



# PHARES

Vol. XV

Hiver 2015

Revue philosophique étudiante

*Nous remercions nos partenaires :*

- Chaire d'enseignement et de recherche  
La philosophie dans le monde actuel
- La Faculté de philosophie de l'Université Laval
- La Confédération des associations d'étudiantes et d'étudiants  
de l'Université Laval (CADEUL)
- L'Association des étudiantes et des étudiants de Laval inscrits  
aux études supérieures (AELIÉS)
- L'Association des chercheurs étudiants en philosophie  
de l'Université Laval (ACEP)
- L'Association générale des étudiantes et étudiants prégradués  
en philosophie de l'Université Laval (AGEEPP)
- Zone Coopérative de l'Université Laval

Revue *Phares*  
Bureau 514  
Pavillon Félix-Antoine-Savard  
Université Laval, Québec  
G1K 7P4

revue.phares@fp.ulaval.ca  
www.revuephares.com

ISSN 1496-8533



## **PHARES**

### **DIRECTION**

Simon Pelletier  
Jean-François Perrier  
Daphnée Savoie

### **RÉDACTION**

Simon Pelletier  
Jean-François Perrier  
Daphnée Savoie

### **RÉDACTION DU DOSSIER**

#### **«L'ESPRIT DES ANIMAUX**

##### **NON HUMAIN»**

Renée Bilodeau  
Simon Pelletier  
Jean-François Perrier  
Daphnée Savoie

### **INFOGRAPHIE ET MISE À**

#### **JOUR DU SITE INTERNET**

Pierre Moreau

### **COMITÉ DE RÉDACTION**

Sébastien Lacroix  
Kate Blais  
Julien Ouellet  
Maxime Marcel Vachon  
Andrée-Anne Bergeron  
Jean-François Perrier  
Philippe Bettez Quessy  
Anne-Sophie Alain  
Pier-Alexandre Tardif  
Antoine Pageau St-Hilaire  
Hugo Tremblay  
Simon Pelletier  
Benjamin Lavoie  
Marie-Christine Lamontagne  
Jeffrey Elawani  
François Lacroix  
Olivier St-Pierre Provencher  
Daphnée Savoie

Comme son nom l'indique, la revue *Phares* essaie de porter quelques lumières sur l'obscur et redoutable océan philosophique. Sans prétendre offrir des réponses aptes à guider ou à éclaircir la navigation en philosophie, cette revue vise, en soulevant des questions et des problèmes, à signaler certaines voies fécondes à l'exploration et à mettre en garde contre les récifs susceptibles de conduire à un naufrage. En outre, le pluriel de *Phares* montre que cette revue entend évoluer dans un cadre aussi varié et contrasté que possible. D'une part, le contenu de la revue est formé d'approches et d'éclaircissements multiples : chaque numéro comporte d'abord un ou plusieurs DOSSIERS, dans lesquels une question philosophique est abordée sous différents angles ; puis, une section COMMENTAIRES, qui regroupe des textes d'analyse, des comptes rendus, des essais, etc. ; et, enfin, une section RÉPLIQUES, par laquelle il est possible de répondre à un texte précédemment publié ou d'en approfondir la problématique. D'autre part, la revue *Phares* se veut un espace d'échanges, de débats et de discussions ouvert à tous les étudiants intéressés par la philosophie. Pour participer aux prochains numéros, voir la politique éditoriale publiée à la fin du présent numéro.

Nous vous invitons à consulter notre site Internet ([www.revuephares.com](http://www.revuephares.com)), où vous aurez accès à tous les articles parus dans *Phares*.

# Table des matières

## ***Dossier*** : Philosophie de l'esprit (l'esprit des animaux non humains)

- 9** Introduction : Que pensent les animaux non humains ?  
RENÉE BILODEAU
- 19** L'attribution de concepts aux animaux non humains  
SAMUEL CASHMAN-KADRI
- 39** Les quatre critères de Newen et Bartels et l'empirisme de Jesse Prinz :  
deux théories des concepts compatibles  
HUGO TREMBLAY
- 65** L'ascension intentionnelle nécessite-t-elle une ascension sémantique :  
une critique de Robert W. Lurz  
JULIEN OUELLET
- 81** La métacognition chez l'animal non humain  
KEBA COLOMA CAMARA
- 105** L'empathie chez les animaux : le modèle des poupées russes  
FÉLIX AUBÉ BEAUDOIN

## ***Commentaires***

- 125** Contexte de découverte et contexte de justification :  
une aporie dans l'empirisme logique ?  
PIER-ALEXANDRE TARDIF
- 147** L'influence décisive de Nietzsche sur le jeune Camus  
OLIVIER MASSÉ
- 169** Le Cas Nietzsche : sur la tentative d'une lecture métaphysique  
par Heidegger  
EMMANUEL CHAPUT
- 191** La complémentarité des formules de l'impératif catégorique  
SIMON PELLETIER
- 213** Autonomie et institution du social selon Debray, Gauchet et Castoriadis  
THIBAUT TRANCHANT
- 237** Utilité scientifique des classes d'espèces : orthodoxie et originalité  
de Leibniz dans les Nouveaux Essais  
LÉA DEROME
- 259** Connaissance et sagesse dans la philosophie de Louis Lavelle  
NICOLAS COMTOIS



## ***Dossier :***

*Philosophie de l'esprit*

*(l'esprit des animaux non humains)*



# Introduction :

## Que pensent les animaux non humains ?

RENÉE BILODEAU, *Professeure à l'Université Laval,*  
*Membre du groupe de recherche interuniversitaire sur*  
*la normativité (GRIN)*

La toile regorge de vidéos d'animaux non humains se comportant de manière sensible et intelligente. On y voit un hippopotame secourant une jeune antilope d'un crocodile, un éléphant venant prêter main-forte (si je puis dire) à un autre éléphant afin de déplacer une table coulissante, ou encore un chimpanzé utilisant une branche pour extraire des termites de leur termitière. Ces anecdotes provoquent l'immédiate sympathie des internautes qui n'hésitent pas à attribuer aux animaux d'autres espèces des états mentaux largement similaires à ceux que possèdent les êtres humains. Il est effectivement tentant de supposer que l'hippopotame était mu par l'empathie lorsqu'il s'est interposé pour sauver de la mort la petite antilope, que l'éléphant a compris que son congénère désirait déplacer la table et avait besoin de son aide pour le faire, et que le singe faisait preuve de rationalité instrumentale en choisissant le meilleur moyen (une fine branche) pour parvenir à ses fins (manger les termites).

La théorie de l'évolution a sonné le glas de l'idée des animaux-machines et nombreux sont nos contemporains qui adoptent spontanément la thèse voulant que les aptitudes de l'esprit humain s'inscrivent dans le prolongement des capacités cognitives des animaux non humains. Déjà en 1871, Charles Darwin affirmait qu'il n'y a qu'une différence de degré, non de nature, entre les facultés mentales des humains et celles des animaux d'autres espèces<sup>1</sup>. Il maintenait, à l'encontre d'Aristote et de sa fameuse définition de l'être humain, que «les animaux peuvent raisonner jusqu'à un certain point, ce qu'ils font évidemment sans l'aide d'aucun langage<sup>2</sup>». L'homologie des caractéristiques morphologiques de diverses espèces d'animaux a incité Darwin à s'intéresser aux

similitudes psychologiques et comportementales chez les membres de ces espèces. Parce qu'il était persuadé que le langage n'est pas la marque de l'intelligence, il a été amené à introduire plusieurs thèmes de recherche qui sont aujourd'hui au centre des études sur l'esprit des animaux : l'apprentissage, l'utilisation des outils, le raisonnement implicite, la conscience, la possession de concepts, la cognition sociale, le sens moral, les habiletés artistiques.

Près de cent vingt-cinq ans plus tard, John Searle défendait des intuitions analogues à celles de Darwin à propos des animaux non humains. Il se fondait, lui aussi, sur les ressemblances anatomiques de multiples espèces, insistant notamment sur les similitudes de fonctionnement du cerveau<sup>3</sup>. Tout comme les humains, nombre d'animaux reçoivent des stimuli perceptuels à travers leurs récepteurs sensoriels et traitent ces informations grâce à leurs processus cérébraux. Les états mentaux conscients sont causés par les mécanismes neuronaux sous-jacents, de sorte qu'ils sont eux-mêmes des propriétés du cerveau, mais des propriétés macroscopiques ou des propriétés d'un niveau supérieur. Ce sont eux qui donnent lieu aux réponses motrices et aux actions intentionnelles, tant chez les humains que chez des animaux non humains. Ainsi se comprend le comportement de Ludwig Wittgenstein, le chien de Searle, lorsque ce dernier rentre à la maison : Ludwig lui fait une fête parce qu'il le reconnaît et qu'il ressent de la joie à le revoir, des états mentaux conscients produits par l'activité neuronale de son cerveau, à l'instar de ce qui se déroulerait dans le cerveau d'un être humain en pareilles circonstances.

Le naturalisme biologique de Searle paraît très intuitif. Il rejoint largement le sens commun, et l'évocation de la théorie évolutionniste lui apporte une caution scientifique. Néanmoins, en y réfléchissant un peu, on est vite amené à douter que les choses soient si simples. Ni l'appareil sensoriel ni le cerveau des animaux non humains ne sont semblables à ceux des humains. Les différences biologiques ont assurément pour effet des différences mentales. En conséquence, ne nous rendons-nous pas coupables d'anthropomorphisme en projetant sur les animaux non humains les états psychologiques avec lesquels nous appréhendons le monde et interagissons avec

nos congénères? Ce problème est de taille et il a de nombreuses ramifications épistémologiques.

D'une part, bien que l'anthropomorphisme soit une erreur méthodologique, l'anthropodéni, pour reprendre l'expression de l'éthologiste Frans de Waal<sup>4</sup>, est une entrave non moins grande à la poursuite des recherches sur l'esprit des animaux. L'anthropodéni est un aveuglement aux caractéristiques humaines que posséderaient les animaux, ou aux caractéristiques animales que partageraient les humains. Plusieurs éthologistes, et c'est le cas de de Waal, considèrent l'anthropomorphisme comme une heuristique susceptible de faire progresser la science et d'aider le chercheur à expliquer le comportement des animaux. Bien sûr, il lui faut éviter de présupposer naïvement que les animaux ont des états psychologiques semblables aux nôtres. Par contre, si on essaie de comprendre les animaux du point de vue de ceux-ci, en s'appuyant sur une connaissance solide de leurs habitudes et de leur environnement, l'usage de notions qui prévalent pour l'explication du comportement humain devrait faciliter la formulation d'hypothèses utiles.

Cette attitude est toutefois décriée par plusieurs, ce qui nous amène à une seconde difficulté épistémologique. L'attribution d'états mentaux aux animaux non humains pour expliquer leur comportement n'a d'intérêt que pour autant que ces animaux exemplifient bel et bien les états que nous leur prêtons. Or, comment pouvons-nous procéder empiriquement pour déterminer ce qui se passe dans l'esprit des membres des autres espèces? Même si l'imagerie par résonance magnétique fonctionnelle permet de connaître quelles zones du cerveau sont sollicitées dans le cadre de diverses tâches, et même si les régions en question peuvent être analogues chez les humains et chez certains animaux non humains, notamment les grands singes, il y a un saut qualitatif entre ce que ces résultats physiologiques nous apprennent et l'affirmation que les perceptions, croyances, désirs ou émotions de ces organismes sont voisins des nôtres.

Nous pourrions en inférer que l'attitude la plus prudente est, en somme, d'éviter toute projection anthropomorphiste et de suspendre notre jugement pour le moment en espérant qu'un jour nous serons mieux outillés pour aller du physiologique au mental. Cette conclusion

est néanmoins trop faible selon de nombreux chercheurs. D'aucuns considèrent, en effet, que l'étude des animaux non humains doit procéder dans le respect d'un principe édicté par C. Lloyd Morgan, et connu sous le nom de Canon de Morgan, principe d'après lequel « en aucun cas nous ne devons interpréter une action comme le résultat de l'exercice d'une faculté psychique supérieure si elle peut être interprétée comme l'exercice d'une faculté se situant plus bas dans l'ordre psychologique<sup>5</sup> ». L'intuition de Morgan peut être clarifiée grâce à un exemple<sup>6</sup>. On a depuis longtemps noté que les chimpanzés suivent le regard de leurs congénères ou de l'expérimentateur avec lequel ils interagissent. Qu'est-ce que cela indique? Une analyse se fondant sur des processus inférieurs dans l'ordre psychologique explique le comportement du chimpanzé comme le produit d'un mécanisme évolutif qui l'amène à diriger la trajectoire de son propre système visuel en fonction de l'orientation de la tête ou des yeux de celui qu'il observe. Une analyse qui a recours à des processus psychiques supérieurs fait valoir, quant à elle, que le chimpanzé croit que celui qui regarde est attentif à quelque chose. Autrement dit, elle suppose que le singe prête un certain état mental à celui qui regarde. Pour qui se conforme au Canon de Morgan, l'hypothèse à favoriser serait la première tant que nous n'avons pas prouvé que les chimpanzés sont capables d'attribuer des états psychologiques à d'autres animaux, humains ou non humains, c'est-à-dire tant que nous n'avons pas de données probantes qu'ils ont une théorie de l'esprit<sup>7</sup>.

Bien que la signification du Canon de Morgan soit sujette à discussion du fait que les notions de faculté psychique inférieure et supérieure sont elles-mêmes litigieuses<sup>8</sup>, nombreux sont ceux qui partagent la conviction qu'on ne peut légitimement affirmer que les animaux sans langage ont des états mentaux. Leur argument est simple : un animal sans langage n'est-il pas un animal sans concepts? Et si l'animal n'a pas de concepts, comment le scientifique pourrait-il déterminer quel contenu il a à l'esprit? Pire, en généralisant l'idée qu'il vaut mieux donner des explications du comportement qui évoquent des processus inférieurs, certains auteurs utilisent les résultats des recherches sur les animaux non humains pour avancer,

à l'encontre de la thèse de la psychologie populaire, qu'une vaste proportion des actions intentionnelles des humains devrait elle aussi être expliquée par des opérations moins complexes plutôt que par des états mentaux et des mécanismes de raisonnement<sup>9</sup>. Que l'on soit ou non prêt à suivre ceux qui s'engagent dans ce paradigme de recherche, il faut reconnaître que les études sur l'esprit des animaux non humains provoquent une profonde remise en question de notre manière de comprendre la plupart des notions que nous employons pour rendre compte de l'esprit humain.

Je mentionnais à l'instant que d'aucuns considèrent qu'un animal sans langage ne peut avoir d'états mentaux faute de concepts pour en individuer le contenu. Cette façon de voir a été défendue avec vigueur par Donald Davidson à la fin du siècle dernier<sup>10</sup>. Les chercheurs réfractaires à abandonner la conviction du sens commun selon laquelle les animaux ont, eux aussi, des désirs et des croyances ne se sont toutefois pas avoués vaincus. Ils ont exploré diverses avenues avec pour objectif de réévaluer le lien traditionnel entre langage et concept. En effet, n'y a-t-il pas de moyens autres que linguistiques pour un contenu mental d'être représenté dans l'esprit d'un organisme ? On peut penser aux images mentales, aux modèles mentaux, ou encore aux cartes mentales, qui sont autant de types de représentation mentale non linguistique. On peut également se demander s'il n'y a pas des voies autres que la formulation verbale qui permettent de discriminer un concept d'un autre. Dans l'article qu'il nous propose, Samuel Cashman-Kadri examine différentes manières d'aborder cette difficulté. Il expose en premier lieu le fameux argument de Davidson, pour ensuite discuter les stratégies et critères suggérés par quelques auteurs afin de déterminer le contenu des éventuelles représentations conceptuelles des animaux sans langage. Bien que la majorité des chercheurs sur lesquels il se penche soit assez sceptique quant à la possibilité de préciser les contenus représentationnels des animaux sans langage, Cashman-Kadri juge la théorie d'Albert Newen et Andreas Bartels prometteuse du fait que les critères d'attribution de concepts qu'ils avancent sont plus souples.

Hugo Tremblay se montre également sympathique à la thèse de Newen et Bartels. Son article défend d'abord les quatre critères épistémiques formulés par ces auteurs dans le but de rendre compte de la possession de concepts. Il entreprend ensuite d'enrichir cette approche en la complétant avec la théorie empiriste des concepts de Jesse Prinz, une position qui se prononce sur le statut métaphysique de ceux-ci plutôt que sur les seules capacités épistémiques requises pour en posséder. Deux études de cas viennent illustrer les conclusions sur la nature des pensées qu'un tel point de vue autorise à tirer des investigations empiriques.

L'article de Julien Ouellet poursuit ces réflexions par l'examen des thèses du philosophe José Luis Bermúdez. Pour Bermúdez, les animaux peuvent avoir des pensées du premier ordre mais non des pensées d'ordre supérieur<sup>1</sup>. La raison en est que seuls les animaux qui ont un langage seraient capables d'avoir des pensées du second ordre. En effet, les images mentales ou les modèles mentaux ne peuvent faire l'objet de pensées d'ordre supérieur, celles-ci ne portant que des contenus linguistiques. Dans cette perspective, bien que les animaux sans langage soient susceptibles d'entretenir certains types de pensées à propos du monde et d'effectuer certains types de raisonnements (des «protoraisonnements», comme les nomme Bermúdez), les états mentaux qui peuvent leur être attribués sont considérablement restreints. C'est à la discussion de quelques arguments qui fondent la théorie de Bermúdez et aux objections qu'ils soulèvent que le texte de Ouellet est consacré.

La question de savoir s'il est possible aux animaux non humains d'avoir des pensées du deuxième ordre est reprise par Keba Coloma Camara dans un article qui aborde ce problème sous l'angle particulier du débat sur la métacognition. De l'avis de certains auteurs, la métacognition requiert d'une créature non seulement qu'elle puisse se représenter la tâche du premier ordre dans laquelle elle est engagée, mais qu'elle ait aussi la capacité de représenter les processus cognitifs sollicités par l'accomplissement de cette tâche. Pour contrer cette approche qu'ils estiment contraignante et erronée, des théoriciens s'inspirent des résultats des neurosciences cognitives et avancent que les connaissances métacognitives

ne sont pas nécessairement métareprésentationnelles. Il suffit pour parler de métacognition que l'organisme puisse évaluer et contrôler sa performance cognitive en fonction des objectifs cognitifs qu'il poursuit. Chacune de ces avenues a des répercussions importantes sur notre façon de concevoir ce qui se passe dans l'esprit des animaux non humains, conséquences que le texte de Camara tente de clarifier.

L'article de Félix Aubé Beaudoin clôt le numéro sur une autre question que l'on se pose à propos des animaux non humains, celle de savoir s'ils sont capables d'empathie. Ce problème n'est pas sans liens avec les précédents puisque, ici encore, une des difficultés concerne la complexité des habiletés cognitives que l'on doit attribuer aux autres animaux pour expliquer certains comportements altruistes. En examinant le modèle des poupées russes de Frans de Waal, Aubé Beaudoin cherche à montrer la plausibilité de l'hypothèse voulant que les formes d'empathie plus sophistiquées se développent à partir de formes plus simples. Il défend également l'idée que non seulement les résultats de diverses expériences empiriques sont compatibles avec l'explication des comportements d'aide des animaux non humains en termes d'états mentaux, mais que de telles explications sont plus vraisemblables que les interprétations concurrentes. À la suite de cette discussion, nous serons mieux à même de juger de ce qui fait agir l'hippopotame qui s'interpose pour sauver la jeune antilope du crocodile.<sup>12,13</sup>

- 
1. Charles Darwin, *The Descent of Man, and Selection in Relation to Sex*, London, John Murray, 1871, 2<sup>e</sup> éd., 1874 ; trad. franç. *La descendance de l'homme et la sélection sexuelle*, Paris, C. Reinwald, 1891, p. 136.
  2. *Ibid.*, p. 93.
  3. John R. Searle, «Animal Minds», dans *Midwest Studies in Philosophy*, vol. 19, n. 1, *Philosophical Naturalism*, sous la dir. de Peter A. French, Theodore E. Uehling et Howard K. Wettstein (1994), pp. 208-209.
  4. Frans De Waal, «Anthropomorphism and Anthropodenial: Consistency in our Thinking About Humans and Other Animals», *Philosophical Topics*, 27 (1999), 255–80.

