgenstein, to Follow a Rule*, London, Routledge and Kegan Paul, 1981, p. 149.

¹⁴ C. WRIGHT, «Wittgenstein's Rule-following Considerations and the Central Project of Theoretical Linguistics», in A. George (ed.), *Reflections on Chomsky*, Oxford, Blackwell, 1989, p. 237-239.

¹⁵ Cf. *ibid.*, p. 236.

2. *Sur quelques pseudo-solutions*

Lorsque, prenant conscience de la faiblesse possible de la règle, il nous semble que le sol de l'objectivité se dérobe sous nos pieds, nous avons tendance, selon McDowell, à nous rabattre sur l'une ou l'autre version de ce qu'il appelle, suivant une suggestion de Wittgenstein¹⁶, «l'image mythique des règles comme rails»:

Ce qui compte comme consistant à faire la même chose, à l'intérieur d' [une] pratique, est fixé par ses règles. Les règles tracent des rails suivant lesquels l'activité correcte à l'intérieur de la pratique doit se dérouler. Ces rails sont là de toute manière, indépendamment de la propension à avoir certaines réponses et réactions, que l'on acquiert lorsqu'on apprend la pratique elle-même [...]. Acquérir la maîtrise de la pratique est représenté comme une chose analogue à une opération consistant à engager des rouages mentaux sur ces rails qui existent objectivement¹⁷.

Cette image est, bien sûr, beaucoup trop grossière pour être par elle-même vraiment séduisante; mais ce que McDowell suggère, c'est qu'en adoptant le point de vue de Wittgenstein, on s'aperçoit que des présentations beaucoup plus sophistiquées et apparemment plus vraisemblables de ce en quoi consiste la maîtrise d'une pratique ne diffèrent pas essentiellement de cette image et sont donc, en un sens, tout aussi primitives.

L'image des règles comme «rails» tend à s'imposer dans deux versions: platonisante et computationnelle. Dans la veine platonisante, une tentation particulièrement tenace est de dire que dans l'obéissance à la règle, c'est la *signification* de la règle qui, une fois appréhendée par celui qui doit la suivre, dicte ou guide le comportement. La signification de la règle semble contenir en germe et régler par avance les applications futures mais, demande Wittgenstein, «même si quelque chose de ce genre était vraiment le cas, en quoi cela m'aiderait-il»¹⁸ Car même si, en appréhendant un universel, l'esprit avait intériorisé les éléments et les préconditions de toute application future, de sorte qu'on pourrait dire qu'il a, d'une certaine manière, franchi tous les pas par anticipation¹⁹ et que dans l'application concrète il ne reste plus qu'à copier cette succession de pas déjà accomplis mentalement (ou «tacitement»), le problème resterait entier puisque non résolu pour l'application de l'instruction «copiez X!» (il pourrait y avoir différentes façons de copier). S'il s'agit

¹⁶ Cf. L. WITTGENSTEIN, *Remarques sur les fondements des mathématiques*, I, 116.

¹⁷ J. McDOWELL, *art. cit.*, p. 145-146. Une des cibles des paragraphes 143-242 des *Recherches philosophiques* est précisément l'idée que tout énoncé décidable possède par définition une valeur de vérité prédéterminée indépendamment de toute enquête possible. Crispin Wright nomme cette conception de la correction «investigation-independence».

¹⁸ L. WITTGENSTEIN, *Philosophische Untersuchungen*, § 258.

¹⁹ *Ibid.*, § 188.

d'*expliquer* – et non simplement de reformuler – la capacité qu'ont les êtres humains d'utiliser les règles de façon concordante dans la pratique, la solution envisagée paraît effectivement dépourvue de mérites.

Si «appréhender une signification» ressemble à «imaginer une certaine formule», le fait d'invoquer cette appréhension ne nous fera pas avancer d'un pouce, car nous pouvons avoir une formule présente à l'esprit sans savoir ce qu'elle signifie; dans un tel cas, par conséquent, il faudrait interpréter la règle – mais alors nous retombons sur le «paradoxe»: toute réponse peut être mise en accord avec la règle sous une certaine interprétation. Pour que l'interprétation me soit utile, il faut qu'elle soit *correcte*. Et il est pour le moins gênant d'avoir à invoquer à ce stade la notion de correction. Mais si l'on conçoit la signification non plus comme une formule interprétable que l'on convoquerait dans l'imagination, mais comme un objet, une propriété ou un processus non interprété qui représente pourtant *in extenso* le contenu de la règle, on ne sera guère plus avancé: dire que la signification (dans ce sens) détermine ses applications reviendrait à peu près à dire, tautologiquement, que c'est le contenu de la règle qui en détermine le contenu.

La solution au «paradoxe» que l'on peut tirer du point de vue «computationnel» consiste à faire remarquer que la raison pour laquelle je ne suis pas libre de donner une réponse comme «5» au problème de savoir combien font $68 + 57$, est que la règle de l'addition peut être «réalisée» dans une machine qui calcule la fonction correspondant à cette opération; la réponse que donne ou donnerait la «réalisation» mécanique est alors la réponse correcte, c'est-à-dire celle qui s'accorde à l'intention que j'ai lorsque j'emploie le signe «+».

Kripke considère cette solution comme une simple variante de la tentative de répondre au défi du sceptique par une analyse dispositionnelle du fait que je veuille dire «plus» plutôt que «quus²⁰». Selon une telle analyse, ce qui distingue le fait que dans le passé j'ai voulu dire l'addition plutôt que la quaddition, ce n'est certes pas un état mental occurrent ou un comportement contemporain de cette intention sémantique. Vouloir dire «plus», c'est avoir une certaine disposition à citer la somme de x et y quand on demande combien font « $x+y$ ». Il y a donc bien un fait – mais un fait dispositionnel – qui est constitutif de mes intentions sémantiques: dire que, dans le passé, lorsque j'employais le signe «+», je faisais référence à l'addition, revient à dire que si on m'avait demandé alors combien font $x+y$ (où « $x+y$ » est une addition encore jamais rencontrée), j'aurais répondu en citant la somme correcte de $x+y$.

Kripke explique que ce compte rendu dispositionnel est inadéquat. Il l'est, en premier lieu, parce que si la perplexité porte sur ce qui *justifie* la réponse «125» à la question «combien font $68+57$?», c'est une condition minimale

²⁰ S. Kripke, *op. cit.*, p. 32.

pour toute solution possible qu'elle permette de sauvegarder la différence entre performance et correction, entre la réponse que l'on donne et celle que l'on doit donner (sans quoi, comme le répète inlassablement Wittgenstein «tout ce qui paraîtra correct le sera»²¹). Or dans le point de vue dispositionnaliste les deux choses paraissent confondues. En effet, si l'on doit pouvoir dire, sur la seule base de mes dispositions, quelle est la fonction à laquelle je faisais référence lorsque j'employais le signe «+», on ne peut écarter sans circularité des dispositions à commettre des erreurs, qui existent également.