5. C. Lloyd Morgan, *An Introduction to Comparative Psychology*, 1<sup>re</sup> éd., Londres, Walter Scott, 1894, p. 53. Une précaution est ajoutée par Morgan dans la seconde édition de son ouvrage pour souligner que ce principe autorise l'interprétation d'une activité en termes de processus supérieurs (des processus psychologiques plus complexes) si nous possédons des données indépendantes démontrant que l'animal observé est capable de tels processus.
6. Cet exemple est emprunté à Daniel J. Povinelli, *Folk Physics for Apes*, Oxford, Oxford University Press, 2000, pp. 24 *sqq.* Povinelli et son groupe ont procédé à diverses expériences qui l'ont conduit à favoriser une explication en termes de processus de niveau inférieur.
7. La littérature sur l'esprit des animaux non humains discute abondamment la question de savoir si seuls les humains sont en mesure d'attribuer des états mentaux à autrui ou de s'en attribuer à eux-mêmes. L'expression « théorie de l'esprit » désigne la première de ces capacités, alors que le terme « métacognition » s'applique à la seconde.
8. Cf. Elliott Sober, « Comparative Psychology Meets Evolutionary Biology: Morgan's Canon and Cladistic Parsimony », dans *Thinking with Animals: New Perspectives on Anthropomorphism*, sous la dir. de Lorraine Datson et Gregg Mitman, New York, Columbia University Press, 2005, 85-99; Sean Allen-Hermanson, « Morgan's Canon Revisited », dans *Philosophy of Science*, 72 (2005), 608-631.
9. Cf. Derek Penn, Keith Holyoak et Daniel Povinelli, « Darwin's Mistake: Explaining the Discontinuity between Human and Nonhuman Minds », *Behavioral and Brain Sciences*, 31 (2008), 109-178; Frans de Waal et Pier Francesco Ferrari, « Towards a Bottom-Up Perspective on Animal and Human Cognition », dans *Trends in Cognitive Science*, 14, 5 (2010), 201-207; Sara J. Shettleworth, « Clever Animals and Killjoy Explanations in Comparative Psychology » dans *Trends in Cognitive Science*, 14, 11 (2010), 477-481; Kristin Andrews, *Do Apes Read Minds? Toward a New Folk Psychology*, Cambridge, The MIT Press, 2012.
10. Donald Davidson, « Rational Animals », dans *Dialectica*, 36, 4 (1982), 317-327; trad. franç. « Animaux rationnels », dans Donald Davidson, *Paradoxes de l'irrationalité*, Nîmes, Éditions de l'Éclat, 1991, 63-75.
11. Les états et processus mentaux du premier ordre sont des états et processus dirigés vers le monde, tandis que les états et processus du second ordre portent sur les états et processus du premier ordre. Ainsi, ma croyance qu'il reste du vin blanc dans le réfrigérateur est une croyance du premier ordre, et ma croyance que je me trompais

lorsque je me rends compte qu'il n'en reste plus est une croyance du deuxième ordre.

12. Les personnes intéressées par la question de l'esprit des animaux non humains peuvent consulter les ouvrages généraux suivants : Colin Allen and Marc Bekoff, *Species of Mind: The Philosophy and Biology of Cognitive Ethology*, Cambridge, The MIT Press, 1997 ; Robert Lurz (sous la dir.), *The Philosophy of Animal Minds: New Essays on Animal Thought and Consciousness*, Cambridge, Cambridge University Press, 2009 ; Robert Lurz, *Mindreading Animals: The Debate over What Animals Know about Other Minds*, Cambridge, The MIT Press, 2011 ; Sara Shettleworth, *Cognition, Evolution, and Behavior*, 2<sup>nd</sup> ed., New York, Oxford University Press, 2010.
13. Je tiens à remercier toutes les personnes sans lesquelles la réalisation de ce dossier thématique n'aurait pas été possible : d'abord, bien entendu, les auteurs des articles publiés ci-après, ensuite, les codirecteurs de la revue *Phares*, et particulièrement Jean-François Perrier, qui ont suggéré d'ouvrir leurs pages à ce projet, et finalement les étudiants du séminaire de philosophie de l'esprit de l'automne 2014, pour les échanges stimulants dont ces textes sont le fruit.



# L'attribution de concepts aux animaux non humains

SAMUEL CASHMAN-KADRI, *Université Laval*

## 1. Introduction

Les questionnements à propos de l'esprit des animaux non humains sont pluridisciplinaires: ils intéressent aussi bien les philosophes de l'esprit que les psychologues cognitifs et les scientifiques qui travaillent en psychologie comparative ou en éthologie. Cela peut probablement s'expliquer, entre autres, par le fait que l'étude de la cognition animale peut nous renseigner sur plusieurs sujets certes reliés, mais tout de même distincts. Par exemple, les philosophes de l'esprit, en se questionnant sur la nature ou le fonctionnement de l'esprit des animaux, pourront être amenés à clarifier certaines notions théoriques fondamentales, comme celles de concept, de croyance ou de conscience. D'un autre côté, les spécialistes de psychologie cognitive pourraient peut-être apporter un éclairage précieux sur le développement de la psychologie de l'être humain au cours de son existence (de sa naissance jusqu'à sa maturité intellectuelle) s'ils sont capables de mieux comprendre le fonctionnement de l'esprit des animaux plus ou moins parents avec l'être humain. Par ailleurs, il importe de noter dès maintenant qu'une étude sérieuse de la cognition des animaux non humains se doit de reposer en grande partie sur des investigations empiriques, sans quoi on risquerait de s'enfermer dans une dynamique de dialogues de sourds, fondés sur des conceptions de l'esprit *a priori*, qui brouillerait la recherche plutôt que de la faire avancer.

Je propose, dans ce travail, d'explorer la notion de concept en lien avec la cognition animale. En effet, la notion de concept est habituellement fortement rattachée à celle de langage ou de capacité linguistique, et il semble donc problématique d'affirmer que les animaux non humains peuvent posséder des concepts, alors même qu'il est habituellement consensuel de dire que seuls les humains

sont des animaux véritablement linguistiques. Plus précisément, je me pencherai sur les difficultés méthodologiques pour l'attribution de concepts aux animaux non humains eu égard à cette apparence de liaison forte entre les concepts et le langage. Je me référerai notamment au texte paradigmatique de Chater et Heyes<sup>1</sup> et à celui de Newen et Bartels<sup>2</sup> pour soutenir la thèse qu'il est probablement justifié de conclure que l'esprit de certains animaux fonctionne par le biais de représentations conceptuelles, mais que certaines difficultés méthodologiques rendent très compliquée l'identification du contenu de ces représentations.

En premier lieu, j'exposerai la problématique générale de l'attribution de concepts aux animaux non humains, en ce qui concerne notamment la liaison entre la pensée et le langage. Pour ce faire, j'essaierai de montrer d'abord pourquoi la question de la possession de concepts chez les animaux se pose, et je présenterai ensuite brièvement le fameux argument de Davidson dont la conclusion est que la possession de concepts et de croyances (et donc de la pensée) implique la possession du langage. Par la suite, en m'appuyant sur ledit texte de Chater et Heyes, je montrerai qu'il existe certaines considérations méthodologiques qui semblent impliquer qu'il n'est pas vraiment possible d'attribuer de concepts aux animaux non humains en tant qu'ils sont non linguistiques. En troisième lieu, je présenterai la thèse soutenue par Newen et Bartels, en passant par celle mise de l'avant par Allen<sup>3</sup>, en montrant qu'elle semble indiquer qu'il est légitime d'attribuer des concepts à certains animaux, tout en notant toutefois que certains problèmes liés à l'identification du contenu particulier de ces concepts demeurent.

## *2. Problématique générale de l'attribution de concepts aux animaux non humains*

Il semble que la première question qui mérite d'être posée est celle de savoir pourquoi les spécialistes de la cognition animale sont portés à vouloir attribuer des concepts aux animaux non humains et à ainsi vouloir mettre de l'avant des expériences empiriques qui viseraient à confirmer ou à infirmer la légitimité d'une telle attribution. D'emblée, on peut remarquer que la notion de concept

fait partie intégrante de notre psychologie populaire, et que nous sommes souvent portés à utiliser les termes de cette psychologie populaire (tels que « croyance », « désir », etc.) pour rendre compte des agissements des animaux. Autrement dit, nous sommes en quelque sorte naturellement portés à appliquer les termes de notre psychologie humaine aux animaux, dont le terme « concept », et il est donc pertinent de mettre à l'épreuve, sur des bases plus sérieuses, cette inclination naturelle.

Par ailleurs, depuis quelques années, l'étude de la cognition animale a pris une dimension plus importante : les capacités cognitives des animaux sont étudiées pour elles-mêmes<sup>4</sup>, c'est-à-dire que certaines personnes sont spécialisées dans l'étude de la cognition animale, à l'instar de certains biologistes spécialisés, par exemple, dans l'évolution des invertébrés. Cela s'explique probablement parce qu'une meilleure compréhension des capacités cognitives des animaux pourrait sans doute mener à une meilleure compréhension de notre propre cognition humaine. Quoi qu'il en soit, ces spécialistes de la cognition animale ont pour but d'identifier et d'expliquer les processus sur lesquels sont fondés certaines habiletés cognitives et certains comportements tout à fait impressionnants d'animaux comme les rats, les oiseaux et les primates<sup>5</sup>. Or, il semble que ces comportements et ces habiletés soient plus facilement explicables si l'on considère que ces animaux sont des « systèmes intentionnels ». Selon Achim Stephan, « this means ascribing to animals the capacity to have beliefs and intentions — and herewith, the capacity to have concepts<sup>6</sup> ». D'aucuns suggèrent même qu'il est nécessaire d'attribuer des états mentaux relativement complexes (comme des concepts) aux animaux pour être en mesure d'expliquer certains de leurs comportements<sup>7</sup>. Bref, c'est d'abord et avant tout la complexité de certains comportements et de certaines prouesses cognitives d'animaux qui justifie de nous pencher sérieusement sur la possibilité de leur attribuer des concepts.

Cependant, l'une des difficultés majeures pour l'idée qu'il soit possible d'attribuer des concepts aux animaux non humains est que la notion de concept est traditionnellement fortement rattachée à celle de langage ou de capacité linguistique, alors que

seuls les êtres humains, parmi le règne animal, semblent être des animaux véritablement langagiers. Bien que j'aie remarqué, dans l'introduction, qu'il serait inadéquat d'appuyer l'étude de la cognition animale sur des considérations et des déductions *a priori*, il n'en reste pas moins qu'un des arguments les plus influents pour nier l'attribution de concepts (et de la pensée, de manière plus générale) aux animaux non humains (en tant que créatures non linguistiques) est de nature analytique, c'est-à-dire qu'il s'appuie sur les relations conceptuelles, notamment entre les notions de concept et de croyance. Penchons-nous un instant sur cet argument, car il permet de bien mettre la table sur la problématique générale de l'attribution de concepts aux animaux non humains, en ce qu'il met bien en évidence la connexion intuitive entre la notion de concept et le langage.

Il faut d'abord avoir en tête que pour Davidson, toute attitude propositionnelle (croyance, désir, intention, etc.) implique un arrière-plan constitué par de nombreuses croyances. En effet, c'est cet arrière-plan de croyances qui permet d'identifier et de distinguer les attitudes propositionnelles les unes des autres. La notion de croyance, pour Davidson, est donc fondamentale lorsqu'il est question d'attitudes propositionnelles. Ainsi, si Davidson montre que la notion de croyance est indissociable de celle du langage, il aura montré que les pensées (en tant qu'attitudes propositionnelles) ne sont possibles que pour une créature langagière. C'est d'ailleurs la thèse qu'il soutient : « My thesis is rather that a creature cannot have a thought unless it has language<sup>8</sup> ».

Sans rentrer dans les détails de l'argument, la première prémisse de Davidson est qu'il n'est pas possible d'avoir une croyance quelconque sans qu'on ne possède le concept de croyance. C'est que de posséder une croyance implique de comprendre qu'elle puisse être vraie ou fausse ; avoir une croyance implique ainsi la possession d'une croyance à propos de cette croyance (en tant que croyance), et donc un certain concept de croyance. Stephan résume efficacement la première partie de l'argument de Davidson ainsi :

Thus, without having quite a lot of beliefs about what a belief really is, one cannot have the concept of belief, and without

that one cannot have any beliefs at all. And from this it follows that without having the concept of belief one cannot have any other concept at all. For, without having quite a lot of beliefs, one cannot have concepts. So we get: *Without having the concept of belief, one has neither beliefs nor concepts*<sup>9</sup>.

La deuxième prémisse de l'argument de Davidson est que la possession du concept de croyance nécessite de posséder le langage. En effet, avoir le concept de croyance, c'est avoir le concept d'un certain état qui peut soit être faux, soit être vrai; il semble donc que le concept de croyance soit très rapproché du concept de vérité objective. Or, pour Davidson, c'est la communication linguistique qui est au fondement de la distinction entre le subjectif et l'objectif, et donc au fondement de la notion de vérité objective. Bien que Davidson admette qu'il n'est pas en mesure de montrer que la communication linguistique est *nécessaire* à la formation du concept de vérité objective, il n'est pas capable d'entrevoir un autre fondement possible à ce concept<sup>10</sup>. C'est pourquoi il conclut que le concept de croyance nécessite la capacité langagière, et donc qu'un animal qui ne possède pas le langage ne peut avoir de croyances ni de concepts, et ne peut donc pas penser.

### *3. Chater & Heyes : difficultés liées à l'attribution de concepts aux animaux non linguistiques*

Chater et Heyes, contrairement à Davidson, ne soutiennent pas que les animaux non humains, en tant que créatures non langagières, ne peuvent pas posséder de concepts par principe. En effet, plutôt que d'élaborer un argument *a priori*, ils se fondent surtout sur des considérations méthodologiques pour soutenir qu'il ne semble pas possible, du moins pour l'instant, de rendre compte de l'affirmation selon laquelle les animaux non humains possèdent des concepts: «Rather, we argue that we simply have no account of what cognitive feat possessing a concept amounts to for nonlinguistic agents, and hence cannot assess whether this or that animal possesses concepts or not, still less ascertain the content of its putative concepts<sup>11</sup>». Comment en arrivent-ils à soutenir cette thèse? D'abord, ils

proposent ces trois critères à propos de la notion de concept : 1) s'applique aux êtres humains et permet d'identifier les concepts des humains avec les termes du langage naturel ; 2) peut s'appliquer aux animaux non linguistiques et 3) permet d'étudier empiriquement la possession de concepts chez les animaux, notamment sous la forme d'investigations à propos du comportement des animaux. Selon eux, si au moins l'un de ces critères n'est pas respecté par une notion de concept, alors il est bien difficile de voir en quoi la recherche sur les concepts des animaux pourrait s'avérer fructueuse.

Or, justement, après avoir évalué plusieurs de ces notions, ils en arrivent à la conclusion qu'aucune d'entre elles ne respecte ces trois critères, car aucune ne permet de penser l'idée de concept indépendamment du langage naturel. On peut évidemment se demander de quelle manière les deux auteurs justifient ces trois critères, notamment la conjonction des deux premiers, qui impose que les caractéristiques des concepts telles que décrites par les théories des concepts humains puissent être appliquées dans l'étude de la cognition des animaux non humains<sup>12</sup>. La réponse de Chater et Heyes, à mon avis, n'est pas particulièrement convaincante : ils affirment qu'on peut remarquer que les recherches sur les concepts des animaux ne concernent pas seulement la structure de leurs représentations mentales, c'est-à-dire la manière dont ces représentations sont liées et organisées entre elles pour produire d'autres représentations ou des comportements, mais aussi le contenu de ces représentations. Or, les catégories habituellement utilisées dans les expériences sur la cognition des animaux sont des exemples paradigmatiques des concepts humains, comme « arbre », « personne » et « chat » ; il semble donc que ces recherches visent à répondre à la question de savoir si le contenu des représentations mentales des animaux est le même que celles des humains, d'où l'idée qu'une théorie des concepts pertinente pour la recherche sur la cognition animale devrait aussi être applicable pour les humains. Cependant, il me semble que l'on est en droit de se demander s'il serait possible d'étudier la structure des représentations mentales des animaux indépendamment de leur contenu ; dans un tel cas, il serait peut-être possible d'affirmer que certains animaux possèdent bien

quelque chose comme des concepts, sans que l'on soit pour autant capable de spécifier leur contenu précis. Par ailleurs, l'exigence selon laquelle une théorie des concepts des animaux devrait être en mesure d'établir si le contenu de leurs représentations mentales est exactement le même que celui des humains me semble excessivement stricte, voire exagérée, car Chater et Heyes ne montrent nullement que c'est le contenu spécifique des représentations humaines qui permet d'affirmer qu'elles sont d'ordre conceptuel.

Quoi qu'il en soit, la première hypothèse qu'explorent Chater et Heyes est celle selon laquelle si les animaux non humains possèdent des concepts, alors cela voudrait dire que leurs comportements de catégorisation sont fondés sur les mêmes structures mentales que celles proposées par les théories des concepts humains : « One suggestion concerning how concepts can be identified nonlinguistically is that a concept is an internal state with a particular structure (a definition, a prototype, or a set of exemplars), which, in the human case, may or may not have the same meaning as terms of a natural language<sup>13</sup> ». C'est pourquoi ils proposent d'évaluer ces trois notions de concept issues de la psychologie cognitive. Sans donner les détails de leur argumentation, ce qui demanderait trop d'espace, il suffit de noter que Chater et Heyes considèrent que les théories définitionnelles et d'ensembles d'exemples ne peuvent être comprises dans un contexte non linguistique, ce pour quoi ils considèrent d'emblée qu'elles ne peuvent pas respecter le deuxième critère énoncé ci-haut. La théorie du prototype veut qu'un stimulus quelconque soit catégorisé, par exemple, comme étant un arbre ou un chat, si ce stimulus est suffisamment similaire à un certain prototype mental d'arbre ou de chat. Cette théorie du concept, contrairement aux deux dernières, peut être formulée facilement de manière non linguistique, car si les différents « tokens » d'une catégorie donnée peuvent être regroupés autour du prototype de manière perceptuelle, alors les catégories peuvent être assimilées sans qu'elles ne correspondent nécessairement à des termes du langage naturel.

Le problème majeur avec cette approche du prototype, c'est qu'elle ne permet pas de distinguer clairement les cas de véritable catégorisation conceptuelle de ceux de simple généralisation

de stimulus: «The predictions of the prototype and stimulus generalization views may be harder to distinguish than originally thought. [...] Thus, as a model of concepts to be used in comparative research, the prototype view has the weakness that it is not readily distinguishable from stimulus generalization accounts, which do not involve concepts at all<sup>14</sup>». Qu'est-ce, exactement, qu'une généralisation de stimulus ? Il s'agit simplement d'une discrimination fondée sur un mécanisme qui n'est sensible qu'à un stimulus particulier quelconque (par exemple, une longueur d'onde précise)<sup>15</sup>. Ainsi, un système cognitif quelconque réalise une généralisation de stimulus lorsqu'il est capable, à partir d'un stimulus X, d'identifier des objets et de réaliser une discrimination de ces objets à partir de ce même stimulus X. Or, comme nous le verrons plus tard, la capacité de généralisation d'un stimulus n'est pas suffisante pour justifier l'attribution de représentations conceptuelles. Bref, comme la notion de concept en tant que prototype ne semble pas, du moins selon Chater et Heyes, permettre de distinguer clairement les cas de catégorisation conceptuelle des simples cas de généralisation de stimulus, cette approche ne respecte par leur troisième critère, car elle ne permet pas d'investiguer précisément la possession de concepts chez les animaux non humains. C'est pourquoi, en fin de compte, les deux auteurs concluent que les théories des concepts fournies par la psychologie cognitive ne sont pas appropriées pour l'étude de la possession de concepts chez les animaux non humains<sup>16</sup>.

Chater et Heyes proposent alors de se tourner vers les théories philosophiques à propos du contenu des concepts. Évidemment, les approches qui lient intimement le contenu des représentations conceptuelles avec le langage naturel ne sont pas appropriées pour les spécialistes de psychologie comparative qui cherchent à déterminer le contenu des représentations mentales des animaux non linguistiques. De même, les théories de la signification holistiques, selon lesquelles la signification ou le contenu d'une représentation mentale particulière ne peut être découvert de manière indépendante des autres représentations (en ce sens que la signification ne peut qu'être rattachée qu'à l'ensemble du système représentationnel, et non aux parties le constituant), sont très problématiques, car la tâche

initiale qui consistait à déterminer si certains animaux possèdent certains concepts particuliers (« arbre », « chat », etc.) est remplacée par la tâche pratiquement insurmontable de découvrir le présumé système conceptuel entier des animaux. Autrement dit, « in looking for leads from philosophy the comparative researcher is looking for theories of content that are not holistic, and that do not make ineliminable reference to external natural language<sup>17</sup> ».

Chater et Heyes sont bien sûr conscients qu'il existe une panoplie de différentes théories du concept émergeant de la philosophie. Ils suggèrent toutefois de se pencher sur une approche en particulier, car elle semble de prime abord se prêter très bien à la recherche expérimentale en psychologie comparative : il s'agit de l'approche de la « sémantique informationnelle » (*informational semantics*), qui fonde la notion de concept sur la discrimination perceptuelle plutôt que sur des habiletés intrinsèquement linguistiques. La plus simple version de cette approche est la version corrélacionnelle, selon laquelle posséder un concept, c'est simplement être capable d'identifier les tokens de ce concept : « According to this view, having the X concept is simply a matter of being able perceptually to discriminate X's from non-X's ; and such discrimination abilities are just what paradigmatic animal concept experiments aim to test<sup>18</sup> ». Chater et Heyes notent que Fodor propose quelque chose du genre avec sa « crude causal theory », mais il n'est pas ici question d'évaluer cette affirmation.