Cette objection concerne, bien sûr, la «normativité» des notions de signification et de règle. Que je veuille dire quelque chose de bien précis en employant une expression ne garantit pas que j'appliquerai celle-ci correctement; cela garantit uniquement que la question de la correction de l'emploi peut recevoir une réponse (plus ou moins) catégorique. On pourrait dire que le problème que posent les dispositions dans ce contexte, c'est que la relation entre une disposition et son exercice n'est pas de type «contractuel» (comme le note McDowell: «une disposition n'est pas quelque chose à quoi son exercice est fidèle»²²). La difficulté générale est donc qu'on imagine mal comment un ensemble de faits convenablement circonscrits concernant la manière dont j'emploie ou j'emploierai une certaine expression pourrait concerner la manière dont je *dois* l'employer.

En second lieu, si le «paradoxe» est lié au caractère fini de mon expérience passée, l'invocation de dispositions ne résoudra rien, car la totalité de mes dispositions est également entachée de finitude: pour des nombres suffisamment grands, l'assertion du dispositionnaliste selon laquelle l'apprentissage a mis en place une disposition à donner des réponses correctes à toute question concernant la somme de deux nombres quelconques n'est plus vraie; et le sceptique peut redéfinir la «quaddition» comme une fonction qui diverge de l'addition ordinaire à partir de ces nombres suffisamment grands pour lesquels je n'ai pas la disposition à donner des réponses correctes.

L'inadéquation des théories dispositionnalistes peut être illustrée en examinant une version tout à fait fruste et en l'appliquant au domaine de la signification des termes lexicaux. Selon une telle théorie, le terme «corbeau», par exemple, désignerait purement et simplement la propriété à laquelle je suis disposé à l'appliquer. A l'évidence, une telle «théorie» est vouée à l'échec car elle est incapable de restituer l'extension de «corbeau»: dans certaines conditions d'éclairage ou d'éloignement et/ou pour certaines pies offrant un aspect plus ou moins corvidésque, j'ai la disposition à appliquer l'expression à des choses qui ne sont pas des corbeaux. L'identification de la propriété que l'expression désigne à la propriété que mes dispositions me font désigner

²¹ L. WITTGENSTEIN, *op. cit.*, § 258.

²² J. McDOWELL, «Wittgenstein on Following a Rule», *Synthese*, 58, 1984, p. 329.

aboutit au résultat absurde que «corbeau» ne désigne pas la propriété CORBEAU mais la propriété disjonctive CORBEAU OU PIE.

Ne pourrait-on pas, cependant, tenter de sauver le compte rendu dispositionnaliste au moyen d'une théorie dispositionnelle modifiée et moins fruste qui en appellerait à des conditions idéalisées («s'il n'y avait des restrictions temporelles, cérébrales, mémorielles, etc., je répondrais correctement à toute question de la forme «combien font x et y »), c'est-à-dire à des dispositions opérant *ceteris paribus* pour déterminer la réponse à tout problème d'addition quels qu'en soient les arguments numériques?

Kripke n'est pas de cet avis. Il écrit, assez expéditivement, que

le dispositionnaliste ne réussira que si l'idéalisation inclut une spécification selon laquelle je continuerai à répondre, sous ces conditions idéalisées, en accord avec la table infinie de la fonction à laquelle je fais référence en réalité. Mais alors la circularité de la procédure est évidente. Les dispositions idéalisées ne sont déterminées que parce que la question de savoir à quelle fonction je me réfère a déjà été réglée²³.

Dans la littérature récente, on trouve des propositions qui paraissent être des variations sur ce thème: il existerait un ensemble de circonstances, un certain type de situation, où les sujets se trouvent être incapables de commettre des erreurs de jugement. Dans ces circonstances qu'on peut appeler «conditions d'optimalité», les propriétés auxquelles le sujet est disposé à appliquer une expression ne sont autres que les propriétés auxquelles l'expression s'applique effectivement.

Ces conditions d'optimalité peuvent être conçues de différentes façons. Dretske par exemple pense que ce sont celles où la signification de l'expression a été acquise²⁴, alors que les tenants d'une approche «téléologique» en philosophie de l'esprit (comme Papineau, Millikan, McGinn²⁵) diront que les conditions d'optimalité sont les conditions, définies par la biologie évolutionnaire, dans lesquelles nos mécanismes cognitifs fonctionnent comme ils *sont censés* fonctionner, la norme invoquée se rapportant aux avantages sélectifs darwiniens ayant présidé à l'incorporation de ces mécanismes dans l'espèce. La proposition de ces auteurs (connus sous l'appellation barbare de «téléosémanticiens») est, en gros, que le fait que nous soyons des «additionneurs» et non des «quadditionneurs» peut être expliqué en se référant aux «fonctions biologiques normales» des mécanismes cognitifs qui sous-tendent notre capacité à raisonner logiquement et qui résultent de la sélection naturelle. Dans

²³ S. Kripke, *op. cit.*, p. 28.

²⁴ Cf. F. DRETSKE, *Knowledge and the Flow of Information*, Oxford, Blackwell, 1981, p. 195 sq.

²⁵ Voir D. PAPINEAU, *Reality and Representation*, Oxford, Blackwell, 1987; R. MILLIKAN, *Language, Thought and Other Biological Categories*, Cambridge, Mass., MIT Press, 1984; C. MC GINN, *Mental Content*, Oxford, Blackwell, 1989.

ceste perspective, c'est la fonction biologique qui détermine le contenu des mécanismes: par exemple, la grenouille possède un mécanisme interne qui actionne les muscles de la langue en direction de la proie, lorsqu'un objet foncé de petite taille – mouche ou balle de plomb – traverse son champ de vision. Mais le dispositif est un attrape-mouche et non un attrape mouche-ou-balle de plomb car sa fonction biologique normale est d'attraper des mouches.

Un examen adéquat de cette tentative de sauver la solution dispositionnaliste au paradoxe de Kripke-Wittgenstein demanderait de longs développements. Aussi dois-je me contenter de signaler la principale difficulté de principe, mais de taille, à laquelle cette tentative semble se heurter, et qui n'est, sur le fond, pas très différente de celle soulevée par Kripke.

Il n'est pas difficile de voir que l'approche téléologique promet en principe de sauvegarder la distinction entre la réponse qu'un sujet donnerait idéalement et celle qu'il donne ou est disposé à donner en fait. Cependant, quels que puissent être par ailleurs les mérites de l'approche téléologique²⁶, la spécification des conditions où la possibilité d'une divergence de type «quaddition» ou corbeau-pie soit exclue ne pourra donner satisfaction que pour autant que cette spécification n'invoque pas des matériaux sémantiques ou intentionnels faute de quoi la théorie prêtera le flanc au reproche de présupposer les notions mêmes dont elle était censée rendre compte.