Cette théorie du concept, telle quelle, est trop simpliste et présente une difficulté manifeste : elle ne permet pas de rendre compte des erreurs de catégorisation conceptuelle, pourtant très fréquentes, même chez les êtres humains. Par exemple, à cause de conditions extérieures qui affectent la perception (neige intense, noirceur, etc.), il peut très bien arriver qu'un être humain confonde une personne avec un petit arbre. Il semblerait cependant fort étrange de dire que pour cette raison, cet individu ne possède pas le concept d'arbre ou de personne. Cela, évidemment, n'est pas acceptable compte tenu du premier critère proposé par Chater et Heyes. On pourrait vouloir contourner cette difficulté en affirmant qu'il est important de distinguer les conditions optimales des conditions suboptimales

lorsqu'il est temps de catégoriser un stimulus perceptuel quelconque. Dans cette optique, ce serait seulement dans des conditions perceptuelles optimales qu'il est possible de fonder la possession d'un concept de X sur la distinction perceptuelle des tokens de X. Mais une présupposition fallacieuse est à la source de cette solution : pour pouvoir spécifier les conditions optimales en lien avec un concept quelconque, il faut d'abord connaître l'identité de ce même concept : «However, as Fodor (1990) forcefully points out, the notion of optimal viewing conditions is inescapably relative to the concept concerned<sup>19</sup>».

Par ailleurs, une autre solution possible pour sauver la théorie corrélationnelle serait de soutenir que la corrélation entre la possession d'un concept et la perception d'un token de ce concept n'est pas parfaite ; ainsi, un système pourrait posséder le concept de X s'il possède une structure mentale particulière qui s'active plus souvent lorsqu'il y a effectivement présence d'un token de X dans l'environnement extérieur. Mais selon Chater et Heyes, cette solution n'est pas viable, car elle implique une laxité beaucoup trop prononcée en ce qui concerne la détermination du contenu des concepts des sujets étudiés, ce qui ne s'accorde pas avec leur troisième critère. Bref, cette théorie corrélationnelle du concept n'est pas appropriée pour l'étude de la possession de concepts chez les animaux non linguistiques, car en voulant éviter la difficulté qui consiste à rendre compte des possibles erreurs de catégorisation conceptuelle, la théorie perd sa possibilité d'être appliquée pour la recherche expérimentale à cause de sa trop grande laxité quant à l'identification du contenu des représentations conceptuelles étudiées.

C'est pourquoi, après avoir montré en quoi ces différentes notions de concepts provenant de la psychologie cognitive et de la philosophie ne sont pas adéquates pour la recherche sur les concepts des animaux en ce qu'aucune ne semble satisfaire leurs trois critères, ils en concluent qu'il n'est, du moins pour l'instant, pas possible d'étudier véritablement la possession de concepts chez les animaux non linguistiques : «no clear sense has been provided for the claim that nonlinguistic animals have concept. [...] we simply do not know how to turn the claim that nonlinguistic animals have concepts into

an empirically substantive question<sup>20</sup>». À mon avis, cependant, cette conclusion pessimiste découle en grande partie du fait que les critères identifiés par Chater et Heyes sont très restrictifs, surtout lorsqu'il est question de l'exigence d'une similarité de contenu entre les représentations conceptuelles des animaux et celles des êtres humains. Néanmoins, ce ne sont pas tous les auteurs qui partagent ce scepticisme. Effectivement, certains ont proposé d'autres critères qui semblent redonner vigueur à l'idée qu'il soit possible d'étudier empiriquement l'attribution de concepts aux animaux non linguistiques. Dans la prochaine section, nous allons nous pencher sur la théorie de Newen et Bartels, en passant brièvement par l'étude des critères que suggère Collin Allen pour être en mesure de justifier l'attribution de concepts aux animaux non humains.

#### *4. L'attribution de concepts aux animaux non humains selon Allen et Newen & Bartels*

Dans leur texte de 1991, Allen et Hauser proposent deux expériences de pensée en lien avec le concept de la mort pour mettre de l'avant des considérations méthodologiques en ce qui concerne l'étude de la possession de concepts chez les animaux non humains. Selon eux, il est justifié d'attribuer des concepts aux animaux «if there is evidence supporting the presence of a mental representation that is independent of solely perceptual information<sup>21</sup>», et les deux expériences qu'ils suggèrent visent justement à fournir cette évidence. Toutefois, dans ce même texte, les deux auteurs ne se préoccupent pas vraiment des fondements théoriques qui permettent justement de conclure quoi que ce soit sur l'attribution de concept aux animaux non humains à partir de ces expériences. C'est dans son texte de 1999 qu'Allen se penche davantage sur cette importante question.

Cette idée d'indépendance face aux stimuli perceptuels semble récurrente lorsqu'il est question de la possession de concepts. Comme le notent Newen et Bartels, cette idée est pertinente, car elle permet d'écarter les cas de simples généralisations de stimulus, en ce sens que l'indépendance face aux stimuli implique l'intégration de plusieurs sources d'informations différentes (et non un seul stimulus bien

particulier) pour produire ou modifier un certain comportement<sup>22</sup>. De cela semble découler que les comportements qui sont produits à partir de représentations conceptuelles sont relativement flexibles, contrairement aux comportements rigides issus de généralisations d'un stimulus. Or, Allen tente de montrer, *contra* Chater et Heyes, qu'il est possible de rendre compte de l'attribution de concepts aux animaux non linguistiques s'il est possible de démontrer que ces animaux possèdent des représentations mentales indépendantes des stimuli perceptifs, dans le sens tout juste présenté.

Allen établit alors une liste de trois critères qui, sans prétendre définir la possession de concepts, fournit des conditions qui justifient l'attribution du concept de X à un organisme O: (i) O discrimine systématiquement certains tokens de X de certains tokens de non-X; (ii) O est en mesure de détecter lui-même certaines de ses erreurs de discrimination entre des tokens de X et des tokens de non-X et (iii) O est capable, à partir de sa capacité telle que décrite par le critère (ii), d'apprendre à améliorer sa capacité de discriminer les tokens de X des tokens de non-X<sup>23</sup>. Ces critères sont particulièrement intéressants, selon Allen, car les capacités qu'ils impliquent peuvent être étudiées empiriquement et identifiables dans des créatures non langagières: «they provide a basis for empirical investigation that allows us to resist the pessimistic conclusion drawn by Chater and Heyes<sup>24</sup>». Allen souligne que le critère (iii) est particulièrement important, car il fait le lien entre les deux premiers critères. En effet, le point sur lequel il veut mettre l'accent est que si la capacité évoquée dans le critère (ii), c'est-à-dire la capacité de détecter ses propres erreurs perceptuelles, est fortement liée avec la capacité (i) de catégorisation perceptuelle, cela montre en quelque sorte que l'organisme étudié possède un système représentationnel au sein duquel s'effectue une comparaison entre le contenu de l'input perceptuel de l'organisme et une représentation indépendante des stimuli de ce qu'est supposée représenter la perception. Cette dernière représentation indépendante des stimuli, selon Allen, est précisément le concept. Bref, pour Allen, c'est cette idée de standard de comparaison interne, «that represents the organism's world independently of its perceptual representation at any given moment<sup>25</sup>», qui justifie l'attribution de concepts chez

les animaux non humains, et qui permet donc d'éviter les difficultés impliquées par la relation étroite entre les concepts humains et le langage.

Newen et Bartels sont d'accord avec Allen qu'un aspect fondamental de la possession de concepts réside dans la capacité de former des représentations indépendantes des stimuli perceptuels immédiats. Toutefois, ils ne considèrent pas que la capacité de correction de ses propres erreurs de catégorisation soit nécessaire pour la possession de concepts<sup>26</sup>. Il ne sera pas ici question d'exposer en détail leurs critères pour la possession de concepts chez les animaux non humains tels que décrits dans leur texte. Je propose plutôt d'expliquer brièvement ce qui semble être le cœur de leurs intuitions. Pour ces deux auteurs, c'est l'extraction de classes à partir des données perceptuelles qui permet de fonder la notion de représentation conceptuelle. Ces classes sont formées selon certaines caractéristiques particulières instanciées par les données perceptuelles. Cela est possible seulement si le système étudié est capable de former une distinction entre l'objet et ses propriétés, ce qui est d'ailleurs de la plus grande importance, selon Newen et Bartels, en ce qui concerne la possession de concepts : «The animal is able to deal with a systematic object-property-representation and this seems to be the core basis for any conceptual attribution<sup>27</sup>». Un animal doit donc être capable de distinguer plusieurs propriétés d'un même objet et de former une représentation stable d'une propriété quelconque en étant confronté à plusieurs objets différents s'il possède une représentation conceptuelle de cette propriété.

Par ailleurs, pour être en mesure d'effectuer des catégorisations à partir de représentations de propriétés indépendantes des stimuli perceptuels immédiats, il faut que ces propriétés soient elles-mêmes inscrites à l'intérieur de ce que Newen et Bartels nomment un «filet sémantique<sup>28</sup>». En effet, la formation de classe, qui est au fondement des représentations conceptuelles, implique que la propriété à partir de laquelle la classe est formée puisse être contrastée avec des propriétés distinctes, qui forment justement le filet sémantique. Par exemple, pour qu'un organisme puisse posséder le concept ROUGE, il doit pouvoir se représenter la propriété d'être rouge dans un filet

sémantique formé par d'autres propriétés appartenant à la même dimension ou au même genre, c'est-à-dire la dimension de couleur. Ainsi, l'organisme doit être capable de se représenter la couleur rouge comme distincte de la couleur jaune ou de la couleur bleue. Autrement dit, les concepts sont constitués par les concepts qui lui sont contrastés: «It is constitutive for a concept that there are at least some relevant “contrasive concepts”, which are constitutive for each of the concepts<sup>29</sup>». Cette manière d'aborder les choses me semble plus valable que celle de Chater et Heyes, car que bien que la notion de filet sémantique ne soit pas indépendante de celle du contenu des représentations mentales, elle n'implique pas le présupposé inutilement restrictif (que Chater et Heyes, d'ailleurs, ne justifient pas explicitement) voulant que le contenu des présumées représentations conceptuelles des animaux devrait être le même que celui des humains.

À l'instar d'Allen, Newen et Bartels croient que leur manière de rendre compte de l'attribution de concepts peut être testée et appliquée pour des animaux non linguistiques. D'ailleurs, le fameux cas du perroquet gris d'Afrique Alex<sup>30</sup>, abondamment traité dans la littérature sur la cognition animale, semble illustrer parfaitement la possession de concepts selon Newen et Bartels. Brièvement, l'expérience consistait à présenter plusieurs objets de formes et de couleurs différentes à Alex. Alex devait ensuite répondre aux questions de Pepperberg, «Quelle couleur?» et «Quelle forme?», ce qui implique qu'Alex devait avoir appris certaines couleurs et formes particulières durant son entraînement. L'un des comportements les plus intéressants était qu'Alex était capable de répondre correctement à ces questions lorsqu'on lui présentait des objets qu'il n'avait jamais vus auparavant; il était donc impossible qu'Alex réponde par un simple mécanisme d'association entre un objet particulier et une couleur ou une forme particulière.

Mais là ne s'arrête pas l'expérience. Pour s'assurer qu'Alex comprenait vraiment les catégories «couleur» et «forme», Pepperberg a testé sa capacité à classer des objets selon les catégories «similaire» et «différent». Deux objets différents étaient présentés

à Alex, et il devait répondre à la question de savoir ce qui était différent ou similaire entre eux. Comme l'explique Pepperberg,

Alex would be presented with two objects that could differ with respect to three categories: color, shape, or material... He would then be queried "What's same?" or "What's different?" The correct response would be the label of the appropriate category... Therefore, to be correct, Alex would have to (1) attend to multiple aspects of two different objects; (2) determine, from a vocal question, if the response was to be based on similarity or difference; (3) determine, for the items he was shown, what was "same" or "different"... and then (4) to produce, vocally, the label for this particular category<sup>31</sup>.

Newen et Bartels pensent que le cas d'Alex illustre clairement qu'il possède des concepts, car ses comportements semblent montrer qu'il est capable de discerner plusieurs propriétés (forme, couleur, matériel, etc.) d'un même objet et d'identifier ces mêmes propriétés dans des objets différents. Les précautions méthodologiques appliquées par Pepperberg permettent aussi de conclure qu'Alex ne faisait pas que des associations ou des généralisations de stimulus, car il était capable de répondre aux questions lorsque lui étaient présentés des objets inconnus. Il y a donc clairement ici la manifestation d'une certaine indépendance aux stimuli immédiats perceptuels. Ainsi, tout semble indiquer qu'Alex était capable de former des classes à partir de propriétés indépendamment des données perceptuelles, ce qui permet d'affirmer, selon Newen et Bartels, qu'il possède des concepts.

Avant de conclure, je pense qu'il est pertinent de noter que tous ne s'entendent pas sur cette interprétation du cas d'Alex. Chater et Heyes, notamment, affirment que cette expérience, bien qu'elle soit impressionnante, ne peut que nous renseigner sur la capacité de discrimination, et non sur la possession de concepts, car le contenu des représentations mentales d'Alex reste, selon eux, indéterminé. À leur avis, rien ne peut assurer qu'Alex, par exemple, possède bel et bien le concept de «similaire» ou de «différent»: «Although Alex's differential responding to ["what's same?"] and ["what's

different ?”] may be taken to indicate that he has the concepts SAME and DIFFERENT, it is also consistent with the ascription of the concept pairs TWO and ONE, DOUBLE and SINGLE, DUPLICATED and NOT DUPLICATED<sup>32</sup>». De même, il est envisageable qu’Alex ne possède pas véritablement le concept de couleur, mais peut-être celui d’éclat ou de brillance (étant donné que les différentes couleurs ne reflètent pas toutes la lumière similairement). Autrement dit, Chater et Heyes pensent que ce genre d’expérience ne permet pas d’identifier comment l’animal se représente les stimuli qui lui sont présentés, ce pour quoi il n’est pas possible de conclure quoi que ce soit sur le contenu de leurs représentations. C’est pour cette raison qu’ils concluent que la question de l’attribution de concepts aux animaux non humains n’est pas éclairée par ce genre d’expérience, car ils considèrent que la question du contenu des concepts des animaux est tout à fait centrale. Personnellement, cette conclusion m’apparaît beaucoup trop forte, pour ne pas dire erronée, pour la simple et bonne raison que je ne vois pas en quoi l’identification du contenu exact des présumées représentations conceptuelles des animaux est nécessaire pour affirmer ou non que ces animaux possèdent bel et bien des concepts. Je pense que l’intuition de Newen et Bartels selon laquelle la possession de concepts requiert la formation d’un filet sémantique contrastant différentes représentations entre elles est très plausible, et cela ne nécessite pas l’identification précise des contenus des représentations conceptuelles.

## *5. Conclusion*

J’ai essayé de montrer que le problème de l’attribution de concepts aux animaux non humains est particulièrement difficile principalement à cause de l’existence d’un lien étroit entre la notion de concept et celle de langage. Davidson, nous l’avons vu, va même jusqu’à dire que ce lien est nécessaire, en ce sens que la possession de toute attitude propositionnelle requiert la possession du concept de croyance, ce concept nécessitant lui-même la possession du langage. Il est donc, de ce point de vue, exclu par principe que les animaux non linguistiques puissent posséder des concepts. Chater et Heyes, quant à eux, sont sceptiques quant à la possibilité d’attribuer des concepts

aux animaux non linguistiques, non pas à cause d'une liaison conceptuelle *a priori* entre la notion de concept et celle de langage, mais plutôt parce qu'ils considèrent qu'aucune notion de concept, qu'elle provienne de la psychologie cognitive ou de la philosophie, ne peut être appliquée à l'étude empirique de la possession de concepts chez les animaux non langagiers. Cela s'explique notamment parce qu'il semble très difficile, voire impossible, de déterminer le contenu des présumées représentations conceptuelles des animaux.

Allen veut justement surpasser cette apparente liaison entre les concepts et le langage en établissant trois critères pour justifier l'attribution de concepts aux animaux non linguistiques. Ces critères reposent sur l'idée que l'organisme doit avoir la capacité de détecter ses propres erreurs de catégorisation perceptuelle pour ainsi améliorer sa capacité de discrimination des stimuli pour pouvoir posséder des concepts. Cette manière de voir les choses implique que la discrimination à partir de représentations conceptuelles se fasse de manière indépendante des stimuli perceptuels immédiats, ce qui permet ainsi de distinguer la catégorisation conceptuelle de la simple généralisation de stimulus. Newen et Bartels défendent aussi l'importance de l'indépendance face aux stimuli perceptuels pour l'attribution de concepts. Toutefois, contrairement à Allen, ils ne pensent pas que c'est la capacité de détection des erreurs de classification qui est centrale pour la possession de concepts. Ils mettent plutôt de l'avant l'idée qu'une représentation conceptuelle implique la formation d'une classe à partir de la différenciation entre un objet quelconque et ses propriétés. C'est justement la représentation stable d'une propriété indépendamment de l'objet particulier qui l'instancie qui est primordiale pour la possession de concepts.

Finalement, nous avons vu que l'expérience de Pepperberg avec le perroquet Alex semblait montrer, du moins selon Newen et Bartels, qu'il possédait des concepts. Chater et Heyes rejettent cette conclusion sur la base de l'idée qu'il ne serait pas possible d'identifier clairement le contenu des représentations conceptuelles d'Alex. Je pense toutefois qu'ils vont trop loin en affirmant que ce genre d'expérience ne permet aucunement de conclure quoi que ce soit à

propos de la possession de concepts chez les animaux non humains. Je pense que le cas d'Alex, en conjonction avec les critères établis par Newen et Bartels, permet au moins d'affirmer que l'oiseau possède certaines représentations d'ordre conceptuel. Il est peut-être vrai que le contenu de ces représentations est difficilement identifiable, mais il n'en reste pas moins que cela n'affecte pas la conclusion qu'il soit justifié d'attribuer des représentations conceptuelles à Alex. Autrement dit, je ne vois pas pourquoi les difficultés rattachées à la détermination du contenu des représentations des animaux non humains devraient nous conduire automatiquement à rejeter la possibilité de leur attribuer des concepts. Il me semble que cette conclusion de Chater et Heyes mériterait d'être appuyée par une argumentation plus ample et plus convaincante pour qu'elle soit considérée sérieusement.

- 
1. N. Chater et C. Heyes, «Animal Concepts: Content and Discontent», dans *Mind & Language*, vol. 9, n° 3, 1994, pp. 209-246.
  2. A. Newen et A. Bartels, «Animal Minds and the Possession of Concepts» dans *Philosophical Psychology*, vol. 20, n° 3, 2007, pp. 283-308.
  3. Collin Allen, «Animal Concepts Revisited: the Use of Self-Monitoring as an Empirical Approach», dans *Erkenntnis*, vol. 51, 1999, pp. 33-40.
  4. Achim Stephan, «Are Animals Capable of Concepts?», dans *Erkenntnis*, vol. 51, 1999, p. 79.
  5. A. Newen et A. Bartels, *loc. cit.*, p. 284.
  6. Achim Stephan, *loc. cit.*, p. 79.
  7. C. Allen et Marc D. Hauser, «Concept Attribution in Nonhuman Animals: Theoretical and Methodological Problems in Ascribing Complex Mental Processes», dans *Philosophy of Science*, vol. 58, n° 2, 1991, p. 224.
  8. Donald Davidson, «Rational Animals», *Dialectica*, vol. 36, n° 4, 1982, p. 322.
  9. Achim Stephan, *loc. cit.*, p. 81.
  10. Donald Davidson, *loc. cit.*, p. 327.
  11. N. Chater et C. Heyes, *loc. cit.*, p. 210.
  12. A. Newen et A. Bartels, *loc. cit.*, p. 285.
  13. N. Chater et C. Heyes, *loc. cit.*, p. 214.
  14. *Ibid.*, pp. 218-219.

15. A. Newen et A. Bartels, *loc. cit.*, p. 287.
16. N. Chater et C. Heyes, *loc. cit.*, p. 221.
17. *Ibid.*, p. 223.
18. *Ibid.*, p. 224.
19. *Ibid.*, p. 225.
20. *Ibid.*, p. 237.
21. C. Allen et Marc D. Hauser, *loc. cit.*, p. 231.
22. A. Newen et A. Bartels, *loc. cit.*, p. 287.
23. Collin Allen, *loc. cit.*, p. 37.
24. *Ibid.*, p. 38.
25. *Ibid.*, p. 39.
26. A. Newen et A. Bartels, *loc. cit.*, p. 291.
27. *Ibid.*, p. 293.
28. *Ibid.*, p. 298.
29. *Ibid.*, p. 298.
30. I. Pepperberg, *The Alex Studies*, Cambridge, Harvard University Press, 1999.
31. *Ibid.*, p. 66.
32. N. Chater et C. Heyes, *loc. cit.*, p. 232.



# Les quatre critères de Newen et Bartels et l'empirisme de Jesse Prinz : deux théories des concepts compatibles

HUGO TREMBLAY, *Université Laval*

## 1. Introduction

Dans leur article «Animal Minds and the Possession of Concepts<sup>1</sup>», Albert Newen et Andreas Bartels proposent quatre critères afin de définir en quoi consiste la possession d'un concept. Mais plutôt que de se prononcer sur la nature métaphysique des concepts, les auteurs soutiennent une théorie épistémique selon laquelle les «concepts sont des types de représentations mentales constituées par des capacités épistémiques<sup>2</sup>», ce qui n'implique pas pour autant que les concepts *soient* des capacités épistémiques. Leur théorie a pour principal avantage d'établir des critères de possession qui ne sont, d'une part, pas trop faibles. Lorsqu'on fait appel à la possession de concepts pour expliquer un comportement, il s'agit de la meilleure explication disponible ; il serait moins approprié (voire impossible) d'expliquer ce comportement par des capacités plus simples, par des capacités non cognitives ou non conceptuelles par exemple. Et d'autre part, ces critères ne sont pas trop spécifiques aux humains ; nous n'excluons pas d'emblée que les animaux puissent avoir une pensée conceptuelle<sup>3</sup> même s'ils ne disposent pas d'un langage naturel<sup>4</sup>.