La difficulté fondamentale, comme l'a bien vu P. Boghossian, peut être mise en rapport avec le caractère holistique des croyances :

Le problème est que, dans des circonstances normales, la fixation de la croyance est typiquement médiatisée par une théorie d'arrière-plan [...]. Etant donné un ensemble adéquat de suppositions d'arrière-plan, à peu près n'importe quel stimulus peut causer à peu près n'importe quelle croyance²⁷.

Le sujet X peut en arriver à croire qu'il est en train d'observer un corbeau alors qu'il s'agit en fait d'une pie, si, parmi ses croyances d'arrière-plan, figure la croyance que les corbeaux ont précisément l'aspect que présente l'oiseau qu'il observe, ou la croyance que Y est un ornithologue compétent conjointement à la croyance que Y croit que l'oiseau que X est en train d'observer est un corbeau, et ainsi de suite, en un nombre potentiellement infini de façons.

Cette observation a une incidence sur la viabilité d'une théorie dispositionnelle de l'espèce considérée : les conditions d'optimalité qu'une telle théorie devrait spécifier sont celles où le sujet aura la disposition à employer un terme ou à appliquer une règle en respectant leurs conditions de correction, c'est-à-dire à employer «corbeau» ou «plus» seulement à propos de corbeaux

²⁶ Voir D. DENNETT, *The Intentional Stance*, Cambridge, Mass., MIT Press, 1987, tr. fr. P. Engel, *La stratégie de l'interprète*, Paris, Gallimard, 1990.

²⁷ P. A. BOGHOSIAN, «The Rule-Following Considerations», *Mind*, vol. 98, 1989, p. 539.

ou d'additions en bonne et due forme. Mais les considérations sur l'«holisticité» montrent que pour spécifier ces conditions d'optimalité, il faudra spécifier, sans recourir au mode de discours sémantique ou intentionnel, une situation caractérisée par l'exclusion de *toutes* les croyances d'arrière-plan pouvant éventuellement occasionner une divergence entre la croyance ou l'emploi effectif et la croyance ou l'emploi correct.

Mais justement, comme le note encore Boghossian, «si nous avions *cela*, nous serions déjà en possession d'une théorie réductrice de la signification. Nous n'aurions pas besoin d'une théorie dispositionnelle»²⁸.

Autrement dit: la notion d'optimalité étant elle-même une notion normative, le problème d'indétermination du contenu, pour la résolution duquel on l'invoque, va se poser de nouveau à son sujet. Ces brèves remarques n'épuisent pas le débat, naturellement, mais j'espère qu'elles indiquent au moins dans les grandes lignes la nature des difficultés que le téléménticien rencontrera pour construire une théorie authentique du contenu qui puisse résoudre de manière satisfaisante le «paradoxe» de Wittgenstein. Pascal Engel a illustré récemment ces difficultés de la manière suivante:

[Que les mécanismes sur lesquels reposent nos croyances et autres états intentionnels soient le produit de la sélection naturelle] n'implique pas que l'on puisse déterminer le contenu des états mentaux à partir de leurs propriétés informationnelles, physiques ou fonctionnelles, pas plus que le fait que la main soit un organe qui a une certaine fonction biologique n'implique que l'on puisse expliquer tous les actes de préhension de la main en termes de leurs fonctions. On peut en dire autant de nos capacités cognitives supérieures²⁹.

3. La «solution» computationnaliste

Revenons à la question du computationnalisme. Selon Kripke, la position dispositionnaliste ne considère pas le sujet autrement que comme une sorte de machine, dont les actions potentielles réalisent *une* fonction telle que son «output» coïncide automatiquement avec la valeur correcte de la fonction visée. Que l'on donne cependant au terme «machine», dans l'affirmation «la machine réalise la fonction *y*», le sens de «programme» ou de «machine physique», des objections parallèles à celles qui militent contre la théorie dispositionnaliste «pure» peuvent être adressées à sa variante mécanique ou computationnelle.

Si c'est d'un programme qu'il s'agit, deux possibilités se présentent: pense-t-on à un programme rédigé, à une liste d'instructions composée de

²⁸ *Ibid.*, p. 540.

²⁹ P. ENGEL, *Etats d'esprit*, Aix-en-Provence, Alinéa, 1992, p. 142.

signes, le raisonnement que le sceptique avait tenu à propos du signe «+» peut évidemment être appliqué de nouveau au programme; pense-t-on à un «objet mathématique abstrait», on n'aura fait que reformuler la question originelle du sceptique (celui-ci demandant à présent que l'on exhibe quelque fait qui permette d'affirmer que le programme rédigé correspond à tel programme – abstrait – plutôt qu'à tel autre) et l'objection reste entière: les raisons que nous pouvons avoir d'affirmer cette correspondance du programme rédigé à un programme abstrait particulier (par exemple une certaine machine de Turing) sont toujours compatibles avec la supposition que la liste d'instructions écrites exemplifie un programme abstrait différent, qui détermine des valeurs coïncidant avec celles du premier sur une section seulement du parcours de valeurs.

Si, d'autre part, on songe à une machine concrète, des problèmes strictement analogues à ceux de la finitude des dispositions et de l'intrusion possible de dispositions à commettre des erreurs se posent: la machine est un objet physique fini à capacité limitée; quel que soit le comportement qu'elle manifeste, celui-ci est compatible avec un nombre indéfini de programmes qui le prolongent de différentes façons. Choisir parmi ceux-ci *le* programme réalisé, en se référant aux intentions du programmeur ou du concepteur, nous ramènerait simplement aux cas précédents. On ne manquera pas d'observer, à ce propos, que nos calculatrices et ordinateurs implémentent toujours des fonctions de type «*quus*», puisqu'ils affichent un message d'erreur, ou un symbole arbitraire comme «1», à partir d'arguments suffisamment grands. Cette remarque explique, incidemment, pourquoi une solution du paradoxe en termes de *simplicité* n'est pas recevable. Cette notion étant éminemment épistémique et pragmatique, son contenu varie en fonction de nos buts. Pour l'ingénieur, par exemple, il est «plus simple» de construire des quadditionneurs. Dire que «*plus*» est plus simple que «*quus*» se réduit dans ce contexte à l'assertion que notre signe «+» dénote le concept d'addition et non celui de quaddition. Ce que personne ne conteste.

Nous touchons ici à un des points sur lesquels la philosophie de la psychologie wittgensteinienne paraît devoir concerner directement les sciences cognitives, ou du moins, le paradigme computationnel dont ses praticiens s'inspirent le plus souvent, ou s'inspiraient jusqu'à une date récente. Le différend fondamental porte sur le statut de ce qu'on pourrait appeler les trois «questions fondatrices» du paradigme computationnel:

- 1) Peut-on dire d'une chose qu'elle pense, qu'elle calcule, qu'elle comprend, simplement en arguant du fait qu'elle manipule (mécaniquement) des symboles, définis formellement, en conformité avec les règles d'un système formel?
- 2) Les capacités cognitives humaines (ou des autres animaux) sont-elles en fait des capacités d'effectuer des opérations computationnelles sur des éléments spécifiés formellement?