Bien que les critères épistémiques proposés par Newen et Bartels soient justes et extrêmement pertinents, nous croyons que cette théorie gagne à être combinée avec une autre théorie des concepts, une théorie se prononçant quant à elle sur la nature métaphysique des concepts : la théorie empiriste des concepts de Jesse Prinz<sup>5</sup>. Il semble en effet que la mise en commun des deux théories permette d'expliquer plus précisément en quoi la possession de concepts implique les capacités épistémiques décrites par Newen et Bartels.

Ce texte aura ainsi un double objectif. Il visera en premier lieu à démontrer comment les quatre critères proposés par Newen et Bartels<sup>6</sup> rendent correctement compte des capacités épistémiques impliquées par la possession de concept — ce que nous ferons en reconstruisant les arguments de Newen et Bartels contre certaines théories des concepts: les théories de Allen et Hauser, Davidson et Dretske. Puis il cherchera par la suite à montrer comment cette théorie épistémique est compatible avec la théorie empiriste de Prinz, et surtout, comment cette dernière permet de mieux comprendre la façon dont les concepts se manifestent dans l'esprit des animaux, et plus particulièrement, dans ceux de Rico le border collie<sup>7</sup> et d'Alex le perroquet<sup>8</sup>.

## *2. La théorie épistémique des concepts de Newen et Bartels*

Une fois le cadre méthodologique de leur théorie exposé — à savoir, il sera principalement question de définir les capacités épistémiques impliquées par la possession de concepts<sup>9</sup> —, Newen et Bartels procèdent à un bref survol de certaines des théories des concepts<sup>10</sup>. Mais avant tout, il importe de mettre en évidence le défi commun auquel font face ces théories.

Lorsque ces théories définissent les critères permettant d'identifier un comportement causé par une représentation conceptuelle, il ne faut pas que ces critères incluent des comportements causés par la généralisation d'un stimulus plutôt que par une véritable représentation conceptuelle. En d'autres mots, il ne faut pas confondre une réaction découlant d'une généralisation de stimuli<sup>11</sup> (*GS*) et une action découlant d'une représentation conceptuelle (*RC*), où :

*GS*: Perception de stimuli → Réaction

*RC*: Perception de stimuli → Représentation conceptuelle  
appropriée → Action

Inversement, il ne faut pas non plus que ces critères soient trop exigeants ou influencés par les capacités humaines au point où ils excluraient certains comportements pourtant véritablement causés par des représentations conceptuelles. Le défi est de taille, car il

faut que ces critères puissent distinguer les types de représentations en fonction des résultats d'observations empiriques, et ce, sans que ces observations fassent nécessairement appel à des compétences linguistiques — du moins, si l'on veut éviter d'exclure d'emblée tous les animaux à l'exception des êtres humains. Une fois ce défi explicité, Newen et Bartels exposent les critères établis par différentes théories des concepts et ils démontrent en quoi ces critères ne répondent pas à ce défi.

### *2.1 Critiques de certaines théories des concepts*

La première théorie examinée est celle de Fred Dretske<sup>12</sup>. En résumé, Dretske soutient qu'un animal fait preuve d'une pensée conceptuelle lorsqu'il y a corrélation causale entre, d'une part, certains de ses comportements, et d'autre part, la perception de stimuli mémorisés et associés à certaines représentations d'objets du monde. Par exemple, un oiseau perçoit un papillon orange, puis, en raison d'expériences passées désagréables qu'il a mémorisées et associées aux papillons orange, il évite de manger ce papillon, alors qu'il mange les papillons d'autres couleurs. En fonction d'un tel comportement, Dretske soutient que l'oiseau *croit* que les papillons orange lui sont néfastes — c'est une croyance qu'il a apprise par ses expériences passées<sup>13</sup> —, et conséquemment, nous pouvons conclure que l'oiseau possède le concept de « papillon orange ».

Mais un tel critère est-il suffisamment exigeant? Bien qu'un oiseau ait appris à réagir d'une façon particulière lorsqu'il perçoit le stimulus d'un papillon orange, pouvons-nous nécessairement conclure que cet oiseau a une représentation conceptuelle du papillon orange? Il semble en fait qu'il soit possible d'expliquer ce comportement par une capacité de représentation non conceptuelle, comme la généralisation de stimulus<sup>14</sup>. Par opposition, on peut s'attendre à ce que la représentation conceptuelle implique une certaine indépendance face à la perception immédiate d'un seul stimulus; une indépendance qui permet à l'animal de modifier ses représentations conceptuelles si certains stimuli supplémentaires pertinents entrent en ligne de compte. La représentation conceptuelle permet ainsi une réponse relativement flexible et changeante face

à un seul stimulus, car elle ne se limite pas nécessairement à une réaction immédiate à la simple perception de ce stimulus.

Pour corriger cette insuffisance et correctement rendre compte du critère de flexibilité, Newen et Bartels se penchent du côté de la théorie de Colin Allen et Marc Hauser<sup>15</sup>. Leur théorie soutient qu'on peut attribuer un concept à un animal seulement dans les cas où la «représentation mentale est indépendante d'une information strictement perceptuelle<sup>16</sup>». Pour Newen et Bartels, cela revient en quelque sorte à appliquer le critère de *plasticité informationnelle* de Pylyshyn<sup>17</sup> et à soutenir qu'une représentation conceptuelle se doit d'être construite à partir *de plus d'une* source d'informations<sup>18</sup>. Leur conclusion à ce sujet: «bien que nous admettons que l'indépendance face à un seul stimulus, au sens où la sensibilité d'un système cognitif est liée à plus d'un seul stimulus, est une condition *nécessaire* pour posséder un concept, nous argumenterons qu'il ne s'agit pas d'un critère *suffisant*<sup>19</sup>». Et selon eux, les tests faits avec le border collie Rico<sup>20</sup> permettent de démontrer la raison d'être de cette affirmation. Leur position se résume ainsi. Bien que Rico soit apte à reconnaître et aller chercher des objets spécifiques après avoir entendu un mot désignant ces objets — son action n'est donc pas dépendante d'un seul stimulus, mais de la *combinaison appropriée* d'informations auditives et visuelles — cela n'implique pas qu'il ait la capacité de classer ces objets en fonction de catégories conceptuelles. En effet, les mots auxquels répond Rico représentent des objets uniques, et non pas des catégories conceptuelles (un ensemble d'objets partageant tous une même propriété<sup>21</sup>). Les auteurs touchent la cible: il semble bien qu'un concept doit pouvoir représenter quelque chose de l'ordre d'une catégorie conceptuelle.

Pendant, cette considération introduit un nouveau problème: peut-on former de telles catégories sans disposer d'un langage naturel? Ou plus précisément: peut-on discerner à quel moment nos représentations conceptuelles représentent adéquatement ou non des catégories sans langage naturel? Ces questions ne sont pas sans rappeler un débat dont l'une des positions les plus connues est celle de Donald Davidson. En résumé, l'argument de Davidson est le suivant: pour être apte à posséder un concept, il faut au moins

comprendre que nous pouvons incorrectement représenter des objets du monde par ce concept. Par exemple, je crois voir passer un chien alors qu'il s'agit en vérité d'un chat, et lorsque je prends conscience qu'il s'agit d'un chat — l'animal s'est rapproché et je le vois maintenant mieux —, je comprends que ma représentation conceptuelle initiale — celle d'un chien — était incorrecte. Toutefois, pour qu'une telle erreur soit possible, il faut que j'aie plusieurs croyances à propos de ce qui caractérise un objet tombant sous un certain concept, sans quoi, ma représentation initiale ne peut pas impliquer des caractéristiques différentes de celles de cet objet<sup>22</sup>. Et pour Davidson, il est impossible de former des croyances — ou toute autre attitude propositionnelle — sans disposer du langage<sup>23</sup>.

Cela dit, on peut être en accord avec Davidson sur le fait que la possession de concepts présuppose la possibilité de discerner les cas de représentations inappropriées sans pour autant conclure à la nécessité de devoir posséder des croyances ou le langage pour qu'un tel discernement soit possible. C'est la stratégie adoptée par Newen et Bartels<sup>24</sup>. Mais alors, si la possession de croyances ou du langage n'est pas ce qui permet d'expliquer comment un animal est capable de discerner des cas de représentations inappropriées, comment peut-on rendre compte de cette capacité ?

Dans « Animal Concept Revisited<sup>25</sup> », Colin Allen soutient qu'il est possible d'identifier trois critères n'impliquant pas le langage et qui permettent néanmoins d'attribuer<sup>26</sup> la compréhension du concept de X à un organisme O<sup>27</sup> :

1. O discrimine systématiquement les X des non-X ;
2. O est capable de détecter certaines de ses propres erreurs de discrimination entre des X et des non-X ;
3. En conséquence de la capacité (2), O est capable d'apprendre à mieux discriminer les X et des non-X.

De ces trois critères, le deuxième tient manifestement compte de l'aspect normatif des concepts. Ce qui est alors intéressant pour Allen, c'est qu'il est « possible pour les animaux de faire preuve d'une capacité à détecter des erreurs sans avoir recours à des indices

extérieurs; ils peuvent démontrer qu'ils ont la capacité d'autoévaluer leurs propres performances<sup>28</sup>». Il donne l'exemple de cochons qui auraient fait preuve d'une telle capacité<sup>29</sup>.

Mais pour Newen et Bartels, la façon dont Allen explique la capacité des animaux à répondre à de tels critères — par l'existence d'un standard de comparaison interne indépendant des représentations perceptuelles<sup>30</sup> — est toutefois trop exigeante. La représentation conceptuelle suppose bel et bien la possibilité de percevoir que nous avons représenté un objet du monde par un concept de façon inappropriée, mais cette possibilité n'implique pas la nécessité d'un standard de comparaison interne permettant d'identifier et de corriger ces erreurs de façon systématique. La possession d'un tel système de correction est peut-être un critère *suffisant* pour posséder un concept, mais il ne s'agit pas d'un critère *nécessaire*<sup>31</sup>. Le véritable critère nécessaire à ce sujet est plutôt le suivant: un système cognitif doit être apte à modifier un concept représenté lorsque des stimuli indépendants du stimulus ayant causé la représentation apportent des informations supplémentaires permettant de modifier la représentation initiale. La précision est pertinente, puisque lorsque nous établirons les critères nécessaires pour définir les capacités épistémiques impliquées par la possession de concepts, il faudra s'assurer de couvrir l'aspect normatif des concepts, sans pour autant inclure la nécessité d'un standard de correction interne.

## *2.2 Quatre critères épistémiques nécessaires*

Avant d'établir explicitement les quatre critères qui découlent des précédentes analyses, Newen et Bartels présentent le cas d'Alex le perroquet<sup>32</sup> afin d'examiner un exemple concret de comportement animal faisant preuve des quatre capacités épistémiques recherchées<sup>33</sup>. La première chose que nous pouvons conclure de cet examen est que la représentation conceptuelle implique la capacité « d'identifier et réidentifier des objets et des propriétés<sup>34</sup> ». Articulée de façon un peu plus précise, cette capacité peut en fait se subdiviser en deux capacités plus spécifiques<sup>35</sup>; un système possède une représentation conceptuelle d'une propriété si :

- C1*: Cette même propriété peut être identifiée dans plusieurs objets ;
- C2*: Un même objet peut être perçu comme n'ayant pas seulement cette propriété, mais aussi d'autres propriétés distinctes.

En distinguant ainsi *C1* et *C2*, on s'assure qu'un concept puisse représenter une propriété spécifique que l'on retrouve dans un ou plusieurs objets, sans que l'objet dans sa totalité soit toujours représenté comme une seule propriété. Ce sont là deux capacités manifestement possédées par Alex, puisque lorsqu'on présente un ensemble d'objets à Alex — par exemple, un plateau sur lequel se trouvent trois cubes en bois rouge et deux triangles en feutre bleu —, il est possible de le questionner sur différentes propriétés de ces objets — par exemple, «Quelle couleur?», «Quelle forme?» ou «Quel matériau?» — et Alex est apte à donner les bonnes réponses à ces différentes questions (Alex est un perroquet et il peut donc produire des sons distincts afin de formuler ses différentes réponses<sup>36</sup>).

Mais si *C1* et *C2* sont des critères nécessaires, ils ne sont pas suffisants. Suite aux critiques des théories précédemment examinées, nous savons que des critères nécessaires supplémentaires se doivent d'être pris en compte afin d'éviter d'inclure des comportements qui s'expliquent en fait par d'autres mécanismes que la représentation conceptuelle. Tel que nous l'avons expliqué précédemment, les critères ne doivent pas avoir pour conséquence d'inclure des cas de généralisation de stimulus. Les comportements qui découlent de la représentation conceptuelle doivent être relativement flexibles et indépendants de la simple perception immédiate d'un seul stimulus. L'animal doit être capable de se représenter un concept et de modifier cette représentation si certains stimuli supplémentaires pertinents entrent en ligne de compte. Alex possède cette capacité car il ne réagit pas immédiatement à la simple perception visuelle d'un objet, c'est en fait la question qu'on lui pose — un son qu'il perçoit —, combiné à ce qu'il voit, qui lui permettent de comprendre la propriété qu'il doit se représenter<sup>37</sup>. Nous avons donc ainsi<sup>38</sup> :

C3: La représentation d'une propriété est relativement indépendante d'un seul stimulus ;

Mais encore une fois, si nous avons là un critère nécessaire, nous n'avons toujours pas couvert l'ensemble de ceux-ci. En effet, nous avons illustré une insuffisance (apparente) avec le cas de Rico : il faut que l'animal soit apte à classer ses représentations en fonction de *dimensions appropriées*. Pour sa part, Alex démontre qu'il a cette capacité lorsqu'en fonction des objets qui se trouvent devant lui, il répond par exemple «rouge», «bleu» ou «jaune» à la question «Quelle couleur?», puis «carré», «rond» ou «triangle» à la question «Quelle forme?». Par ailleurs, Alex est aussi capable d'identifier en fonction de quelle catégorie — la forme, la couleur, le matériau — les objets devant lui sont pareils ou différents. Il semble bien que ses réponses démontrent que les concepts auxquels il fait appel sont classés en fonction de catégories appropriées<sup>39</sup>. Ceci nous donne le dernier critère nécessaire<sup>40</sup> :

C4: La propriété est représentée comme une instance d'une catégorie particulière et appropriée. Elle appartient à un réseau de concept minimal.

Notons finalement que lorsqu'on considère l'ensemble de ces capacités, nous pouvons expliquer comment un animal peut modifier une de ses représentations inappropriées sans faire appel au langage ou à un standard de correction interne. En effet, par C1 et C2, un animal peut se représenter une propriété particulière, puis par C3, il est apte à modifier sa représentation dans l'éventualité où une information supplémentaire lui indiquerait que l'objet perçu ne correspond finalement pas au concept auquel il a été associé, par C4.

À ce point, il semble bien que nous ayons correctement identifié les capacités épistémiques découlant de la possession de concepts. Du moins, nous avons tenté de démontrer que les arguments en faveur de ces quatre critères sont convaincants. Mais même si ces capacités faisaient consensus, cela ne serait pas suffisant pour clore le débat

sur la nature des concepts. Nous n'avons en effet jusqu'à maintenant rien dit sur la nature métaphysique des concepts. C'est maintenant à cette question que nous nous intéresserons, car il nous semble que la théorie empiriste de Jesse Prinz offre les assises nécessaires à de telles capacités épistémiques. En fait, elle permet même de mieux les comprendre.

### *3. La théorie empiriste des concepts de Jesse Prinz*

Jesse Prinz est un ardent défenseur d'une théorie empiriste des concepts<sup>41</sup>. Il conçoit ses travaux sur ce sujet comme une extension des travaux de Lawrence Barsalou, psychologue et chercheur en science cognitive<sup>42</sup>, et Prinz fait régulièrement référence à ses travaux afin de supporter sa théorie par différentes observations empiriques<sup>43</sup>.

De façon quelque peu simplifiée, la théorie des concepts de Prinz se définit de la façon suivante : « Tous les concepts (humains) sont des copies ou des combinaisons de copies de représentations perceptuelles<sup>44</sup> ». En d'autres mots, nos concepts sont intrinsèquement liés à nos expériences perceptuelles passées et c'est à partir de ces expériences que nous pouvons nous représenter conceptuellement le monde. Mais pour Prinz, la représentation conceptuelle n'a pas pour première fonction de permettre le raisonnement théorique, elle est plutôt liée à l'action : elle a pour rôle de nous permettre d'agir plus efficacement dans un monde complexe et en constant changement<sup>45</sup>. En résumé, la pensée conceptuelle nous permet de représenter nos expériences perceptuelles passées, ce qui a pour conséquence d'améliorer l'efficacité de nos actions.

Mais au-delà du lien entre représentation conceptuelle et expérience perceptuelle, la théorie de Prinz implique aussi la capacité de représenter des catégories d'objets du monde à partir de constructions de notre mémoire de travail<sup>46</sup> (*working-memory*) ; les concepts représentent des catégories mentales d'objets, et « penser à une catégorie, pour un empiriste, c'est simuler la manifestation de cette catégorie à l'aide de nos sens<sup>47</sup> » dans notre mémoire de travail. Pour Prinz, les catégories s'expliquent ainsi : la première fois que nous rencontrons une instance d'un objet que nous ne connaissons

pas, une catégorie est créée dans notre mémoire à long terme — Prinz compare cela à une base de données<sup>48</sup>. Cette catégorie est réactivée et représentée de façon généralement fiable chaque fois que nous rencontrons ou nous imaginons une instance de cette catégorie — soit un objet ayant certaines propriétés (*features*) appartenant à cette catégorie<sup>49</sup> —, ce qui nous permet par le fait même de modifier cette catégorie lorsque nous constatons, par exemple, que les objets associés à une catégorie possèdent ou ne possèdent pas, finalement, certaines propriétés<sup>50</sup>. Ceci implique qu'en fonction du contexte dans lequel nous sommes et des propriétés que nous percevons ou imaginons, un même concept peut être constitué de plusieurs représentations relativement différentes<sup>51</sup> — par exemple, si vous vous trouvez dans un restaurant et qu'on vous demande si vous voulez du canard, il est probable que votre représentation du canard à ce moment sera différente de celle que vous aurez si vous vous promenez dans un parc et qu'on vous demande si vous avez vu le canard dans l'étang<sup>52</sup>.

Notons finalement que cette théorie se prononce seulement sur le *format* de représentation de ces catégories, et ce faisant, elle n'impose pas que les informations soient encodées d'une façon particulière, alors que c'est précisément ce que cherchent à démontrer d'autres théories des concepts comme la théorie des prototypes, la théorie des modèles et la théorie des théories. Il s'ensuit que :

L'on pourrait avoir une version empiriste de chacune de ces théories pour laquelle l'information associée serait encodée sous un format perceptuel. Conséquemment, l'empirisme n'est pas en compétition avec ces autres approches. Les empiristes ne commettent pas l'erreur de soutenir que seule leur théorie mérite d'être appelée une théorie des concepts alors que les autres théories seraient dans l'erreur<sup>53</sup>.

En résumé, Prinz soutient que sa théorie implique les quatre principes fondamentaux suivants<sup>54</sup> :

1. Les concepts représentent des catégories via des causations étiologiques et nomologiques<sup>55</sup> ;

2. Les concepts sont des constructions variables de la mémoire de travail ;
3. Les concepts sont construits à partir de modalités sensorielles spécifiques dont la mémoire a conservé des traces ;
4. Les concepts et les divisions de base à partir desquelles ils sont organisés sont tous appris (plutôt qu'innés).

### *3.1 L'application aux animaux*

Dans ses différents travaux sur le sujet, Prinz ne s'intéresse pas spécifiquement à l'application de cette théorie aux animaux. Mais lorsqu'on s'y arrête, rien n'empêche une telle application ; en fait l'extension de cette théorie aux animaux s'avère même naturelle.

D'une part, il est clair que pour Prinz, cette théorie n'implique aucunement que le langage soit une condition nécessaire à la possession de concepts<sup>56</sup>. D'autre part, les quatre principes que l'on vient d'énumérer peuvent très bien être respectés par au moins quelques espèces d'animaux (pensons aux primates et à d'autres mammifères). En effet, pour autant que certains animaux disposent de facultés qui s'apparentent à la mémoire à long terme — c'est là une possibilité assez peu controversée, pensons à Rico qui peut retenir des mots pour au moins quatre semaines<sup>57</sup> — et à la mémoire de travail — ce que semblent impliquer certaines études<sup>58</sup> —, et pour autant que les perceptions de ces animaux puissent être traitées par ces deux types de mémoire, alors rien n'empêche qu'ils puissent former des concepts. Cela n'implique pas nécessairement qu'ils le peuvent, mais ils disposent à tout le moins des facultés minimales pour y parvenir.