- 3) Nos descriptions ordinaires de la vie mentale des êtres humains, à la première ou à la troisième personne, sont-elles des descriptions, certes à un niveau d'abstraction élevé, des états et événements du cerveau?

Pour la psychologie computationnelle, ce sont là des questions qui ont une chance raisonnable d'être résolues expérimentalement alors que du point de vue wittgensteinien, il s'agit avant tout de confusions conceptuelles qui ne constituent qu'en apparence des perplexités devant les causes de phénomènes naturels. Pour le dire tout à fait brutalement: la thèse du mécanisme computationnel serait incohérente.

Une manière de rendre cette affirmation plausible est d'attirer l'attention sur ce que Kenny appelle «le sophisme de l'homoncule»³⁰ résultant de la pratique courante qui consiste à appliquer des descriptions réservées en principe à l'être humain pris dans son intégralité à des objets qui ne lui ressemblent pas suffisamment ou plus du tout. Le paragraphe 281 des *Recherches philosophiques* est régulièrement cité à ce propos:

C'est seulement de l'homme vivant et de ce qui lui est semblable (se comporte de façon semblable à lui) que l'on peut dire qu'il a des sensations, qu'il voit; qu'il est aveugle; qu'il entend; qu'il est sourd; qu'il est conscient ou inconscient.

Mais le philosophe wittgensteinien ne peut assurément se contenter d'accuser le tenant de la thèse mécaniste d'avoir commis une version du «sophisme de l'homoncule».

La solution computationnaliste au paradoxe de Kripke-Wittgenstein revient à affirmer que la source de la normativité et la garantie du caractère déterminé du résultat de l'application d'une règle sont à situer dans l'univocité des symboles du programme et/ou dans la contrainte du mécanisme qui l'implémente. Une certaine conception de ce qu'est un algorithme sous-tend cette tentative de résorption du normatif dans le mécanique. D. Knuth propose cette définition typique:

Un algorithme est un ensemble de règles ou d'instructions pour l'obtention d'une certaine sortie pour une certaine entrée. Le trait distinctif d'un algorithme est que toute indétermination doit être éliminée. Les règles doivent décrire des opérations si simples et si bien définies qu'elles peuvent être exécutées par une machine³¹.

Du point de vue wittgensteinien, avant même de discuter de la *nature* des règles que la machine est censée pouvoir calculer, il convient de poser la question du sens qu'il faut donner au terme «règle» dans le contexte de la manipulation mécanique des symboles non interprétés. Autrement dit, de ce

³⁰ Cf. A. KENNY, «The Homunculus Fallacy» in *The Legacy of Wittgenstein*, Oxford, Blackwell, 1987.

³¹ D. KNUTH, «Algorithms», *Scientific American*, vol. 236, avril 1977, p. 63.

que la phrase «la machine suit la règle» signifie (le wittgensteinien a sûrement raison de se méfier d'une vision naïve de la compositionnalité sémantique : ce n'est pas parce que nous savons ce que «machine» et «règle» veulent dire en général que nous sommes au clair au sujet de ce que «la machine suit la règle» veut dire).

Dans les *Remarques sur la philosophie de la psychologie*, Wittgenstein écrit, assez énigmatiquement :

Les ‘machines de Turing’. Ces machines sont bel et bien des *hommes*, qui calculent³².

Tentons d'expliquer cette remarque. Compte tenu du fait qu'une des leçons du célèbre article de Turing de 1937 est que le domaine des fonctions récursives est idoine pour l'implémentation mécanique, ou, pour le dire de façon encore moins précise, que la manipulation de symboles spécifiés formellement *conformément* à des règles peut être mécanisée, que s'ensuit-il? Faut-il dire que la machine suit mécaniquement les règles qui gouvernent la tâche à exécuter? Que, littéralement, c'est elle qui calcule? N'est-il pas plutôt préférable de dire – même si, de prime abord, cela peut paraître paradoxal – que la machine à calculer fournit le résultat d'un calcul sans calculer, un peu comme on préférerait dire d'un piano mécanique qu'il peut produire de la musique sans en jouer, plutôt que de dire qu'il la produit en «se jouant lui-même»?

On peut localiser une des sources de la difficulté dans l'idée, suggérée par l'exposé de Turing lui-même, qu'il y aurait donné «une analyse logique convaincante et complète de la notion de calcul³³» ou, comme l'explique Hao Wang que

ce que fit Turing, ce fut d'analyser l'acte humain de calcul pour arriver à un petit nombre d'opérations simples dont la nature mécanique est évidente et dont on peut cependant montrer qu'elles peuvent être combinées pour exécuter des opérations mécaniques arbitrairement complexes³⁴.

³² L. WITTGENSTEIN, *Remarques sur la philosophie de la psychologie*, I, § 1096. Dans son célèbre article de 1937 («On Computable Numbers with an Application to the Entscheidungsproblem», *Proceedings of the London Mathematical Society*, Vol. 42, 1937) Turing établit un résultat analogue à celui du premier théorème d'incomplétude de Gödel. Il procède en deux étapes. D'abord, il montre qu'un certain type de machine abstraite (dont il définit la structure formelle) est capable d'exécuter n'importe quelle procédure effective binairement codable. Cette première étape garantit que la validité de toute conclusion ultérieure ne sera pas limitée par une conception trop étroite de ce qu'est un mécanisme. Et il montre ensuite qu'il n'existe pas de machine de ce type capable d'accepter comme entrée une proposition mathématique arbitraire et de fournir en sortie la réponse correcte à une question qui peut être interprétée comme : «la proposition est-elle démontrable?»

³³ M. DAVIES, «What is a Computation?», in L.A. STEEN (ed.), *Mathematics Today*, New York, Springer Verlag, 1978, p. 241.

³⁴ H. WANG, *From Mathematics to Philosophy*, London, Routledge and Kegan Paul, 1976, p. 91.

Un wittgensteinien comme Stuart Shanker objecte cependant que

l'analyse de Turing ne nous a pas fait passer d'un concept vague de calcul à un concept précis, mais a plutôt construit un concept entièrement nouveau auquel Turing a donné, de façon potentiellement trompeuse, le même nom. [...] Le danger réside dans la supposition que cette nouvelle version est, d'une certaine manière, plus fidèle aux faits que notre ancien concept (normatif) de calcul. Notre tâche, par conséquent, est de distinguer ces deux concepts différents, dont la similitude est uniquement d'homophonie³⁵.

Dans les *Remarques sur les fondements des mathématiques*, Wittgenstein fait une suggestion (qui n'est pas sans rappeler la fameuse Chambre chinoise de Searle) qui revient à imaginer que l'on ait entraîné un être humain à manipuler un système formel non interprété dont les théorèmes se trouvent être isomorphes aux résultats correctement obtenus par application des règles de l'arithmétique. On imaginera, par exemple, que cet individu ait mémorisé la table d'une machine de Turing correspondant aux opérations élémentaires de l'arithmétique. Wittgenstein demande si un tel individu a acquis la même capacité mathématique que quelqu'un qui connaît l'arithmétique pour l'avoir apprise par la voie normale³⁶.

La réponse sera vraisemblablement négative, mais qu'est-ce qui lui manque, au juste ? (L'idée que ce qui lui manque n'est pas de l'ordre d'une expérience psychologique ineffable représente un des thèmes les mieux assimilés de l'enseignement wittgensteinien.) Certes *nous* pouvons utiliser un tel individu, qui serait de fait une sorte de machine à calculer humaine, pour vérifier des calculs ou pour nous épargner des calculs fastidieux, mais il semble évident qu'un tel «idiot savant» serait dépourvu des capacités les plus élémentaires que nous associons normalement à une compréhension, même rudimentaire, de l'arithmétique, comme de savoir compter, savoir ce qu'est un nombre, ce qu'est une opération, être capable d'appliquer ces savoirs pour faire des inférences simples à propos d'objets observés, etc. Même s'il parvient à des résultats arithmétiquement corrects, il lui manque, pour ainsi dire, la saisie de l'inférence en tant qu'inférence.

On ne saurait cependant nier que cette machine à calculer humaine suive, d'une certaine façon, des règles, même si ce ne sont pas les règles de l'arithmétique. Pourquoi dès lors refuser à la machine (ordinateur ou calculatrice) cette faculté ? J. Hyman propose à ce sujet une analogie assez éclairante :

Supposez que le directeur d'une école décrète que les élèves ne doivent pas quitter le terrain de l'école pendant la pause de midi. Une manière de faire respecter cette règle serait de punir ceux qui la transgessent. Pourrait-on aussi la faire

³⁵ S. SHANKER, Introduction au vol. IV de S. SHANKER (ed.), *Ludwig Wittgenstein. A Critical Assessment*, London, Croom Helm, 1986, p. 18.

³⁶ L. WITTGENSTEIN, *Remarques sur les fondements des mathématiques*, § 258.

respecter en fermant à clé le portail de l'école pendant la pause de midi? Il est tentant de dire que ce serait la meilleure façon possible de faire respecter la règle, mais ceci serait certainement une erreur. La fermeture à clé du portail de l'école ne fait pas respecter la règle, elle rend la règle superflue. [...] La mécanique a remplacé les surveillants, mais une mécanique n'est pas un genre particulier de surveillant³⁷.

La conclusion de cet exemple nous ramène à la question de la normativité que les comptes rendus dispositionnels ne parviennent pas à restituer: si je suis causalement empêché de transgresser une règle (ou, d'ailleurs, de lui obéir), la question de mon obéissance ou de ma transgression est déplacée. On en revient toujours à ceci que la simple *conformité* avec la règle assurée par un mécanisme ne fournit pas en elle-même un exemple représentatif de ce qu'est suivre une règle; car, comme l'écrivent Baker et Hacker:

Suivre une règle c'est l'*employer* pour guider la conduite (et non être causalement constraint de faire quelque chose). C'est la consulter pour évaluer la correction ou la non-correction des opérations qui *devraient* lui être conformes. C'est justifier la conduite, lorsqu'on est sommé de le faire, en s'y référant, identifier la conduite sous son aspect normatif approprié et expliquer les identifications et descriptions normatives par référence à la règle. Aucun de ces actes, activités, raisons et justifications n'entrent en compte dans les mécanismes que nous construisons pour nous épargner des tâches normatives ennuyeuses³⁸.

Pour résumer: le wittgensteinien insiste sur la différence, qu'il tient pour fondamentale et indéracinable, entre l'idiot savant, aussi déshumanisé puisse-t-il paraître, et la machine programmée pour obtenir le résultat d'un calcul. Et il insiste également sur le fait qu'il ne s'agit pas, en l'occurrence, d'une différence «sentimentale». Ce qui est nié a priori, ce n'est pas la possibilité de fabriquer des créatures capables de suivre des règles, mais la possibilité d'en fabriquer sans fabriquer du même coup des créatures capables d'employer des règles d'une façon qui coïncide suffisamment avec le genre d'emploi (dans l'explication, la justification, l'instruction ou l'évaluation) auquel nous soumettons les règles. En tout état de cause, aucun «résultat net» ne suffit à légitimer l'attribution d'une capacité aussi générale et ramifiée que celle qui consiste à suivre des règles.

Une objection possible serait d'accuser Wittgenstein (ou les wittgensteiniens) de pétition de principe: le calculateur humain et la simulation mécanique ne différeraient fondamentalement que si l'on présuppose, comme critère, la possession d'une série de capacités intentionnelles supplémentaires qui ne soient pas à leur tour le produit causal de processus dont la nature est

³⁷ J. HYMAN, Introduction à J. HYMAN (ed.) *Investigating Psychology*, London, Routledge, 1991, p. 15.

³⁸ G. P. BAKER, P. M. S. HACKER, *Language, Sense and Nonsense. A Critical Investigation into Modern Theories of Language*, Oxford, Blackwell, 1984, p. 257.

computationnelle. Je reviendrai sur cette importante objection à la fin de mon texte.

Ayant passé en revue quelques-unes des principales solutions du «paradoxe» dont on peut penser qu'elles sont des pseudo-solutions, nous aboutissons à ce dilemme : ou bien la règle constraint, ou bien il n'y a pas de règle. Pour Bouveresse, «il semble que nous ne trouvions à notre disposition que des formes d'action qui sont trop rigides, comme celle de la contrainte, ou trop lâches, comme celle qui suppose une interprétation ajoutée à chaque fois à la règle»³⁹.

La lecture de Wittgenstein par Kripke admet que le refus d'une des branches du dilemme – celle des pseudo-solutions «hyper-rigides» ou causales – constitue *ipso facto* un argument en faveur de l'autre branche – celle de la «faiblesse de la règle» révélée par le paradoxe sceptique.

Pourtant, le point visé par Wittgenstein semble être plutôt qu'un tel dilemme ne surgit que lorsque nous acceptons certaines présuppositions ou adoptons certaines images de ce que sont des choses comme la compréhension, l'intention, etc. Il ne cherche pas à nous convaincre de la supériorité d'un des termes de l'alternative sur l'autre ; il cherche à nous persuader de rejeter le dilemme en rejettant les présuppositions et les images qui en conditionnent le surgissement.

4. Contre l'interprétation de Kripke

L'interprétation de Kripke est-elle fidèle ? C'est ce qu'il faut voir maintenant. Kripke ne s'est-il pas fourvoyé 1) au sujet de la nature du problème posé par le paradoxe sceptique et 2) au sujet de la solution (ou dissolution) qu'il convient d'y apporter ?

1) Pour la plupart des commentateurs, Wittgenstein ne cautionne pas en tant que tel le «paradoxe» selon lequel «aucune façon de faire ne peut être déterminée par une règle, puisque n'importe quelle façon de faire peut être mise en accord avec la règle»⁴⁰, étant donné qu'il déclare un peu plus loin que ce paradoxe provient d'un *malentendu*.