Par ailleurs, dans la continuité de Barsalou<sup>59</sup>, Prinz reconnaît que l'hypothèse selon laquelle les concepts sont construits à partir de modalités sensorielles a l'avantage d'impliquer une continuité dans l'évolution du cerveau des animaux :

Au début, les créatures étaient des machines de type «entrées-sorties». Le monde causait des stimulations sensorielles, et ces stimulations causaient des réponses programmées. [...] Avec le temps, l'évolution a entraîné chez certaines créatures la capacité d'enregistrer des informations

perceptuelles à propos des objets qu'ils rencontraient et des conséquences qui suivaient la rencontre de ces objets. [...] Finalement, l'évolution a entraîné la capacité de réactiver les informations perceptuelles enregistrées, même en l'absence de stimulations sensorielles, et la capacité de manipuler ces informations dans la mémoire de travail. C'est ce que l'humain fait, et il le fait mieux que n'importe quelle autre créature sur terre<sup>60</sup>.

Bref, rien n'empêche que la théorie de Prinz s'applique aux animaux. Il s'agit même d'une implication probable. Si les humains sont doués d'une pensée conceptuelle beaucoup plus développée que celles des autres animaux — une pensée permettant un raisonnement purement théorique par moment —, cela n'empêche pas que certains animaux peuvent être doués d'une pensée conceptuelle plus rudimentaire : une pensée rendant possible un certain raisonnement pratique qui donne aux animaux un moyen efficace de s'adapter au monde qui les entoure.

#### *4. Établir le lien entre les deux théories*

En admettant que la théorie des concepts de Prinz soit avérée, nous sommes maintenant en mesure de mieux comprendre comment la possession de concepts entraîne les capacités épistémiques identifiées par Newen et Bartels. Il faut en effet bien voir que les deux théories ne sont pas en opposition ; elles sont tout à fait compatibles. Elles visent à rendre compte de caractéristiques différentes qui ne sont pas mutuellement exclusives. Avec la théorie de Prinz, nous avons un postulat métaphysique sur la nature des représentations conceptuelles : les concepts sont des représentations de catégories, ils sont construits dans notre mémoire de travail et leur format est perceptuel. Alors qu'avec la théorie de Newen et Bartels, la stratégie consiste à ne pas se prononcer sur la nature métaphysique des concepts, mais plutôt à rendre compte des capacités épistémiques impliquées par la possession de concepts. En d'autres mots, cette dernière stratégie est indirecte : peu importe la façon dont les concepts existent dans l'esprit des animaux, si ces animaux sont aptes à effectuer certaines « tâches épistémiques » — où ces tâches

consistent à respecter les quatre critères épistémiques définis par les auteurs — nous pouvons conclure que leur comportement découle d'une représentation conceptuelle.

Pour faire le lien entre ces deux théories, nous ferons appel à l'analyse de Newen et Bartels quant aux comportements de Rico le border collier et d'Alex le perroquet. En ce qui concerne les capacités épistémiques de Rico, Newen et Bartels expliquent pourquoi Rico possède *C3* (indépendance relative d'une représentation face à un seul stimulus) et qu'il ne possède pas *C4* (réseau sémantique minimal), mais ils ne développent pas sur *C1* et *C2* (identification d'une même propriété dans plusieurs objets et de plusieurs propriétés dans un même objet). Ils reconnaissent qu'il est plausible que Rico possède *C1* et *C2*, mais ils ne se prononcent pas sur ces capacités, car les tests menés avec Rico ne permettent pas de le faire<sup>61</sup>. En effet, les tests à partir desquels nous évaluons ses capacités supposent que la propriété conceptualisée par Rico représente toujours un seul objet dans sa totalité, et non pas une partie d'un objet qui pourrait être retrouvée dans plusieurs objets distincts. Ainsi, pour expliquer comment l'empirisme de Prinz rend compte des capacités *C1* et *C2*, il sera préférable d'utiliser le cas d'Alex le perroquet.

Mais avant de s'attaquer à l'explication de *C1* et *C2*, commençons par expliquer comment nous pouvons rendre compte de *C3* et *C4* chez Rico. Si pour les auteurs le comportement de Rico ne permet pas de conclure qu'il agit suite à une représentation conceptuelle, en raison de son incapacité à répondre correctement à *C4*, voyons en quoi la théorie de Prinz nous permet d'interpréter la chose de façon différente et de reconnaître que Rico fait preuve d'une pensée conceptuelle.

#### *4.1 Le critère C3 : indépendance relative d'une représentation face à un seul stimulus*

Le premier critère que nous permet de bien comprendre la théorie de Prinz est le critère *C3* : la représentation conceptuelle d'une propriété est relativement indépendante d'un seul stimulus. Nous l'avons vu, il faut éviter de confondre une réaction qui découle

d'une généralisation de stimuli (*GS*) d'une action qui découle d'une représentation conceptuelle (*RC*):

*GS*: Perception de stimuli → Réaction

*RC*: Perception de stimuli → Représentation conceptuelle  
appropriée → Action

Dans le cas de Rico, Newen et Bartels expliquent que le chien répond à ce critère puisque son comportement n'est pas déclenché par un seul stimulus : il ne va pas chercher un objet à la simple perception visuelle de celui-ci, et il ne va pas non plus chercher n'importe quel objet dès qu'il entend un mot. L'action de Rico est plutôt déterminée par la *combinaison appropriée* d'informations visuelles et auditives. Rico entend premièrement un mot — il traite une information auditive — et il cherche par la suite l'objet approprié — il traite une information visuelle en tenant compte de l'information auditive précédente. Il fait donc preuve de la capacité épistémique *C3*<sup>62</sup>.

Comment expliquer ce qui se passe dans l'esprit de Rico à la lumière de la théorie de Prinz? La première remarque d'intérêt est que dans les tests documentés, Rico ne voit pas l'objet qu'on lui nomme; lorsqu'il entend un mot désignant un objet, Rico se trouve dans une autre pièce que la pièce où se trouve l'objet qu'il doit chercher<sup>63</sup>. On peut donc supposer que lorsque Rico entend le mot, une certaine représentation perceptuelle de ce mot construite à partir de ces expériences passées lui vient à l'esprit — à quoi l'objet associé à ce son (le mot prononcé par son maître) ressemble-t-il visuellement, etc. — ce qui permet ensuite à Rico de chercher l'objet qui correspond à cette représentation perceptuelle. Dans les cas où Rico entend un mot qu'il ne connaît pas, aucune représentation adéquate n'est liée à ce son (le mot). Mais lorsqu'il entre dans la pièce et constate éventuellement qu'un seul des objets n'est pas associé à une représentation déjà connue, une nouvelle catégorie est alors créée et associée à ce son et aux propriétés perçues du nouvel objet<sup>64</sup> — ce à quoi il ressemble visuellement, etc. Cette nouvelle catégorie pourra à partir de ce point faire l'objet d'une représentation conceptuelle.

#### *4.2 Le critère C4: nécessité d'un réseau sémantique minimal*

Pour Newen et Bartels, si l'on s'en tient aux tests conduits avec Rico, rien n'indique que ce chien soit apte à classer les mots qu'il entend et les objets que ces mots représentent en fonction de dimensions correctement organisées dans un réseau sémantique minimal; en d'autres mots, rien n'indique qu'il possède C4. Conséquemment, puisqu'il s'agit d'un critère nécessaire, nous ne pouvons pas conclure que Rico est doué d'une pensée conceptuelle. À strictement parler, la conclusion des auteurs est adéquate: sur la base des tests faits avec Rico, nous ne pouvons pas conclure que Rico possède C4. Néanmoins, à la lumière de la théorie de Prinz et d'une recherche menée avec une autre border collie, Chaser<sup>65</sup>, nous pouvons spéculer sur ce qui se produit dans l'esprit de Rico et voir en quoi il pourrait être doué d'une pensée conceptuelle.

Nous l'avons vu dans la section 4.1, en fonction de la théorie de Prinz, il semble plausible de croire que Rico répond au critère C3: lorsqu'il entend un mot qu'il connaît, il se représente perceptuellement l'objet en question dans sa mémoire de travail et il peut ensuite chercher cet objet. Mais dans les tests sur lesquels se sont basés Newen et Bartels, les objets représentés sont tous des objets uniques du monde; lorsque Rico se représente perceptuellement un objet suite à l'audition d'un mot, il n'existe qu'un seul objet qui correspond à cette représentation. Ce que nous pouvons conclure à partir de ces tests, c'est que les représentations que peut avoir Rico s'apparentent plus à la référence des noms propres — elles renvoient à des objets uniques — plutôt qu'à des catégories conceptuelles représentant une pluralité d'objets du monde partageant une même propriété. Mais bien que ces expériences ne le démontrent pas, cela n'empêche pas qu'à partir de cette capacité — celle de représenter ce à quoi réfèrent des noms propres —, Rico peut avoir la capacité de former un réseau sémantique minimal. En effet, les représentations qu'il peut avoir n'ont pas forcément à être limitées à des objets uniques du monde. D'ailleurs, Newen et Bartels mentionnent qu'en fonction d'observations qui n'ont pas fait l'objet d'une publication, il semble que «Rico était capable d'apprendre à catégoriser les balles comme une classe d'objets, même s'il ne semble pas faire appel à cette

habileté plus complexe lorsqu'il identifie des objets uniques<sup>66</sup>). En fait, dans une étude publiée ultérieurement sur une autre border collie, Chaser, il fut précisément démontré que Chaser était capable d'identifier autant des objets uniques — un total de 1022 objets<sup>67</sup> — que des catégories d'objets : les jouets, les balles et les frisbees<sup>68</sup>.

Bref, bien que les tests faits avec Rico ne démontrent pas qu'il a la capacité *C4*, il semble bien que la capacité de se représenter des objets uniques en référence aux mots qu'il entend puisse être interprétée comme un indice qu'il est possible pour l'animal, s'il est entraîné adéquatement, de développer un réseau sémantique minimal dans lequel certains concepts représentent des catégories d'objets et non pas seulement des objets uniques.

#### *4.3 Les critères C1 et C2 : une même propriété dans plusieurs objets/plusieurs propriétés dans un seul objet*

Lorsqu'on présente un ensemble d'objets à Alex — par exemple, un plateau sur lequel se trouve un cube rouge en bois —, il est possible de questionner Alex sur différentes propriétés de ce même objet — par exemple, «Quelle couleur?», «Quelle forme?» ou «Quel matériau?» — et Alex est apte à donner les bonnes réponses à ces différentes questions. Il peut aussi correctement identifier une même propriété dans plusieurs objets.

Dans ces cas, nous pouvons supposer que lorsqu'Alex entend une information sonore — par exemple, «Quelle couleur?» — et qu'il perçoit visuellement la couleur de l'objet devant lui — le cube est rouge — ses expériences passées lui permettent de se représenter que parmi les informations visuelles perçues, l'effet de rougeur perçu est une des instances appropriées de la catégorie associée au son «couleur». Mais Alex a aussi appris à reconnaître l'effet de rougeur comme une instance d'une catégorie plus spécifique, une catégorie associée au son «rouge»; il peut donc répondre par ce mot. De la même manière, si Alex avait plutôt entendu le son «Quelle forme?», Alex aurait pu se représenter que la forme carrée du cube est une des instances appropriées de la catégorie associée au son «forme». Et il aurait aussi pu répondre par le son «carré», car ce son est

associé à cette catégorie plus spécifique dont la forme perçue est une instance appropriée<sup>69</sup>.

Bref, plusieurs propriétés différentes peuvent être distinguées dans un même objet, car selon ce sur quoi notre attention se porte (ce qui varie en fonction des contextes dans lesquels nous sommes ; dans le cas d'Alex, son attention varie en fonction du mot entendu), un objet peut tomber sous plusieurs représentations, en fonction de diverses catégories. Et une fois que l'on possède un concept représentant une propriété, il est vraisemblable que nous sommes alors aussi aptes à percevoir que cette même propriété peut se retrouver dans différents objets distincts.

Notons en terminant que si les tests faits avec Rico ne permettaient pas de déterminer s'il possédait les capacités *C1* et *C2*, les tests faits avec Chaser sont quant à eux extrêmement convaincants à sujet :

En formant des catégories représentées par des noms communs, Chaser associe une même dénomination à plusieurs objets. Chaser a aussi démontré qu'elle pouvait associer jusqu'à trois dénominations à un même objet sans faire d'erreur. Par exemple, l'expérience 1 démontre que Chaser connaissait les noms propres de tous les objets utilisés dans l'étude. Mais Chaser a aussi associé le nom commun «jouet» à certains de ces objets. Le fait qu'elle ait en plus correctement réussi à associer les noms communs «balle» et «frisbee» à certains de ces objets démontre qu'elle leur a associé une troisième dénomination. [...] Chaser a démontré que les mots peuvent représenter des catégories. La démonstration de sa capacité à faire des associations un-à-plusieurs et plusieurs-à-un entre des noms et des objets accroît notre compréhension de la nature référentielle des mots chez les border collie<sup>70</sup>.

Il semble donc que lorsqu'un animal possède *C3* et la capacité de se représenter ce à quoi réfèrent des noms propres — comme dans le cas de Rico —, nous avons possiblement là un indice que l'animal peut aussi développer les capacités *C1*, *C2* et *C4* s'il est correctement entraîné.

## 5. Conclusion

Ce texte avait un double objectif. Nous avons premièrement cherché à démontrer que les quatre critères épistémiques identifiés par Newen et Bartels rendaient correctement compte des capacités épistémiques impliquées par la possession de concepts. Puis une fois cette démonstration faite, nous avons ensuite tenté de démontrer que la théorie empiriste des concepts de Jesse Prinz — une théorie se prononçant sur la nature métaphysique des concepts, plutôt que sur les capacités épistémiques impliquées — permet de donner les assises nécessaires à de telles capacités. À la lumière de la théorie de Prinz, nous avons cherché à mieux comprendre la façon dont les capacités épistémiques peuvent être expliquées chez les animaux, et plus particulièrement, chez Rico le border collie et Alex le perroquet. Bien sûr, il est impossible (ou du moins, extrêmement difficile) de savoir si une analyse du genre rend véritablement compte de ce qui se déroule dans la tête de ces animaux ; c'est un problème auquel toutes les théories du type se heurtent. Mais nous espérons à tout le moins avoir offert une explication plausible.

- 
1. Albert Newen et Andreas Bartels, «Animal Minds and the Possession of Concepts» dans *Philosophical Psychology*, vol.20, no.3 (2007), pp.283-308.
  2. *Ibid.*, p. 284 (citation traduite de l'anglais).
  3. Dans ce texte, «pensée conceptuelle» et «représentation conceptuelle» sont considérées comme des expressions synonymes.
  4. *Ibid.*, p. 299.
  5. Les principaux travaux de Jesse Prinz portant sur sa théorie des concepts sont les suivants : (1)*Furnishing the Mind: Concepts and their Perceptual Basis*, Cambridge, MIT Press, 2002 ; (2)«Sensible Ideas: A Reply to Sarnecki and Markman and Stilwell» dans *Philosophical Psychology*, vol. 17, no.3 (2004), pp.419-430 ; (3)«The Return of Concept Empiricism» dans Henri Cohen et Claire Lefebvre, *Categorization and Cognitive Science*, Amsterdam, Elsevier, 2005 ; (4)«Can Concept Empiricism Forestall Eliminativism?» dans *Mind and Language*, vol. 25, no. 5 (2010), pp. 612-621.
  6. Albert Newen et Andreas Bartels, *op. cit.* p. 296.

7. Juliane Kaminski, Josep Call et Julia Fischer, «Word Learning in a Domestic Dog: Evidence for “Fast Mapping”», dans *Science*, vol. 304, no. 5677 (2004), pp. 1682-1683.
8. Irene Pepperberg, *The Alex studies*, Cambridge, Harvard University Press, 1999.
9. Albert Newen et Andreas Bartels, *op. cit.*, pp.284-285. À la lumière de leur théorie des concepts, les auteurs cherchent aussi à distinguer la représentation conceptuelle des autres formes de représentation comme la représentation non conceptuelle et la représentation propositionnelle. Nous n'examinerons pas ces développements dans ce travail.
10. Albert Newen et Andreas Bartels, *op. cit.* pp. 285-291.
11. Afin de mettre en évidence la différence entre *GS* et *RS*, le schéma en deux étapes de la généralisation de stimuli «*GS*: Perception de stimuli → Réaction» vise à expliciter la relation mécaniste et quasi immédiate entre la perception de stimuli et la réaction. Dans les faits, un schéma du type «*GS*: Perception de stimuli → Traitement non conceptuel de l'information → Réaction» serait probablement plus près du fonctionnement réel du mécanisme de généralisation de stimuli.
12. Fred Dretske, «Machines, Plants and Animals: The Origins of Agency» dans *Erkenntnis*, vol. 51, no. 1 (1999), pp. 19-31.
13. *Ibid.*, pp. 27-28.
14. Albert Newen et Andreas Bartels, *op. cit.*, pp. 286-287.
15. Colin Allen, «Animal Concepts Revisited: the use of Self-Monitoring as an Empirical Approach» dans *Erkenntnis*, vol. 51, no. 1 (1999), pp. 33-40; Colin Allen et Marc Hauser, «Concept Attribution in Non-Human Animals: Theoretical and Methodological Problems in Ascribing Complex Mental Processes» dans *Philosophy of Science*, vol. 58, no. 2 (1991), pp. 221-240.
16. Colin Allen et Marc Hauser, *op. cit.*, p. 231 (citation traduite de l'anglais).
17. Zenon Pylyshyn, «Computation and Cognition: Issues in the Foundations of Cognitive Science» dans Jay Garfield, *Foundations of Cognitive Science: The Essential Readings*, New York, Paragon House, 1990, pp. 18-74.
18. Albert Newen et Andreas Bartels, *op. cit.*, p. 287.
19. *Ibid.*, p. 288 (nous soulignons et la citation est traduite de l'anglais).
20. Juliane Kaminski, Josep Call et Julia Fischer, *op. cit.*
21. Albert Newen et Andreas Bartels, *op. cit.*, p. 288.
22. Donald Davidson, «The emergence of thought», dans *Erkenntnis*, vol. 51, no. 1 (1999), p. 8.

23. Donald Davidson, « Rational Animals », dans *Dialectica*, vol. 36, no. 4 (1982), p. 322 ; Donald Davidson, « The emergence of thought », *op. cit.*, p. 14.
24. Albert Newen et Andreas Bartels, *op. cit.*, p. 289.
25. Colin Allen, *op. cit.*
26. Allen précise toutefois qu'il faut distinguer les critères d'*attribution* de concepts des critères de *possession*, car respecter les critères d'*attribution* « peut fournir de bonnes justifications pour attribuer des concepts aux animaux, sans pour autant que ces critères ne soient nécessaires ou suffisants pour parler de possession de concepts », cf. Colin Allen, *op. cit.*, p. 37 (citation traduite de l'anglais).
27. *Ibid.*, p. 37 (citation traduite de l'anglais).
28. *Ibid.*, p. 38 (citation traduite de l'anglais).
29. *Ibid.*, p. 38.
30. *Ibid.*, p. 39.
31. Albert Newen et Andreas Bartels, *op. cit.*, pp. 290-291.
32. Irene Pepperberg, *op. cit.*
33. Albert Newen et Andreas Bartels, *op. cit.*, section 4.
34. *Ibid.*, p. 295 (citation traduite de l'anglais).
35. *Ibid.*, p. 296.
36. *Ibid.*, pp. 292-293.
37. Les tests ont été conduits de façon à éliminer la possibilité qu'Alex généralise la combinaison de deux stimuli déjà rencontrés — (1) les mêmes objets perçus visuellement (2) associés à des questions déjà entendues — à partir desquels une réponse rigide serait développée. Une des observations qui nous encourage à juger que ce n'est pas le cas est qu'Alex peut correctement répondre à de nouvelles combinaisons, cf. Irene Pepperberg, *op. cit.*, p. 68.
38. Albert Newen et Andreas Bartels, *op. cit.*, p. 296.
39. *Ibid.*, pp. 293-294.
40. *Ibid.*, p. 296.
41. Il ne faut pas confondre une théorie empiriste des concepts — qui porte sur la façon dont nos pensées conceptuelles sont véhiculées lorsque nous réfléchissons — avec d'autres types de théories empiristes, comme l'empirisme épistémologique, sémantique ou le béhaviorisme, cf. Jesse Prinz, *Furnishing the Mind: Concepts and their Perceptual Basis*, *op. cit.*, p. 109. Voir la note 5 du présent texte pour un résumé des principaux ouvrages de Jesse Prinz portant sur sa théorie des concepts.
42. Jesse Prinz, *Furnishing the Mind: Concepts and their Perceptual Basis*, *op. cit.*, p. ix.