Kripke identifie le paradoxe sceptique qu'il attribue à Wittgenstein avec le paradoxe que Wittgenstein formule au début du paragraphe 201 des *Recherches philosophiques* :

Ceci était notre paradoxe : aucune façon de faire ne peut être déterminée par une règle, puisque n'importe quelle façon de faire peut être mise en accord avec la règle. La réponse était : si n'importe quoi peut être mis en accord avec la règle,

³⁹ J. BOUVERESSE, «Le 'paradoxe de Wittgenstein', ou comment peut-on suivre une règle?» in *Ludwig Wittgenstein*, numéro spécial de la revue *Sud*, 1986, p. 22.

⁴⁰ L. WITTGENSTEIN, *Philosophische Untersuchungen*, § 201.

alors n'importe quoi peut être mis en conflit avec elle. Et par conséquent il ne saurait être question ici ni d'accord, ni de conflit.

Mais la suite de cette remarque, dont Kripke ne semble pas tenir compte, dit :

On peut voir qu'il s'agit ici d'un malentendu du simple fait qu'au cours de notre argument nous donnons une interprétation après l'autre ; comme si chacune nous donnait satisfaction temporairement, jusqu'à ce qu'une autre, qui se tient derrière elle, nous vienne à l'esprit. Ce que cela montre, c'est qu'il y a une manière de saisir une règle qui n'est pas une interprétation mais qui se manifeste dans l'application au cas par cas dans ce que nous appelons «suivre une règle» ou «aller à l'encontre de la règle».

Le «paradoxe» a donc plutôt l'allure d'une réduction à l'absurde d'une certaine manière de concevoir ce qu'est la compréhension d'une règle. La conclusion inacceptable qui contraindrait à reconsidérer l'une des prémisses qui y aboutissent est explicitement énoncée par Wittgenstein : la vérité de la proposition paradoxale impliquerait d'une manière générale qu'il n'y aurait plus de sens à parler d'accord ou de désaccord ou, ce qui dans le présent contexte revient à peu près au même, que «l'idée même de signification part en fumée⁴¹». Et il désigne également, dans le même paragraphe, la prémissse à réexaminer, à savoir : la seule manière de comprendre une règle consiste à l'*interpréter*. La manière dont Wittgenstein est tenté de rectifier le mode d'expression de l'interlocuteur imaginaire qui dit que l'accord de l'application avec la règle se fait toujours par l'intermédiaire d'une interprétation en témoin-gne : «Chaque interprétation est, en même temps que ce qu'elle interprète, suspendue en l'air; elle ne peut lui servir de support. Les seules interprétations ne déterminent pas la signification.»⁴²

Kripke, confronté aux arguments de Wittgenstein montrant que la signification ne peut faire ce qu'elle est censée faire (guider l'application du signe dans tous les cas futurs) si l'on s'en tient aux interprétations, conceptions et images traditionnelles de la signification, sans «contenir d'une manière mystérieuse la totalité des pas futurs»⁴³, n'en conclut pas simplement que la conception philosophique qui s'élabore à partir de l'image du guidage est à rejeter. Il conclut, en outre, que la démystification d'une notion de signification qui, à l'examen, se révèle impliquer des pouvoirs miraculeux ou magiques, compromet *toute* notion de signification. Cependant la conclusion qu'on peut tirer n'est pas que nul mot n'a de signification (à part celle que lui confère le consensus des usagers compétents), mais que pour qu'on puisse dire qu'il en a une, il n'est pas nécessaire qu'il existe quelque fait ou état mental qui dirige l'application du mot.

⁴¹ S. Kripke, *op. cit.*, p. 22.

⁴² L. WITTGENSTEIN, *Philosophische Untersuchungen*, § 198.

⁴³ *Ibid.*, § 188.

Parler d'un paradoxe sceptique qui minerait la conception ordinaire de ce qu'est suivre une règle, appliquer un mot, etc. induit en erreur car, comme l'écrit McDowell,

le but [de l'argument de Wittgenstein] n'est pas de suggérer que nous devrions être remplis d'inquiétude face à l'éventualité que des possibilités du type «quaddition» se réalisent [...]. En fait nous avons l'assurance qu'il n'en sera rien, et l'argument vise non à saper cette confiance, mais à modifier notre conception de son fondement et de sa nature⁴⁴.

Selon Kripke, la «conclusion sceptique de Wittgenstein» est qu'il n'y a pas de fait, pas de conditions de vérité qui correspondent à un énoncé tel que «Dupond veut dire, par «+», l'addition». En réalité, il semble plus exact de dire que la conclusion de Wittgenstein est, *en un sens*, qu'il n'y a pas de «fact of the matter» qui détermine comment on comprend une règle; et pour expliquer en quel sens il n'y en a pas, on pourrait expliquer que si on part d'une certaine préconception, sublimée et uniforme, de ce que l'on doit compter comme un fait, alors on peut être certain qu'aucun fait de ce genre ne pourra jouer le rôle escompté. Si, en niant que dans le cas d'une phrase comme «Dupond suit la règle R» il y ait un «fact of the matter», on veut dire non seulement qu'il est toujours concevable que des indices ultérieurs nous amènent à considérer que Dupond est en train de suivre R', mais que le doute à ce propos est toujours justifié, la conclusion est clairement erronée. En tout cas, la première chose à remarquer contre l'interprétation de Kripke est que le texte de Wittgenstein ne cautionne ni la négation de la *réalité* des règles, ni celle de la contrainte imposée par leur adoption, et qu'on n'est amené à penser autrement qu'à la faveur d'une confusion entre le caractère réductible d'une propriété et son caractère factuel.

2) S'il n'y a aucun état de choses qui soit constitutif du fait qu'un signe ou une règle vise une signification bien précise, et si l'on récuse la référence à la règle elle-même dans ses applications «normales», on peut effectivement être tenté de tirer les conclusions sceptiques les plus radicales au sujet du bien-fondé de notre notion de correction objective d'une application. Selon Kripke, Wittgenstein parvient à éviter ces conséquences sans avoir à renoncer pour autant à sa critique d'une notion de correction platonisante ou «indépendante par rapport à toute recherche possible». Mais il n'y parvient qu'en faisant de la communauté des usagers à laquelle appartient le sujet qui utilise une règle donnée, le dépositaire et le pourvoyeur de l'objectivité nécessaire pour qu'il y ait encore un sens à parler de correction ou d'erreur. La solution serait que si le sujet en question donne, dans un nombre suffisant de cas, des réponses qui concordent avec celles que donnent ou sont disposés à donner les membres de la communauté, on peut légitimement attribuer à ce sujet la compréhension

⁴⁴ J. McDowell, «Non-cognitivism and Rule-Following», p. 148.

de la règle ou la maîtrise du concept et qualifier ses réponses de correctes. Il n'y aurait aucun autre fait que l'on pourrait invoquer pour justifier qu'on revienne sur ce verdict: la correction de la réponse dépend intégralement (et trivialement) du commun accord, du consensus communautaire, puisque la leçon qu'il y a lieu de tirer du «paradoxe» est précisément que quiconque s'aventurerait à citer des considérations censées prévaloir contre le verdict communautaire ne pourrait s'appuyer, en définitive, sur rien.