43. À titre d'exemple, dans les travaux de Prinz mentionnés dans ce texte, on trouve de nombreuses références aux travaux suivants de Lawrence Barsalou: (1) «The Instability of Graded Structure: Implications for the Nature of Concepts», dans Ulric Neisser, *Concepts and Conceptual Development: Ecological and Intellectual Factors in Categorization*, Cambridge, Cambridge University Press, 1987; (2) «Flexibility, Structure, and Linguistic Vagary in Concepts: Manifestations of a Compositional System of Perceptual Symbols», dans Alan Collins, *Theories of Memory*, Hillsdale, Lawrence Erlbaum Associates, 1993; (3) «Perceptual Symbol Systems», dans *Behavioral and Brain Sciences*, vol. 22, no. 4 (1999), pp. 577-660. Si ces travaux peuvent paraître quelque peu datés, la page Web de Lawrence Barsalou (<http://psychology.emory.edu/cognition/barsalou/onlinepapers.html>) fait état de plusieurs travaux récents sur les concepts, et plus précisément, sur la relation entre concepts et modalités sensorielles. Cf. (1) Lawrence Barsalou, «Cognitive and Neural Contributions to Understanding the Conceptual System», dans *Current Directions in Psychological Science*, vol. 17, no. 2 (2008), pp. 91-95; (2) Lawrence Barsalou, «Simulation, Situated Conceptualization, and Prediction», dans *Philosophical Transactions of the Royal Society of London: Biological Sciences*, vol. 364 (2009), pp. 1281-1289; (3) Markus Kiefer et Lawrence Barsalou, «Grounding the Human Conceptual System in Perception, Action, and Internal States», dans Wolfgang Prinz, Miriam Beisert et Arvid Herwig, *Action Science: Foundations of an Emerging Discipline*, Cambridge, MIT Press, 2013, pp. 381-407.
44. Jesse Prinz, *Furnishing the Mind: Concepts and their Perceptual Basis*, *op. cit.*, p. 108 (citation traduite de l'anglais). On reproche souvent à un tel empirisme de ne pas pouvoir rendre compte des concepts abstraits, car ces derniers ne peuvent pas vraiment faire l'objet d'expériences perceptuelles. Newen et Bartels font d'ailleurs ce reproche à la théorie de Prinz, cf. Albert Newen et Andreas Bartels, *op. cit.*, p. 302. Il est à noter que Jesse Prinz et Lawrence Barsalou offrent des réponses à ces objections, mais nous n'aurons pas l'occasion de les examiner dans le cadre de ce texte. Cf. (1) Jesse Prinz, *Furnishing the Mind: Concepts and their Perceptual Basis*, *op. cit.*; (2) Jesse Prinz, «Sensible Ideas: A Reply to Sarnecki and Markman and Stilwell», *op. cit.*; (3) Jesse Prinz, «The Return of Concept Empiricism», *op. cit.*, section 3; (4) Jesse Prinz, «Beyond Appearances: The Content of Sensation and Perception», dans Tamar Gendler et John Hawthorne, *Perceptual Experience*, Oxford, Oxford University Press, 2006, pp. 434-460; (5) Lawrence Barsalou,

- «Perceptual Symbol Systems», *op. cit.*, section 3.4; (6) Lawrence Barsalou et Katja Wiemer-Hastings, «Situating Abstract Concepts», dans Diane Pecher et Rolf Zwaan, *Grounding Cognition: The role of Perception and Action in Memory, Language, and Thought*, New York, Cambridge University Press, 2005, pp.129-163; (7) Christine Wilson-Mendenhall, Kyle Simmons, Alex Martin et Lawrence Barsalou, «Contextual Processing of Abstract Concepts Reveals Neural Representations of Non-Linguistic Semantic Content», dans *Journal of Cognitive Neuroscience*, vol. 25, no. 6 (2013), pp. 920-935.
45. Jesse Prinz, «The Return of Concept Empiricism», *op. cit.*, pp. 681-683; Jesse Prinz et Andy Clark, «Putting Concepts to Work: Some Thoughts for the Twentyfirst Century», *Mind & Language*, vol. 19, no. 1 (2004), pp. 57-69.
46. La mémoire de travail est un des différents types de mémoire (par opposition à la mémoire à long terme, par exemple). Elle n'est pas dédiée à un type d'information spécifique et «elle permet à un individu d'activer et de maintenir (parfois suite à un entraînement actif) un ensemble de représentations mentales afin de manipuler et traiter ces représentations. Le contenu de la mémoire de travail est généralement considéré comme étant conscient», cf. Peter Carruther, «Evolution of Working Memory» dans *Proceedings of the National Academy of Sciences*, vol. 110, supplément 2 (2013), p. 10371.
47. Jesse Prinz, «Can Concept Empiricism Forestall Eliminativism?», *op. cit.*, p. 618 (citation traduite de l'anglais).
48. *Ibid.*, pp. 618-619.
49. Jesse Prinz, «The Return of Concept Empiricism», *op. cit.*, pp. 687-688.
50. *Ibid.*, p. 685.
51. *Ibid.*, pp. 685-686.
52. Jesse Prinz, «Can Concept Empiricism Forestall Eliminativism?», *op. cit.*, p. 618.
53. *Ibid.*, p. 617.
54. Jesse Prinz, «The Return of Concept Empiricism», *op. cit.*, p. 691 (passage traduit de l'anglais).
55. Par causation étologique, il est entendu que la création d'une catégorie est causée par la perception d'une première instance associée à cette catégorie, puis qu'un concept représente toujours par la suite une instance du type de représentations associées à cette catégorie. Par causation nomologique, il est entendu qu'il y a un lien causal fiable entre les perceptions d'instances d'une catégorie et les représentations

- conceptuelles de cette catégorie, cf. Jesse Prinz, «The Return of Concept Empiricism», *op. cit.*, p. 685.
56. Jesse Prinz, *Furnishing the Mind: Concepts and their Perceptual Basis*, *op. cit.*, pp.21-22. Mais Prinz admet toutefois que le langage peut améliorer nos habiletés conceptuelles. Cf. *Furnishing the Mind: Concepts and their Perceptual Basis*, *op. cit.*, p.21 ; «The Return of Concept Empiricism», p.692.
57. Juliane Kaminski, Josep Call et Julia Fischer, *op. cit.*
58. Peter Carruther, *op. cit.*
59. Lawrence Barsalou, «Perceptual Symbol Systems», *op. cit.*, pp. 606-607 ; Lawrence Barsalou et Katja Wiemer-Hastings, «Situating Abstract Concepts», *op. cit.*
60. Jesse Prinz, «The Return of Concept Empiricism», *op. cit.*, p.688 (citation traduite de l'anglais).
61. Albert Newen et Andreas Bartels, *op. cit.*, pp. 297-298.
62. *Ibid.*, p. 297.
63. Juliane Kaminski, Josep Call et Julia Fischer, *op. cit.*, p. 1682.
64. Les tests conduits avec Rico avaient entre autres pour but de démontrer qu'il était apte à faire de telles inférences. Ces tests se sont montrés concluants, cf. Juliane Kaminski, Josep Call et Julia Fischer, *op. cit.*
65. John Pilley et Alliston Reid, «Border Collie Comprehends Object Names as Verbal Referents», *Behavioural Processes*, vol.86, no.2 (2010), pp. 184-195.
66. Albert Newen et Andreas Bartels, *op. cit.*, p. 306, note 6 (citation traduite de l'anglais).
67. John Pilley et Alliston Reid, *op. cit.*, p. 188.
68. *Ibid.*, pp.190-192. Les tests faits avec Chaser ont aussi permis de démontrer qu'il était capable de distinguer les mots désignant les objets des commandements associés à ces mots : prendre, toucher avec la patte, toucher avec le museau (*Ibid.*, pp. 189-190).
69. Cela dit, les tests avec Alex sont particuliers, car bien qu'ils font état de ce qui semble être des exemples manifestes de possession de concepts chez un animal, ils ne sont probablement pas représentatifs de la façon dont les concepts sont habituellement «utilisés» chez les animaux. En effet, outre les cas où certains animaux associent un cri d'alarme à la présence d'un «objet» spécifique — comme dans le cas des singes vervet, cf. Robert Seyfarth, Dorothy Cheney et Peter Marler, «Vervet Monkeys Alarm Calls: Semantic Communication in a Free-Ranging Primate», dans *Animal Behaviour*, vol.28, no.4 (1980), pp.1070-1094 — , on peut supposer que les animaux n'identifient généralement pas les

catégories que visent à représenter leurs concepts par des sons servant spécifiquement à identifier la catégorie — comme le son « couleur » identifie la catégorie des couleurs. L'une des principales utilités de la représentation conceptuelle pour un animal doit plutôt venir du fait que suite à la perception d'une seule ou de quelques propriétés associées à une catégorie, l'animal peut réagir avant même d'avoir perçu la totalité des propriétés de la catégorie (ce qui est particulièrement avantageux pour se sauver de prédateurs). Les actions de l'animal peuvent donc être plus rapides et efficaces.

70. John Pilley et Alliston Reid, *op. cit.*, p. 192 (citation traduite de l'anglais).





# L'ascension intentionnelle nécessite-t-elle une ascension sémantique : une critique de Robert W. Lurz

JULIEN OUELLET, *Université Laval*

## *Introduction*

Les animaux pensent-ils? Cette question a fait et continue de faire couler beaucoup d'encre. Certains philosophes tendent à refuser toutes pensées aux animaux<sup>1</sup>, alors que d'autres considèrent que l'existence de la pensée animale va de soi<sup>2</sup>. Heureusement, il existe une myriade de positions nuancées entre ces deux pôles. José Luis Bermudez, par exemple, permet aux animaux d'avoir des pensées à propos du monde. Par contre, il refuse que ces derniers aient des pensées sur des pensées. En effet, Bermudez, dans *Thinking without Words*<sup>3</sup>, présente un argument selon lequel la pensée du second ordre (*second-order thoughts*), soit le fait d'avoir des pensées sur des pensées, nécessite de penser sur des mots. Si l'argument est fondé, les animaux dénués de langage se retrouvent dans l'impossibilité d'avoir des pensées du second ordre. Dans cet article, je tenterai donc de savoir si cet argument est fondé. En premier lieu, je remettrai en contexte l'argument de Bermudez dans le cadre plus général de sa théorie de la pensée et du raisonnement non linguistiques. En deuxième lieu, je présenterai la reconstruction que fait Robert W. Lurz de cet argument<sup>4</sup>. Bien que Lurz ne respecte pas parfaitement la forme de l'argument présenté par Bermudez, la reconstruction qu'il propose semble être essentiellement conforme à sa pensée. En troisième lieu, j'examinerai les critiques qu'avance Lurz contre l'argument de Bermudez. Finalement, je mettrai de l'avant les désaccords entre Bermudez et Lurz, afin de mettre en évidence les enjeux théoriques et empiriques inhérents au débat et afin de déterminer si la critique de Lurz atteint bien sa cible.

### Mise en contexte de l'argument

Dans *Thinking without Words*, Bermudez développe un cadre théorique permettant de rendre compte de la pensée des animaux non linguistiques. Considérant que, selon Bermudez<sup>5</sup>, les notions de *pensée* et de *raisonnement* sont inextricablement liées, il doit expliquer comment des animaux peuvent être capables de raisonner, et ce même s'ils ne possèdent pas de langage<sup>6</sup>. En effet, la notion classique de *raisonnement* se base généralement sur l'application de certaines règles logiques, qui elles, s'appliquent uniquement à des énoncés. Or, les énoncés sont des entités linguistiques. Ainsi, force est de constater que les animaux non linguistiques, par définition, sont incapables de faire des raisonnements en bonne et due forme. L'explication de Bermudez se base essentiellement sur la protologique, une forme primitive de raisonnement qu'auraient les animaux non linguistiques et qui leur permettraient d'effectuer certains types de raisonnements analogues aux *modus ponens* et au *modus tollens*. Par exemple, un animal pourrait s'engager dans le raisonnement suivant<sup>7</sup> :

- (1) Si la gazelle est présente au point d'eau, alors le lion est absent au point d'eau.
- (2) Le lion est présent au point d'eau.
- (3) Donc, la gazelle est absente au point d'eau.

Dans cet exemple de proto-*modus tollens*, la prémisse (1) est justifiée par le proto-conditionnel, une capacité qu'auraient les animaux d'associer certains événements en observant des régularités empiriques<sup>8</sup>. Le protoconditionnel, en associant des événements (contrairement au conditionnel logique, qui associe des énoncés), serait accessible aux animaux non linguistiques. La prémisse (2), quant à elle, est le simple résultat d'une perception<sup>9</sup>. La conclusion (3) s'appuie à la fois sur la notion de *protonégation* et sur une notion non syntaxique de *validité*. La protonégation correspond à la capacité de lier ensemble une paire de concepts contraires<sup>10</sup> (présent/absent, visible/invisible). En faisant ainsi, l'animal en question serait capable de reconnaître qu'un objet ne peut pas à la fois être présent et être absent au point d'eau. De plus, le raisonnement qu'attribue Bermudez aux animaux non linguistiques n'est pas

syntactiquement valide, c'est-à-dire qu'il n'est pas valide en vertu de sa forme<sup>11</sup>. Il est toutefois valide dans la mesure où les prémisses ne peuvent être vraies et la conclusion fautive<sup>12</sup>.

Par contre, Bermudez défend que la protologique n'est pas suffisante pour permettre aux animaux non linguistiques d'avoir des pensées du second ordre (des pensées sur des pensées). En effet, Bermudez nous dit que certains types de pensées sont, par principe, uniquement accessibles aux créatures disposant d'un langage<sup>13</sup>. Bermudez justifie cette affirmation à l'aide d'un argument selon lequel l'ascension intentionnelle nécessite une ascension sémantique<sup>14</sup>. Que veut-il dire par là ? Par « ascension intentionnelle », Bermudez fait référence à la capacité d'avoir des pensées sur des pensées, et par « ascension sémantique », Bermudez fait référence à la capacité d'avoir des pensées sur des mots. Ainsi, Bermudez cherche à défendre que les animaux non linguistiques sont incapables d'avoir des pensées sur des pensées<sup>15</sup>, car cela nécessiterait qu'ils soient aussi capables d'avoir des pensées sur des mots (et donc, de posséder un langage). Considérant que, par définition, les animaux non linguistiques ne possèdent pas de langage, on peut en conclure que les animaux non linguistiques ne peuvent pas avoir de pensées du second ordre. À des fins pratiques, je nommerai l'argument de Bermudez « argument de la nécessité sémantique ».

Ainsi, même si les animaux non linguistiques sont capables d'avoir certains types de pensées à propos du monde et d'effectuer certains types de raisonnements, l'argument de la nécessité sémantique limite considérablement les pensées que l'on peut leur attribuer. Évidemment, la conclusion à laquelle arrive Bermudez n'est pas sans conséquence. En effet, certaines capacités, comme la métacognition (avoir une pensée sur ses propres pensées<sup>16</sup>) et l'attribution d'états mentaux (*mind-reading*) sont considérées comme des pensées du second ordre. Ainsi, il semble que Bermudez refuse, à l'aide d'un argument *a priori*<sup>17</sup>, plusieurs capacités cognitives aux animaux non linguistiques, et ce, en dépit de certains résultats provenant des sciences cognitives.

*Argument de la nécessité sémantique*

Dans *In Defense of Wordless Thoughts About Thoughts*<sup>18</sup>, Lurz reconstruit et critique l'argument de Bermudez<sup>19</sup>. Je commencerai par exposer la structure logique de l'argument, pour ensuite me pencher davantage sur chacune des prémisses. Bermudez présente essentiellement deux thèses. Premièrement, (T1) les animaux non linguistiques peuvent avoir des pensées à propos du monde, comme nous l'avons vu plus tôt. Deuxièmement, (T2) les animaux non linguistiques ne peuvent avoir de pensées du second ordre. La thèse (T2) est supportée essentiellement par deux prémisses<sup>20</sup>. La première défend que (P2) les pensées du second ordre impliquent des pensées à propos de certains énoncés du langage public<sup>21 22</sup>. La deuxième prémisses énonce une vérité logique selon laquelle (P3) les animaux non linguistiques ne peuvent pas avoir de pensées à propos des énoncés du langage public (par définition). Ainsi, toute la force de la thèse (T2) repose sur la prémisses (P2).

La prémisses (P2), qui représente la conclusion de l'argument de la nécessité sémantique, est supportée par deux autres prémisses. La première défend que (L2<sup>23</sup>) les pensées du second ordre impliquent d'entretenir (*entertaining*) des pensées de manière consciente et de considérer leurs relations logiques et évidentielles<sup>24</sup> (*evidential*). La deuxième défend que (L3) si l'on entretient une pensée de manière consciente et que l'on considère ses relations logiques et évidentielles, alors le véhicule de la pensée doit être personnel et linguistique<sup>25</sup>. La prémisses (L3) mérite d'être clarifiée davantage. Il est important de comprendre que l'argument de Bermudez se concentre sur les véhicules de la pensée : « The basic premise from which we need to start here is that reflexive thoughts can only be possible if the target thoughts have vehicles that allow them to be the objects of further thoughts<sup>26</sup>. » Les exemples classiques de véhicule sont le langage de la pensée, le langage naturel et la représentation picturale (des cartes et des modèles, par exemple). Bien que Bermudez ne définit pas exactement ce qu'il entend par « véhicule », il semble approprié de comprendre la distinction entre pensée et véhicule à l'aide de la distinction entre prédicat et propriété. Présenté ainsi, un prédicat est

une entité linguistique représentant une propriété. De la même façon, un véhicule serait l'entité par lequel une pensée est représentée.

Nous voyons maintenant que la thèse (T2), selon laquelle les animaux non linguistiques ne peuvent pas avoir de pensées du second ordre, repose entièrement sur la plausibilité de (L2) et de (L3). Examinons comment Bermudez défend ces deux prémisses. Afin de défendre (L2), Bermudez reprend essentiellement l'analyse de Andy Clark, selon laquelle la pensée du second ordre<sup>27</sup> nous permet de nous engager dans des *dynamiques cognitives du second ordre*<sup>28</sup>. De quoi est-il question ? Clark écrit : «By second-order cognitive dynamics I mean a cluster of powerful capacities involving self-evaluation, self-criticism and finely honed remedial responses<sup>29</sup>.» Ainsi, la pensée du second ordre doit permettre une série de capacités passablement complexes, sans quoi il n'est pas question de pensée du second ordre. Avant d'explorer la défense de (L3), il est intéressant de mentionner que, selon Lurz, (T2) est défendue sur le terrain de la métaphysique : «given that Bermudez's argument for T2 is largely *a priori* and does not rely upon any contingent facts about nonlinguistic creatures or laws of nature, it is plausible to take the modalities in question to be metaphysical possibility and necessity<sup>30</sup>». Le caractère métaphysique que Lurz attribue à (T2) est passablement douteux. Premièrement, ni Bermudez ni Clark ne prétendent mettre de l'avant une thèse métaphysique. Deuxièmement, Lurz lui-même attaquera (L2) en faisant appel à des arguments empiriques, ce qui nous permet de douter du caractère «largement *a priori*» de (T2).

La prémisses (L3), quant à elle, est plus longuement défendue par Bermudez. En d'autres mots, cette prémisses énonce qu'il n'existe qu'un seul type de véhicule qui puisse permettre à une pensée du premier ordre d'être la cible d'une pensée du second ordre, soit le langage naturel. Pour ce faire, Bermudez considère qu'un véhicule doit remplir deux conditions. Premièrement, le véhicule doit être au niveau personnel. Qu'est-ce que cela veut dire ? Bermudez oppose les véhicules personnels aux véhicules sous-personnels (*subpersonnal*). Les véhicules sous-personnels correspondent explicitement au langage de la pensée de Fodor. Or, comme le dit Bermudez, il semble difficile de voir comment il serait possible d'avoir un accès conscient

au langage de la pensée, considérant que le langage de la pensée est généralement conçu comme une structure cognitive inconsciente<sup>31</sup>. Par élimination, il semble donc qu'une pensée doit avoir un véhicule personnel pour être la cible d'une pensée du second ordre<sup>32</sup>. J'aurai l'occasion de revenir plus en détail sur cette partie de l'argument lorsque j'exposerai la critique de Lurz.

Deuxièmement, Bermudez défend que le véhicule doit être linguistique. Pour en arriver à cela, Bermudez fonctionne encore une fois par élimination. Selon lui, il existe deux types de véhicule au niveau personnel : les symboles complexes du langage naturel et les modèles picturaux<sup>33</sup>. Parmi les modèles picturaux, les plus importants sont les cartes et les modèles mentaux. Ainsi, Bermudez doit nous convaincre que ni les cartes mentales ni les modèles mentaux ne sont des véhicules convenables. Pour ce faire, Bermudez fera appel à la notion d'*isomorphisme structurel*, qui renvoie grosso modo à la notion de *ressemblance*. Bermudez distingue ensuite deux types d'isomorphisme : l'isomorphisme faible et l'isomorphisme fort. L'isomorphisme faible sera attribué à une pensée s'il existe une ressemblance entre le contenu d'une pensée et de ce qu'elle représente. Par contre, lorsqu'il est question d'isomorphisme fort, il faut non seulement qu'il y ait une ressemblance entre le contenu d'une pensée et de ce qu'elle représente, mais il faut aussi qu'il existe des unités représentationnelles de base qui peuvent être combinées selon des règles combinatoires indépendantes entre elles<sup>34</sup>. Selon Bermudez, le langage naturel possède un isomorphisme structurel fort, alors que les cartes et les modèles mentaux ne possèdent qu'un isomorphisme structurel faible. C'est précisément à cause du manque d'isomorphisme structurel fort que les cartes et les modèles mentaux ne peuvent être la cible d'une pensée du second ordre : « they do not seem to be structured in the right sort of way to permit the reflexive type of second-order cognitive dynamics under discussion<sup>35</sup> ». Un exemple aidera à comprendre. Les dynamiques cognitives du second ordre, selon Clark, impliquent la capacité de voir les bases évidentielles de nos croyances. Par exemple, les dynamiques cognitives du second ordre nous permettent de réaliser que nous avons fait une induction hâtive. Or, selon Bermudez, les cartes mentales

permettent d'effectuer une telle évaluation uniquement si elles sont interprétées en termes propositionnels : « Those very feature of maps (their analogue nature and structural isomorphism with what they represent) that make them so useful for guiding action serve to make them inappropriate for the type of inferential evaluation characteristic of second-order cognitive dynamics<sup>36</sup> ». Ainsi, le seul candidat en lice pour le titre de *véhicule permettant la pensée du second ordre* est le langage naturel.

### *Critique de l'argument de la nécessité sémantique*

Lurz présente des arguments autant contre (L2) que contre (L3). Commençons par examiner la critique de (L2). Cette prémisse énonce que les pensées du second ordre impliquent d'entretenir des pensées de manière consciente et de considérer leurs relations logiques et évidentielles. Lurz distingue deux manières de comprendre cette thèse : une interprétation forte et une interprétation faible. Lurz défend que si l'on prend une interprétation forte de (L2), alors la thèse est empiriquement fausse. D'un autre côté, si l'on prend une interprétation faible de (L2), la thèse est inconsistante avec le projet de Bermudez.

Afin de montrer qu'une interprétation forte de (L2) est empiriquement fausse, Lurz présente plusieurs études en psychologie développementale qui tendent à montrer que (E1) les enfants montrent des signes explicites de pensées du second ordre dès l'âge de 5 ans<sup>37</sup>. Par exemple, dans une étude sur l'explication des fausses croyances<sup>38</sup>, on dit à des enfants (3-4 ans) qu'un protagoniste cherche un objet et que cet objet est caché sous le piano. Après avoir demandé à ces enfants pourquoi le protagoniste regarde sous le sofa, les enfants expliquent ce comportement en disant que le protagoniste regarde sous le sofa, car il croit que l'objet y est caché. Il semble ainsi que les enfants attribuent une croyance au protagoniste, ce qui démontre qu'ils ont des pensées du second ordre (une croyance à propos d'une croyance). Malgré cela, d'autres études en psychologie tendent à démontrer que (E2) les enfants ne s'engagent dans aucun processus de mise en relation consciente des relations logiques et évidentielles avant le début de l'adolescence<sup>39</sup>. Sur la base de ces expériences en

psychologie du développement, il semble ainsi qu'une interprétation forte de (L2) soit empiriquement fausse.

Examinons maintenant l'interprétation faible de (L2). Lurz nous dit qu'il est possible que (L2) fasse référence uniquement aux pensées *explicités* (conscientes<sup>40</sup>) du second ordre<sup>41</sup>. De cette manière, il semble plausible que les pensées explicites du second ordre requièrent effectivement d'entretenir des pensées de manière consciente et de considérer leurs relations logiques et évidentielles. Par contre, si cette interprétation est vraie, Bermudez ouvrirait la porte à la possibilité que les animaux non linguistiques puissent avoir des pensées *implicites* du second ordre, ce qui impliquerait que les animaux non linguistiques soient capables de s'engager dans des raisonnements (au sens classique du terme) implicites. Certes, Bermudez considère que les animaux non linguistiques peuvent avoir des pensées sur des états du monde. Par contre, il semble exagéré, pour Bermudez, d'admettre que les animaux non linguistiques soient capables de raisonner (au sens classique), même de manière implicite, sur des pensées. En effet, Bermudez prend le temps de développer sa protologie afin de permettre aux animaux d'être capables de s'engager dans un type *primitif* de raisonnement. De plus, Bermudez critique, à plusieurs endroits dans son livre, la conception de la rationalité pratique basée sur l'inférence<sup>42</sup>. Ainsi, il semblerait étrange qu'il admette du même coup que les animaux soient capables de s'engager inconsciemment dans des raisonnements classiques. Admettant cette interprétation faible de (L2), Lurz soulève une possible solution : si Bermudez veut montrer que les animaux ne peuvent pas, non plus, avoir de pensées implicites, il doit défendre une dépendance entre la pensée explicite et la pensée implicite<sup>43</sup> selon laquelle la pensée implicite *nécessite* la pensée explicite. Lurz développe un argument relativement complexe afin de montrer qu'une telle dépendance est inconsistante avec le projet de Bermudez<sup>44</sup>, mais il n'en sera pas question dans cet article<sup>45</sup>.

Examinons la critique de (L3). Lurz soulève la plausibilité qu'une pensée puisse être la cible d'une pensée du second ordre même si la pensée est au niveau sous-personnel, ce qui viendrait attaquer (L3), qui défend qu'une pensée doit nécessairement avoir un véhicule

personnel afin d'être la cible d'une pensée du second ordre. Lurz présente un exemple dans lequel une personne essaie de déterminer si l'inférence suivante est valide : (1) Andrew est plus vieux que Barb, (2) James est plus jeune que Barb, (3) donc Andrew est le plus vieux<sup>46</sup>. Selon Lurz, il est plausible que la personne en question entretienne des pensées (une par prémisses et une pour la conclusion) dans le langage de la pensée (au niveau sous-personnel), tout en utilisant des images mentales afin de tenir ces pensées assez longtemps pour pouvoir évaluer leurs relations logiques et évidentielles de manière consciente. Évidemment Bermudez n'accepterait pas ce genre d'exemples. Lurz soulève deux considérations afin de défendre la plausibilité de son exemple.

Premièrement, Lurz accuse Bermudez de commettre une confusion entre *pensée* et *véhicule de la pensée* : « Bermudez's reasoning here appears to rest upon the dubious assumption that if the vehicles of thought are subpersonal, the thoughts themselves those vehicles represent are as well<sup>47</sup>. » Lurz, considère que les pensées peuvent provenir du langage de la pensée et être représentées autrement à l'esprit de l'agent. Ainsi, rien n'empêche qu'une pensée soit, en soi, sous-personnelle, mais qu'une personne se la représente à travers un véhicule personnel. Lurz écrit : « The idea here is akin to that of the use of diagrams and shapes in the evaluation of geometrical proofs : the proofs are, of course, sets of propositions, while the diagrams and shapes are helpful aids in the evaluation and construction of the proofs<sup>48</sup>. » On comprend ici que la représentation picturale n'est pas constitutive de la pensée, mais elle vient supporter cette dernière afin de permettre à un organisme d'entretenir cette pensée à l'esprit.

Deuxièmement, il apparaît clair que Lurz croit que les véhicules picturaux ont la structure nécessaire afin de permettre d'entretenir une pensée au sens de la dynamique cognitive du second ordre (contrairement à ce que croit Bermudez). En effet, dans l'exemple de Lurz, c'est grâce à la représentation picturale qu'il devient possible d'évaluer consciemment les relations logiques et évidentielles des pensées. Lurz fonde la plausibilité de cette affirmation sur la théorie des modèles mentaux de Johnson-Laird<sup>49</sup>. Il semble s'agir d'un enjeu

empirique, et vraisemblablement, Lurz et Bermudez ne s'entendent pas. Nous aurons l'occasion d'y revenir sous peu.

### *Désaccords entre Bermudez et Lurz*

Dans cette section, je considérerai deux désaccords importants entre Bermudez et Lurz. Premièrement, j'examinerai si l'interprétation forte de (L2) est, effectivement, empiriquement fautive. Lurz base sa critique sur deux types d'expériences, l'un (E1) qui démontre que les enfants de 4 ou 5 ans sont capables d'avoir des pensées du second ordre, et un autre (E2) qui démontre que les enfants sont malgré tout incapables d'entretenir consciemment des pensées et de considérer leurs relations logiques et évidentielles. Afin de contrer cette objection, Bermudez peut, entre autres, fournir une interprétation des expériences (E1) qui n'attribue pas de pensées du second ordre aux enfants. Lurz rejette les interprétations de ce type, car il considère que la majorité des expériences (E1) ne peuvent être réinterprétées de la sorte (du moins, de manière plausible<sup>50</sup>).

Comme Lurz, j'ai de la difficulté à expliquer autrement le comportement de ces enfants sans faire appel à des pensées du second ordre (quoique cela ne semble pas impossible). Par contre, faut-il rappeler que la totalité des expériences (E1) que soulève Lurz porte sur le comportement d'enfants *humains*<sup>51</sup>. Bien qu'il ne s'agisse pas d'une objection très puissante, il faut se rappeler que l'argument de Bermudez vise essentiellement les animaux non humains. Il semble que des expériences empiriques sur des animaux non linguistiques auraient été plus appropriées. Qui plus est, les expériences (E1) qu'utilise Lurz portent sur des organismes en *développement*. Bermudez pourrait, il me semble, se détacher de ces expériences en défendant que la capacité d'avoir des pensées du second ordre, dans le cas des enfants de 4-5 ans, n'est que partielle, mais qu'une fois pleinement développée, elle impliquera effectivement une mise en relation consciente des relations logiques et évidentielles. De plus, les expériences (E2) tendent à montrer que les enfants, même en très bas âge, considèrent effectivement les relations logiques et évidentielles de leurs pensées, mais ils ne semblent pas le faire de manière *consciente*<sup>52</sup>. C'est sur ce point que Lurz fonde sa critique.

Or, il semble raisonnable d'affirmer que le phénomène de conscience est, encore aujourd'hui, bien mal compris, autant par les philosophes que les scientifiques, de quoi nous permettrait de remettre en question la force de la critique qu'adresse Lurz envers (L2)<sup>53</sup>.

Deuxièmement, Lurz et Bermudez ne s'entendent pas à savoir si les modèles picturaux possèdent bel et bien une structure adéquate afin de permettre une mise en relation logique et évidentielle entre les pensées. Lurz répond positivement en se fondant sur la théorie de Johnson-Laird<sup>54</sup>, alors que Bermudez répond négativement en se fondant sur la théorie de Braddon-Mitchell et Jackson<sup>55</sup>. Comme le dit Lurz lui-même, ce désaccord nous ramène à un débat en psychologie cognitive entre les pictoralistes et les descriptionnistes. Ce débat dépasse largement l'objectif de cet article, mais il semble adéquat d'affirmer qu'il n'existe pas encore de consensus par rapport à la question<sup>56</sup>. Le contre-exemple que présente Lurz, même s'il est plausible, ne semble pas être concluant à l'égard de (L3). L'avenir nous dira qui de Lurz ou de Bermudez a raison. Sur la base des deux éléments mentionnés, il me semble que la critique de Lurz ne soit pas aussi puissante qu'il ne le croit. Ainsi, considérant que la critique de (L2) et de (L3) soulève certains problèmes non résolus, il semble que l'argument de Bermudez ne s'en tire pas trop mal.

### *Conclusion*

Nous avons vu que l'argument de la nécessité logique prétend poser une limite aux types de pensées que peuvent entretenir les animaux non linguistiques. Cet argument (P2), selon lequel les pensées du second ordre impliquent des pensées à propos d'entité linguistique, se fonde sur deux prémisses. Premièrement, (L2) les pensées du second ordre impliquent d'entretenir des pensées de manière consciente et de considérer leurs relations logiques et évidentielles. Lurz critique cette prémisse en défendant que, si l'on conserve une interprétation forte de cette prémisse, elle est empiriquement fautive. En jetant un doute sur les expériences empiriques que soulève Lurz, j'ai tenté de montrer que cette affirmation n'est pas aussi évidente qu'il n'y paraît. Deuxièmement, l'argument de la nécessité sémantique se fonde sur la prémisse (L3), selon laquelle si l'on entretient une pensée de manière

consciente et que l'on considère ses relations logiques et évidentielles, alors le véhicule de la pensée doit être personnel et linguistique. Lurz critiquait cette prémisse en défendant, entre autres, que les représentations picturales permettent bel et bien de s'engager dans la mise en relation consciente des relations logiques et évidentielles de nos pensées. En mettant en lumière que la réponse à cette question (non résolue) est essentiellement empirique, je souhaitais encore une fois lever un doute quant à la force de la critique de Lurz.

- 
1. D. Davidson, «Rational Animals» *Dialectica*, vol.36, n°4, 1982, pp.317-327.
  2. J.R. Searle, «Animal Minds» *Midwest Studies in Philosophy*, vol. 19, n° 1, 1994, pp.206-219.
  3. J.L. Bermudez, *Thinking about Words*, Oxford, Oxford University Press, 2003.
  4. R. W. Lurz, «In Defense of Wordless Thoughts About Thoughts» *Mind & Language*, vol. 22, n° 3, 2007, pp.270-296.
  5. J.L. Bermudez, *Thinking about Words*, Oxford, Oxford University Press, 2003, p. 140.
  6. Le langage, au sens fort, demande l'existence d'un système de signes, doté d'une sémantique et d'une syntaxe. Évidemment, les animaux sont capables de communiquer, mais la notion de langage est plus exigeante que cela.
  7. *Ibid.*, p. 144.
  8. *Ibid.*, p. 145.
  9. Bermudez emprunte la théorie de l'affordance de J.J. Gibson selon laquelle la perception n'est pas neutre. La nature offre, dans le contenu même de la perception, certaines alternatives d'action, et ce, sans avoir à s'engager dans un quelconque raisonnement. Voir J.J. Gibson, «The Theory of Affordances» dans *Perceiving, Acting, and Knowing: Toward an Ecological Psychology*, pp. 67-82.
  10. *Ibid.*, p. 143.
  11. *Ibid.*, p. 148.
  12. Il existe des notions plus faibles de validité, comme la validité sémantique, mais Bermudez n'identifie pas clairement à quel type de validité il fait référence.
  13. *Ibid.*, p. 140.
  14. *Ibid.*, p. 158.

15. Par souci de clarté, j'utiliserai « pensées du second ordre » afin de faire référence à des pensées sur des pensées et à l'ascension intentionnelle.
16. La notion de métacognition est plus complexe que cela, mais considérant qu'elle ne sera pas discutée davantage dans cet article, je me permets de la définir sommairement.
17. J'aurai l'occasion de revenir sur le caractère *a priori* de l'argument dans ce même article.
18. R. W. Lurz, « In Defense of Wordless Thoughts About Thoughts » dans *Mind & Language*, vol. 22, n° 3, 2007, pp. 270-296.
19. La reconstruction faite par Lurz est passablement différente de l'argument original présenté par Bermudez, mais je considère qu'elle est fidèle à la pensée de Bermudez.
20. En vérité, la reconstruction faite par Lurz expose une prémisse supplémentaire, selon laquelle (P1) l'attribution d'états mentaux implique des pensées du second ordre. J'ai décidé d'ignorer cette prémisse pour deux raisons. Premièrement, le présent article ne traitera pas spécifiquement de l'attribution d'attitudes propositionnelles (*propositional attitude ascriptions*). Deuxièmement, Lurz considère lui-même que cette prémisse est plausible. Voir J. Fodor, « Propositional attitudes » dans *The Monist*, vol. 61, 1978, pp. 501-523 et N. Salmon et S. Soames, *Propositional Attitudes*, Oxford, Oxford University Press, 1988.
21. Contrairement au langage de la pensée, par exemple.
22. Lurz considère qu'il s'agit d'un énoncé métaphysique, alors il ajoute un opérateur modal de nécessité dans la formalisation de chacune des prémisses. L'aspect métaphysique des prémisses ne nous intéressera pas dans le présent article, alors je me suis permis de l'ignorer.
23. (L1) était la prémisse qui supportait (P1), prémisse que j'ai décidé de laisser de côté.
24. Bien que le mot n'existe pas en français, je n'ai trouvé aucune traduction satisfaisante. J'entends donc par « évidentielle » la propriété d'être fondé sur des preuves.
25. Considérant que le seul véhicule qui soit à la fois personnel et linguistique est le langage naturel, il en découle que le véhicule doit être le langage naturel.
26. J. L. Bermudez, *Thinking about Words*, Oxford, Oxford University Press, 2003, p. 158.
27. Bermudez utilise aussi le terme « pensée réflexive ».
28. J. L. Bermudez, *Thinking about Words*, Oxford, Oxford University Press, 2003, p. 158.

29. A. Clark, «Dealing in futures: Folk psychology and the role of representations in cognitive science» dans *The Churchlands and Their Critics*, Oxford, 1996, p. 177.
30. R. W. Lurz, «In Defense of Wordless Thoughts About Thoughts» dans *Mind & Language*, vol. 22, n° 3, 2007, p. 275.
31. J.L. Bermudez, *Thinking about Words*, Oxford, Oxford University Press, 2003, p. 159.
32. De manière implicite, Bermudez présuppose ici que les véhicules sont soit personnels, soit sous-personnels, et donc qu'il n'existe pas d'autres types de véhicules.
33. *Ibid.*, p. 160.
34. *Ibid.*, p. 161.
35. *Ibid.*, p. 161.
36. *Ibid.*, p. 162.
37. Lurz soulève quatre grands types d'étude à ce sujet: les *standard false-belief studies* (H. Wimmer et J. Perner, «Beliefs about beliefs: Representation and constraining function of wrong beliefs in young children's understanding of deception», dans *Cognition*, vol. 13, 1983, pp. 103-125), les *false-belief explanation studies* (K. Bartsch et H. Wellman, «Young children's attribution of action to belief and desires», dans *Child Development*, vol. 60, 1989, pp. 946-964), les *false-belief deception studies* (M. Chandler, A. Fritz et S. Hala, «Small-scale deceit: Deception as a marker of two-, three-, and four-year-olds' early theories of mind», dans *Child Development*, vol. 60, 1989, pp. 1263-1277) et les *belief-desire prediction studies* (H. Wellman et H. Bartsch, «Young children's reasoning about beliefs» dans *Cognition*, vol. 30, 1989, pp. 239-277).
38. H. Wellman et H. Bartsch, «Young children's reasoning about beliefs» dans *Cognition*, vol. 30, 1989, pp. 239-277.
39. Pour ce type d'études, Lurz fait référence, entre autres, à A. Morris, «Development of logical reasoning: Children's ability to verbally explain the nature of the distinction between logical and nonlogical forms of argument», dans *Developmental Psychology*, vol. 36, 2000, pp. 741-758., D. Moshman, «Development of metalinguistic understanding», dans W.F. Overton (ed.), *Reasoning, Necessity, and Logic: Developmental Perspectives*, Hillsdale, 1990., B. Pillow, «Children's understanding of inferential knowledge», dans *Journal of Genetic Psychology*, vol. 160, 1999, pp. 419-428.
40. Lurz semble utiliser les termes «explicite» et «conscient» de manière interchangeable, tout comme les termes «implicite» et «inconscient».

41. R. W. Lurz, «In Defense of Wordless Thoughts About Thoughts» dans *Mind & Language*, vol. 22, n° 3, 2007, p. 281.
42. J. L. Bermudez, *Thinking about Words*, Oxford, Oxford University Press, 2003, p. 237.
43. R. W. Lurz, «In Defense of Wordless Thoughts About Thoughts» dans *Mind & Language*, vol. 22, n° 3, 2007, p. 282.
44. *Ibid.*, p. 283-286.
45. Je considère l'interprétation faible de (L2) comme étant trop éloignée de la thèse de Bermudez. De plus, le caractère contre-intuitif de la thèse de dépendance proposée par Lurz semble suffire à montrer que Bermudez n'accepterait pas une telle thèse.
46. *Ibid.*, p. 292.
47. *Ibid.*, p. 288.
48. *Ibid.*, p. 292.
49. P. N. Johnson-Laird, *Mental Models: Toward a Cognitive Science of Language, Inference and Consciousness*, Harvard University Press, 1983.
50. Plus précisément, Lurz accorde à Bermudez qu'une des quatre expériences pourrait être réinterprétée de la sorte, mais Lurz considère qu'une explication unifiée de ces quatre expériences est préférable. Lurz semble donc utiliser un argument classique en philosophie, soit celui de la parcimonie.
51. Lurz mentionne aussi d'autres types d'expériences portant sur des animaux non humains, mais pour critiquer (L2), il n'est question que d'enfants humains.
52. David Moshman, *Adolescent rationality and development*, Psychology Press, 2011.
53. Évidemment, Bermudez ne développe pas davantage la notion de conscience qu'il utilise. Ainsi, la critique que je propose ici ne renforce évidemment pas l'argument de Bermudez, au contraire.
54. P. N. Johnson-Laird, *Mental Models: Toward a Cognitive Science of Language, Inference and Consciousness*, Harvard University Press, 1983.
55. D. Braddon-Mitchell et F. Jackson, *Philosophy of Mind and Cognition*, Oxford: Blackwell, 1996.
56. Pour plus d'information sur le débat entre les pictorialistes et les descriptionnistes, voir l'entrée «Mental Imagery» dans le *Stanford Encyclopedia of Philosophy*.



# La métacognition chez l'animal non humain

KEBA COLOMA CAMARA, *Université Laval*

## *Introduction*

L'homme est depuis très longtemps considéré comme un être supérieur à l'animal non humain. Cette opinion est largement défendue par Descartes<sup>1</sup> lorsqu'il considère l'animal comme une simple machine dépourvue de raison qui n'agit que par instinct; contrairement à l'homme qui possède une âme pensante. L'anthropomorphisme défend également la même opinion : il consiste à attribuer certaines capacités humaines aux animaux non humains sans pour autant accepter que les animaux possèdent réellement ces capacités<sup>2</sup>. À la question les animaux pensent-ils?, certains philosophes tels que Donald Davidson<sup>3</sup> répondent par la négative, pour la simple raison que les animaux ne possèdent pas de croyance. Toutes ces thèses considèrent l'homme comme une exception dans la nature et l'animal non humain, privé de raison, lui reste inférieur. Mais, avec les recherches et découvertes contemporaines sur la cognition animale et l'organisation sociale des animaux non humains, il semble que nous assistons à la fin du mythe de l'exception humaine professée par Jean-Marie Schaeffer<sup>4</sup> et le début de la redéfinition de l'homme proposée par Joëlle Proust<sup>5</sup>.