Kripke conteste que son interprétation de la «théorie» de Wittgenstein infléchisse celle-ci vers la conception statistique selon laquelle la valeur d'une fonction est correcte si et seulement si elle est celle que trouveraient (presque) tous les membres de la communauté linguistique⁴⁵. La différence cruciale entre les deux «théories» serait que la seconde reste une théorie des conditions de vérité d'énoncés comme «par 'plus' nous voulons dire telle fonction», alors que selon le Wittgenstein de Kripke, de telles phrases n'ont pas à proprement parler de conditions de vérité mais seulement des conditions d'«assertabilité» ou de justification⁴⁶. Cela signifie, entre autres choses, que malgré l'indétermination présumée de toute imputation de règles ou de concepts, Wittgenstein «trouve un rôle utile dans nos vies à un 'jeu de langage' qui autorise, sous certaines conditions, à affirmer que quelqu'un 'veut dire telle ou telle chose' et que son application actuelle d'un mot 's'accorde' avec ce qu'il voulait dire dans le passé»⁴⁷. Que l'on explique cependant ce qu'il y a de litigieux dans l'impression de Kripke que Wittgenstein a abandonné, dans sa deuxième philosophie, le concept de condition de vérité au profit de celui de condition d'assertabilité, et cette manière de prendre ses distances d'avec la conception statistique perdra beaucoup de sa vraisemblance.

Si le consensus communautaire ou majoritaire jouait réellement ce rôle d'arbitre ultime et de source première dont Kripke le revêt, le contraste entre la réponse consensuelle et la réponse correcte serait aboli. Ceci constituerait un avatar de plus de l'abrogation de la distinction, dont Wittgenstein ne se lasse pas de souligner qu'elle est essentielle au concept de règle, entre être conforme et le sembler seulement. Dans une théorie de ce genre tout se passe comme si le verdict sur ce qu'est suivre une règle correctement ne pouvait s'appuyer sur aucune considération autre que celle, d'ordre statistique et non pas normatif, selon laquelle «il est en quelque sorte moins *vraisemblable* qu'un grand nombre de gens soient dans l'erreur»⁴⁸.

Mais pour Wittgenstein, il semble bien que la possibilité de la divergence collective généralisée est exclue en ce sens seulement qu'un fond d'accord,

⁴⁵ Cf. S. Kripke, *op. cit.*, p. 11.

⁴⁶ *Ibid.*, p. 74.

⁴⁷ *Ibid.*, p. 79.

⁴⁸ C. WRIGHT, «Rule Following, Objectivity and the Theory of Meaning», in S.H. HOLTZMANN, C.M. LEICH (eds), *op. cit.*, p. 104.

permettant des désaccords exceptionnels, est présupposé par le «jeu de langage» de l'attribution de concepts ou de règles. L'accord en question n'est pas du tout de l'ordre d'une convention que l'on pourrait instituer ou abroger délibérément, par exemple pour des raisons de commodité ou d'étiquette. Wittgenstein est parfaitement explicite sur ce point:

‘Vous êtes donc en train de dire que l'accord des hommes décide de ce qui est vrai et de ce qui est faux?’ – c'est ce que les hommes *disent* qui est vrai ou faux ; et ils sont d'accord sur le *langage* qu'ils utilisent. Ce n'est point un accord sur des opinions mais sur une forme de vie⁴⁹.

Il me semble que Wittgenstein entendrait l'expression «instituer (ou abroger) une forme de vie» à peu près comme «se défaire de sa condition humaine et de ses propensions et réactions naturelles», c'est-à-dire comme désignant un changement dont il ne nous est pas facile de nous faire une conception cohérente.

Kripke s'exprime comme si l'identification wittgensteinienne de la signification à l'usage et le remplacement de conditions de vérité par des conditions d'assertabilité étaient les deux faces d'une même médaille. Cependant, la référence à l'usage implique seulement qu'il n'y a pas de sens à chercher des conditions de vérité *générales* pour une séquence grammaticale de mots en dehors d'un certain contexte (ou d'un certain type de contexte) d'utilisation, mais en aucun cas que la distinction entre justification et vérité ne peut être maintenue quand on pose des questions concernant ce que quelqu'un veut dire en employant un signe ou quelle règle il applique. Le renvoi à l'usage est destiné, dans ce cas, à contrecarrer la tendance à identifier les conditions de vérité de «X veut dire telle ou telle chose» à l'occurrence d'un état ou événement mental, mais non pas à retrancher des mots comme «vrai» ou «fait» du «jeu de langage» de l'attribution de concepts, d'intentions ou de compréhension, ni à les revêtir d'une signification qui, au vu de l'usage que nous faisons de phrases contenant de tels vocables, ne saurait être la leur. En somme, Wittgenstein n'est nullement contraint de nier le truisme selon lequel «si un énoncé est vrai, il doit y avoir quelque chose en vertu de quoi il est vrai»⁵⁰, sauf à prendre la locution *quelque chose* dans le sens de *une chose*.

Par conséquent, renoncer à l'idée que la règle fonctionne comme une sorte d'étalement supra-empirique, ou qu'elle agit comme un déterminant empirique se propageant dans la collectivité par une sorte de contagion, ce n'est pas encore renoncer à l'idée de normes objectives pour jauger la correction des applications de la règle.

⁴⁹ L. WITTGENSTEIN, *op. cit.*, § 241.

⁵⁰ M. DUMMETT, *Truth and Other Enigmas*, London, Duckworth, 1978.

5. Sur l'abandon des «questions constitutives»

Voyons maintenant brièvement comment Wittgenstein caractérise positivement la «force de la règle»:

1) L'accord constaté dans la pratique ne peut pas être *expliqué* en invoquant des notions comme celle de «suivre une règle», ou de «comprendre la signification de la règle», car elles presupposent ce que l'on cherche à expliquer. Postuler des entités ou des processus pour rendre à la règle le degré de détermination qui paraît nécessaire pour expliquer l'accord de ceux qui la suivent ne résoud rien car les mêmes problèmes d'indétermination ressurgissent.

2) La perplexité philosophique qui s'exprime dans la question «comment peut-on suivre une règle?» ne saurait être dissipée en empruntant des images, des analogies ou des modèles à d'autres domaines, car la source de cette perplexité réside moins dans l'inadéquation des solutions proposées que dans l'incapacité où l'on se trouve, en philosophant, de se contenter d'explications qui, pour être terre à terre, n'en sont pas moins vraies et peut-être les meilleures possibles. Toutes les solutions envisagées s'avèrent être des pseudo-solutions parce qu'en cherchant un fondement ultime à la pratique, elles n'aboutissent qu'à une réification de celle-ci. Wittgenstein exprime cela par un renversement frappant:

En réalité, au lieu de dire que le sens de la règle détermine l'application, nous pourrions aussi bien dire que l'application régulière détermine le sens de la règle. Et cela signifie également que la règle n'entraîne pas son application de façon causale, au sens où un rouage en entraîne un autre dans un mécanisme⁵¹.

3) Le travail philosophique le plus utile sur le concept de «suivre une règle» est essentiellement négatif, et l'idée qu'il y ait *un* phénomène général correspondant à ce concept doit être rejetée. Qu'on n'apprenne pas ce qu'est suivre une règle en apprenant à suivre une Règle qui régit l'emploi des règles en général, mais qu'on le fasse en apprenant à suivre des règles déterminées, n'est pas simplement un fait accidentel concernant cet apprentissage, mais un de ces faits cruciaux dont la simplicité et la familiarité contribuent à occulter l'importance⁵².

L'impression que Wittgenstein esquive, en définitive, les «vraies questions» est fort répandue. Comme l'écrit Malcolm:

[On] veut savoir ce que *c'est*, pour une personne, que de *vouloir dire* (se représenter, penser, se souvenir, comprendre, etc.). Mais Wittgenstein parle seulement de la manière dont diverses expressions du langage sont employées. Il

⁵¹ L. WITTGENSTEIN, *op. cit.* § 190.

⁵² *Ibid.*, § 129.

n'aborde donc pas la question de la *nature* de ces choses – de *ce qui arrive* quand on veut dire ou qu'on comprend quelque chose⁵³.

Wittgenstein répond d'avance à cette critique en soutenant que les enquêtes qui portent sur des *natures psychologiques* ne peuvent pas (contrairement à ce que ses contradicteurs semblent supposer) s'achever par un simple acte d'ostension, interne ou externe : attirer l'attention sur l'usage des mots

ne signifie pas que je veuille seulement parler des mots. Car l'interrogation sur la nature de la représentation porte sur le mot «représentation» tout autant que ma question. Et je dis seulement que cette question ne saurait être décidée – ni par la personne qui se représente quelque chose, ni pour qui que ce soit d'autre en montrant («nicht durch ein Zeigen»); ni en décrivant quelque occurrence⁵⁴.

L'erreur, écrit-il ailleurs, c'est de dire que la signification consiste en quelque chose («Der Irrtum ist zu sagen, Meinen bestehe in etwas»⁵⁵).

Un tel abandon des questions constitutives comporte cependant des risques. Il y a une manière caricaturale et dogmatique de prétendre, en se réclamant de Wittgenstein, que l'«explication et la justification arrivent à leur terme plus tôt qu'on ne le pense» ; elle peut ressembler, à s'y méprendre, à de la paresse intellectuelle, attitude que Dennett stigmatise ainsi :

Tout se passe comme si les philosophes [wittgensteiniens] se déclaraient experts dans l'art d'expliquer les méthodes d'un prestidigitateur puis, lorsque nous leur demandons d'expliquer comment celui-ci s'y prend pour exécuter le tour de la-femme-que-l'on-scie-en-deux, ils nous expliquent que cela est réellement assez évident : le prestidigitateur ne scie pas vraiment la femme en deux ; il fait simplement en sorte que cela semble être le cas. ‘Mais comment fait-il *cela*’, demandons-nous. ‘Ce n’est pas notre rayon’, disent les philosophes – et certains d’ajouter, d’une voix sonore : ‘l’explication doit bien s’arrêter quelque part’⁵⁶.

Il veut dire que la Nature exécute, d'une manière ou d'une autre, le «tour» consistant à faire sortir le lapin de l'esprit du chapeau cérébral ; il doit donc exister une tâche conceptuellement cohérente qui consiste à essayer de diminuer notre ignorance des mécanismes responsables de ce «tour». Dans le cas de la-femme-que-l'on-scie-en-deux, nous avons peu de chances d'identifier le «truc» en restant assis à scruter la scène avec des yeux de lynx ; de même, il est peu probable que le sens commun nous fournisse les idées les plus éclairantes sur l'asymétrie entre l'enseignement et l'apprentissage.

⁵³ M. MALCOLM, *Nothing is Hidden*, Oxford, Blackwell, 1986, p. 168.

⁵⁴ L. WITTGENSTEIN, *op. cit.*, § 370.

⁵⁵ L. WITTGENSTEIN, *Zettel*, § 16.

⁵⁶ D. DENNETT, «Cognitive Wheels. The Frame Problem of AI» in C. HOOKWAY (ed.), *Minds, Machines and Evolution*, Cambridge, Mass., Cambridge University Press, 1984, p. 133.

Il existe donc quelques raisons d'accuser Wittgenstein de préjugés anti-computationnalistes car, dira-t-on, il faut bien que la capacité humaine à suivre et à inventer des règles puisse être expliquée d'une manière ou d'une autre; mais la proposition inverse, selon laquelle la thèse du computationnalisme est entachée d'une pétition de principe à l'encontre de l'approche wittgensteinienne, n'est pas non plus, me semble-t-il, dénuée de plausibilité. C'est, du moins, ce que j'ai essayé de montrer.

Entre ces pétitions de principe mutuelles, on perçoit tout de même une asymétrie. D'un côté nous avons un *programme* de recherche qui ne semble pas paralysé par le manque de clarté conceptuelle. Comme le dit Wittgenstein dans un contexte quelque peu différent :

la clarté philosophique aura le même effet sur la croissance des mathématiques que la clarté du jour sur celle des germes de pommes de terre; dans l'obscurité de la cave, elles atteignent des longueurs impressionnantes⁵⁷.

Or les chercheurs peuvent *désirer* la croissance des germes, en se prévalant du rôle productif pour la recherche d'un certain désordre conceptuel. L'image de la «machine pensante», par exemple, a clairement joué un tel rôle, au début du moins, dans le développement de l'Intelligence Artificielle et des sciences cognitives. D'un autre côté, Wittgenstein nous invite à décrire clairement l'agencement de nos concepts, leurs ramifications dans la pratique, en nous abstenant de chercher des explications plus «profondes», invitation que d'aucuns considèrent comme une forme de puritanisme intellectuel stérile, anachronique et obscurantiste.

Il n'est pas exclu que le choix entre ces deux directions s'adresse moins à notre entendement qu'à notre volonté, ce qui n'est peut-être pas une conclusion très satisfaisante.

Remerciements

Ce travail a été rédigé grâce à une bourse de la Fondation Fyssen que je remercie sincèrement. Les Professeurs Marie-Jeanne Borel et Daniel Schulthess ont bien voulu revoir mon texte et y apporter des améliorations. Stefan Imhoof m'a aidé par ses remarques judicieuses et Michel Kuhne m'a fait observer que les calculatrices «quadditionnent». Je les en remercie. Je suis redevable à la Société Romande de Philosophie de m'avoir fait l'honneur de me confier sa conférence annuelle. Je remercie enfin les participants pour leurs remarques.

⁵⁷ L. WITTGENSTEIN, *Philosophische Grammatik*, § 381.