La rationalité qui représentait jusque-là la frontière entre l'homme et l'animal peut être conçue, selon certains chercheurs, comme un produit de l'évolution<sup>6</sup> et il existe des formes primitives de rationalité chez les animaux non humains<sup>7</sup>. De nombreuses expériences accréditent cette thèse en s'appuyant sur les principes de l'évolution biologique. La raison n'est plus identifiée à l'âme, elle est plutôt considérée comme un système de commande cognitif qui permet à l'animal et à l'homme de réaliser et de contrôler leurs cognitions. Si les chercheurs, dans le domaine de la cognition animale, acceptent cette définition minimale de la rationalité; ils

ne sont pas, pour la plupart, convaincus que l'animal non humain possède la capacité métacognitive : c'est-à-dire la capacité à suivre et à évaluer les cognitions. Pour certains, cette aptitude relève du niveau supérieur de la rationalité, exclusif à l'homme, et l'animal qui n'a atteint que le niveau primitif est incapable de manifester des aptitudes métacognitives. Est-ce un mythe philosophique ou un raisonnement scientifique ? La métacognition<sup>8</sup> représente, pour le moment, une frontière cognitive qui sépare l'homme de l'animal non humain.

Les animaux non humains ont-ils des aptitudes métacognitives ? La réponse à cette question varie selon les théories et les philosophes. Deux conceptions s'opposent : la première, la conception auto-attributive, considère que pour qu'un organisme soit capable d'aptitude métacognitive, il doit posséder au préalable une théorie de l'esprit et le mindreading<sup>9</sup> (la capacité d'attribution d'états mentaux à autrui). Selon cette conception, si l'animal non humain ne dispose pas de ces éléments alors il ne possède tout simplement pas la capacité métacognitive. Cette hypothèse est rejetée par la seconde conception : la conception auto-évaluative. Cette conception, en s'appuyant sur des résultats en neurosciences cognitives, estime que les animaux non humains ont la capacité métacognitive ; car la métacognition est un système de commande qui ne nécessite pas l'acquisition du mindreading ni d'une théorie de l'esprit.

Dans les sections qui suivent, nous proposerons, d'abord, une définition du concept de métacognition. Ensuite, nous examinerons les arguments de la conception auto-attributive de la métacognition et, notamment, l'approche de Peter Carruthers qui considère que les preuves métacognitives découvertes chez les animaux non humains, défendues par la conception auto-évaluative, ne sont pas suffisantes pour attribuer la capacité métacognitive aux animaux non humains. Dans la troisième section, nous discuterons des arguments de la conception auto-évaluative qui défend l'existence des capacités métacognitives chez les animaux et les problèmes auxquels ils sont confrontés, avant d'examiner les différentes formes de métacognition. Nous terminerons par les perspectives de Joëlle Proust qui mettent

en valeur les enjeux de l'attribution de la métacognition chez l'animal non humain.

### *1. Qu'est-ce que la métacognition ?*

La définition la plus simple à retenir de la métacognition est la suivante : c'est une cognition de la cognition. Lorsqu'on parle de cognition, on fait référence aux processus cérébraux et mentaux tels que le traitement de l'information, la mémoire, la connaissance, le raisonnement, l'apprentissage. La métacognition est alors la cognition qui permet d'évaluer, de contrôler et de réguler le fonctionnement des autres cognitions.

Le concept de métacognition est utilisé tout aussi bien en psychologie cognitive, en philosophie qu'en sciences de l'éducation. Même si son usage est relatif par rapport aux préoccupations de chaque domaine, il y a tout au moins deux aspects qui entrent en jeu lorsqu'on utilise ce concept<sup>10</sup> : il s'agit du savoir métacognitif et de l'expérience métacognitive. Le premier représente les métaconnaissances c'est-à-dire les savoirs et les croyances stockées dans la mémoire à long terme qui affectent les processus cognitifs. Ces métaconnaissances permettent aux individus, par exemple, de savoir comment :

- agir dans le cadre des relations humaines,
- trouver la disponibilité d'une information,
- opter pour les meilleures stratégies face aux situations, etc.

Le second aspect, les expériences métacognitives, renvoie à l'évaluation « des expériences cognitives et affectives conscientes<sup>11</sup> ». Le contrôle, le suivi et l'évaluation de ces expériences cognitives et affectives permettent le contrôle des émotions et d'anticiper ou de planifier l'action du sujet face aux situations qu'il rencontre. Ainsi l'individu est-il généralement en mesure de contrôler son niveau de confiance par rapport aux situations qu'il rencontre.

Proust<sup>12</sup> distingue deux fonctions de l'activité métacognitive. La première, écrit-elle, permet de décider de la faisabilité d'une activité mentale. Par exemple, si nous proposons à un enfant de

l'école primaire de résoudre une équation mathématique à deux inconnus, il n'a certainement pas encore appris à résoudre ce type d'équation ; il doit être en mesure de savoir, après une rapide évaluation métacognitive, s'il est compétent ou non à résoudre ce type d'équation avant même de tenter quoi que ce soit. Si on propose la même équation, par ailleurs, à un étudiant en mathématique, qui a probablement rencontré ce type de problème mathématique dans son cursus scolaire, et qu'il accepte de le résoudre, c'est parce qu'il sait, après évaluation métacognitive, qu'il en est capable. La décision métacognitive permet, en effet, d'évaluer les aptitudes cognitives de l'individu. L'individu sait ce qu'il est capable de réaliser ou non, après avoir évalué la situation dans laquelle il se trouve par rapport à ces aptitudes cognitives.

La seconde fonction de la métacognition consiste à évaluer si l'activité cognitive ou mentale a été réussie ou non. Reprenons notre exemple : l'enfant va certainement refuser de résoudre une équation qu'il ne connaît pas ; par contre, l'étudiant en mathématique qui s'engage à résoudre le problème doit être en mesure de savoir si le résultat qu'il a trouvé est conforme au raisonnement mathématique pour lequel il a opté, faute de quoi, il peut douter de son résultat.

La métacognition est une activité cognitive qui permet, d'une part, d'évaluer, de contrôler ou de faire le suivi des activités cognitives prospectives, c'est-à-dire prendre une décision sur la faisabilité ou non d'une action cognitive (première fonction) ; d'autre part, elle permet de faire un suivi rétrospectif de l'activité cognitive, par exemple l'évaluation de la conformité du résultat par l'étudiant en mathématique (seconde fonction). La métacognition ne se limite pas, cependant, à ces deux cas de figure ; elle se manifeste également dans l'évaluation de la cognition dans les tâches quotidiennes telles que l'apprentissage, le jugement ; etc.

Il existe, en outre, un autre élément essentiel dans l'activité métacognitive : c'est le rôle des sentiments. Certains philosophent<sup>13</sup>, en effet, ont défendu la thèse selon laquelle les émotions et la raison représentent deux niveaux distincts, les émotions sont un obstacle à la raison<sup>14</sup>. Pour parvenir à un raisonnement ou un jugement rationnel, l'individu, croient-ils, doit pouvoir contrôler ses émotions

en s'en détachant, car c'est seule la raison qui permet d'atteindre l'objectivité dans la réflexion. Les sentiments représentent quelquefois des obstacles au raisonnement rationnel et objectif; mais est-ce une raison de les négliger dans les activités rationnelles? D'autres philosophes<sup>15</sup> ont montré, par contre, l'importance des sentiments dans le raisonnement. Les sentiments jouent, en effet, un rôle épistémique dans l'évaluation d'une activité cognitive. Proust<sup>16</sup> distingue différentes fonctions des sentiments dans l'activité métacognitive: par exemple, dans une tâche métacognitive prospective ou rétrospective, la réponse métacognitive peut être un sentiment d'incertitude ou de confiance, de familiarité, de savoir, d'avoir un mot sur le bout de la langue, d'échec, de cohérence, d'incohérence, etc.

La raison et les sentiments entretiennent alors une relation complexe et complémentaire. L'être rationnel est tout aussi bien l'être sentimental. La raison n'a pas tout à fait le pouvoir d'exclure les sentiments; ces dernières sont tout aussi bien en mesure de l'orienter pour les choix meilleurs dans une tâche cognitive. Le rôle des sentiments métacognitifs dans les évaluations cognitives doit nous orienter à penser que la capacité métacognitive n'est pas l'apanage exclusif de l'homme, l'animal non humain doué de sentiment et, au moins, une forme primitive de raison peut aussi en être capable. Car l'évaluation métacognitive est, d'ailleurs, motivée à la fois par les émotions et par la raison.

Le problème que nous soulevons à propos du concept de métacognition est le suivant: à l'instar du sujet humain qui fait usage des aptitudes métacognitives pour évaluer et contrôler sa cognition, l'animal non humain a-t-il des aptitudes métacognitives? Il existe une hypothèse à cette question: si l'animal non humain a des aptitudes cognitives, il est tout à fait possible que ses activités cognitives soient accompagnées d'activités métacognitives, et ces dernières permettent de contrôler et d'évaluer les premières. Mais, cette hypothèse, aussi cohérente qu'elle soit, fait face à des problèmes conceptuels tels que le *mindreading* et la théorie de l'esprit.

Deux approches vont s'opposer dans l'étude de la métacognition chez l'animal non humain. En effet, les défenseurs du *mindreading*<sup>17</sup>

estiment qu'il ne peut exister une activité métacognitive chez un individu qui ne possède pas une théorie de l'esprit<sup>18</sup> : c'est la conception auto-attributive ou la Self-Ascriptive View (SAV). Le mindreading et la métacognition, défendent-ils, jouent des rôles analogues : le mindreading à propos d'autrui et la métacognition à propos de soi. Alors la théorie de l'esprit est une condition nécessaire et suffisante pour la métacognition. Contrairement aux partisans de la SAV, ceux de la conception auto-évaluative ou la Self-evaluative View (SEV) estiment que la métacognition et le mindreading jouent deux fonctions<sup>19</sup> distinctes qui peuvent néanmoins se compléter. Alors les animaux non humains qui ne disposent pas d'une théorie de l'esprit sont tout à fait capables de posséder des aptitudes métacognitives.

L'enjeu du problème de la métacognition est d'abord théorique : la définition de la métacognition est différente dans la SAV et la SEV. Pour la première conception, défendue par les spécialistes de la théorie de l'esprit, la métacognition renvoie à la métareprésentation de ses propres états mentaux<sup>20</sup>. Cela signifie que la capacité métacognitive est déterminée par la théorie de l'esprit c'est-à-dire les connaissances dont on dispose sur son propre esprit. Or, ces connaissances sont formées par des représentations portant sur d'autres représentations acquises culturellement ou de façon innée. La métacognition a, selon cette conception, une structure métareprésentationnelle puisque les connaissances métacognitives portent nécessairement sur des contenus représentationnelles. La métacognition et la métareprésentation font donc appel à des capacités similaires. Cependant, pour la seconde conception, acceptée principalement en sciences cognitives, les connaissances métacognitives ne sont pas nécessairement métareprésentationnelles. Selon cette conception, la métacognition est la capacité dont dispose un sujet pour évaluer sa performance cognitive<sup>21</sup>. La capacité métareprésentationnelle ou la théorie de l'esprit n'est donc pas une condition nécessaire à l'activité métacognitive. Les animaux non humains, en effet, qui ne disposent pas de théorie de l'esprit sont tout à fait en mesure de posséder la capacité métacognitive. C'est la perspective de la SEV que nous espérons défendre et approfondir dans les sections qui suivent. Mais, pour le moment, examinons l'orientation de la SAV.

## *2. La SAV: Carruthers et la HOT.*

### *2.1 L'approche auto-attributive (SAV): la métacognition une capacité strictement humaine*

Même si les versions sont différentes, il existe néanmoins un point sur lequel les théoriciens de la SAV sont unanimes: pour qu'un individu puisse posséder une capacité métacognitive, il doit nécessairement posséder une théorie de l'esprit et avoir la capacité de lire l'esprit des autres. La métacognition passe par la métareprésentation et cette dernière nécessite l'acquisition du mindreading. Ils considèrent, en général, la métacognition comme une activité du second ordre.

Certains théoriciens de la SAV s'inspirent, en effet, de ce schéma de Nelson et Narens<sup>22</sup>: le fonctionnement de tout système, précisent-ils, a besoin nécessairement de deux niveaux, un premier niveau orienté dans l'exécution des tâches et un second niveau qui permet le suivi et le contrôle des tâches exécutées au premier niveau. Tout ce qui est réalisé, en effet, au premier niveau est vérifié et contrôlé au second niveau, cela permet d'assurer le bon fonctionnement du système, c'est-à-dire la correspondance entre la commande, l'exécution et le résultat attendu. La conception auto-attributive, s'appuie sur ce modèle pour expliquer que la métacognition est une activité cognitive du second ordre qui doit nécessairement représenter la tâche cognitive du premier ordre. Cela signifie que la métacognition passe par le mindreading qui implique la métareprésentation. En effet, la représentation métacognitive s'appuie nécessairement, selon les théoriciens de la SAV, sur d'autres représentations mentales c'est-à-dire d'autres états mentaux du sujet qui s'auto-attribue, finalement, le résultat métacognitif. Si je crois, par exemple, que P est Y c'est que j'ai déjà des croyances à propos de P ce qui me permet de le qualifier de Y.

Selon John H. Flavell<sup>23</sup>, un des premiers théoriciens de la SAV, il y a une relation entre la métacognition et l'attribution d'états mentaux à autrui. Les enfants âgés de moins de 4 ans, selon cet auteur, ne possèdent pas le mindreading car ils sont incapables de métareprésenter leurs croyances et celles des autres. Aussi sont-ils incapables de distinguer l'apparence et la réalité des objets et

phénomènes qui les entourent. Mais, dès qu'ils comprennent, après 4 ans, qu'ils possèdent un esprit capable de se représenter soi-même, les autres et le monde, alors ils parviennent à discriminer les croyances et à distinguer ce qui relève de la réalité et de l'apparence. Flavell considère en fait, à l'instar des praticiens des sciences de l'éducation, que la métacognition fait partie de la cognition sociale. Les connaissances métacognitives sont inséparables de la connaissance de soi et des autres. Il est donc nécessaire de posséder une théorie de l'esprit qui permet d'évaluer les connaissances métacognitives, car sans elle, il n'y aura pas de capacité métacognitive. Un autre théoricien de la SAV Joseph Perner<sup>24</sup> abonde dans le même sens : pour lui, l'enfant n'acquiert une théorie de l'esprit et une capacité métacognitive qu'au moment où il est capable de se faire une représentation du rapport entre une représentation mentale X et l'état d'un objet Z. L'activité métacognitive a, selon la SAV, la même structure cognitive que celle du mindreading<sup>25</sup> (connaître ses propres états mentaux et inférer ceux des autres).

Si tel est le cas, il est impossible d'attribuer aux animaux la capacité métacognitive étant donné que l'animal non humain ne possède pas une théorie de l'esprit. La métacognition est, selon la SAV, une capacité strictement humaine. L'animal peut, toutefois, avoir une pensée du premier ordre pour agir dans son environnement, mais il ne possède pas la pensée du second ordre, c'est-à-dire l'aptitude métacognitive, puisqu'il ne possède pas de mindreading.

## *2.2 Carruthers, la théorie de la pensée d'ordre supérieur (HOT) et la métacognition*

Peter Carruthers<sup>26</sup>, un partisan de la SAV, accepte l'argument selon lequel l'activité métacognitive a la même structure cognitive que le mindreading ; autrement dit, la métacognition et le mindreading n'utilisent qu'une seule voie d'accès : l'agent s'attribue des représentations mentales qu'il infère d'autres représentations. Mais il propose, avec sa théorie de la pensée d'ordre supérieur<sup>27</sup> (Higher-order Thought ou HOT), une explication du premier ordre de la plupart des activités métacognitives. Selon cet auteur, la métacognition et le mindreading se font de façon conceptuelle,

c'est-à-dire inférentielle ou par métareprésentation. En effet, si l'accès à la mémoire ou à la perception se fait de façon directe et non inférentielle, la métacognition, cependant, passe nécessairement par inférence ou par métareprésentation. L'idée est que les connaissances métacognitives passent nécessairement par le mindreading et que ce dernier requiert des capacités métareprésentationnelles. Or, sans le mindreading, il ne peut y avoir de métacognition. Les connaissances métacognitives représentent les capacités représentationnelles dont on dispose : après inférence le sujet s'attribue ces connaissances.

Carruthers défend la thèse suivante : la métacognition est une représentation contenue dans une autre représentation. Il est possible à l'homme de recourir à l'explication métacognitive, puisqu'il possède une pensée d'ordre supérieur, c'est-à-dire une théorie de l'esprit ; mais l'animal non humain qui ne possède pas ce type de pensée est, du coup, incapable d'exercer une activité métacognitive. Tout ce que l'animal est capable de faire relève de la cognition et il n'existe pas du tout d'activité métacognitive. Pourtant les études récentes ont montré que les singes et les dauphins ont des aptitudes métacognitives<sup>28</sup>. En outre, selon certaines expériences, qui sont pour le moment discutables, les singes possèdent la capacité du mindreading<sup>29</sup>. Carruthers écrit :

I shall subject the main studies referred to above to sustained critique, arguing that there is no need to postulate *meta-cognitive*<sup>30</sup> processing in order to explain the data. Rather, I shall show how those data admit of explanation in first-order terms, appealing only to states and processes that are world-directed rather than self-directed. I shall argue, in consequence, that we should, at present, refuse to attribute *meta-cognitive*<sup>31</sup> processes to animals<sup>32</sup>.

Carruthers parvient, en effet, à proposer une explication du premier ordre à tous les exemples<sup>33</sup> que certains auteurs<sup>34</sup> qualifient de preuves métacognitives chez l'animal non humain. Mais l'explication qu'il propose est finalement très sophistiquée puisqu'elle traite de différents types de cas. Étant donné qu'il n'accorde pas l'aptitude métacognitive à l'animal non humain, il applique la même rigueur

d'explication du premier ordre aux croyances de l'homme dans le but de réduire la frontière cognitive entre l'homme et l'animal. En d'autres termes, il est possible d'expliquer à la fois le comportement de l'animal comme celui de l'homme en faisant référence au mécanisme du premier ordre c'est-à-dire comment l'homme et l'animal parviennent à avoir des désirs et des croyances, à les évaluer et à réaliser des actions cognitives.

Même si la démarche de Carruthers paraît quelquefois convaincante, il n'en demeure pas moins que la thèse qu'il défend dans sa théorie de la HOT fait face à un enjeu déterminant quant au statut de la cognition animale. Deux éléments nous paraissent suffisants pour s'écarter de la perspective des théoriciens de la SAV et notamment de celle de Carruthers. Premièrement, les études récentes en sciences cognitives et en éthologie démontrent de plus en plus que l'action cognitive des animaux non humains est tout aussi complexe que celle des humains ; et il existe assez de preuves convaincantes que les animaux non humains possèdent une capacité métacognitive.

Le second élément est épistémologique, il est structuré à deux niveaux. D'abord, il nous semble impossible qu'un système puisse causer des activités cognitives sans pour autant être capable de réaliser des activités métacognitives qui sont censées évaluer les activités cognitives. Alors il est tout à fait cohérent de présumer que les animaux non humains possèdent la capacité métacognitive ; sinon comment parviendront-ils à discriminer les phénomènes, à choisir leur proie et à vivre dans leur milieu ? Or toutes ces situations nécessitent, à notre avis, l'activité métacognitive, même si les partisans de la SAV, principalement Carruthers, préfèrent analyser ces situations sous l'angle de la cognition plutôt que la métacognition. Ensuite, l'outil conceptuel utilisé par les théoriciens de la SAV et principalement Carruthers (HOT) impose des critères censés être simples et suffisants pour évaluer la cognition animale, mais nous constatons qu'ils sont, paradoxalement, complexes et leur application s'avère difficile dans l'étude de l'esprit des animaux non humains. Ces auteurs commettent une erreur de principe lorsqu'ils estiment que si l'animal ne possède pas, par exemple, le mindreading

alors il n'a pas de capacité métacognitive. Comme nous le verrons dans la section suivante, le mindreading n'est pas nécessaire pour la métacognition.

### *3. La SEV et l'approche de Proust*

#### *3.1 L'approche auto-évaluative (SEV)*

Contrairement à la thèse défendue par les partisans de la SAV, les auteurs de la SEV tels que Proust et Koriat<sup>35</sup> considèrent que la fonction de la métacognition est différente de celle du mindreading. Cela signifie qu'il est possible pour un organisme de posséder la capacité métacognitive sans pour autant acquérir une théorie de l'esprit et le mindreading. En ce sens, les animaux non humains peuvent avoir la capacité métacognitive. Mais comment parviendront-ils à évaluer et à contrôler leurs états mentaux sans pour autant lire le contenu des croyances? L'enjeu de cette question représente l'essentiel du défi lancé par la SAV. Il existe quelques arguments pour répondre à cette question, examinons-les.

D'abord, les progrès réalisés dans la dernière décennie en psychologie comparative ont été favorables à la conception auto-évaluative. Les études sur la cognition animale faites par Hampton<sup>36</sup> et par Smith<sup>37</sup> ont permis de démontrer que certains animaux tels que le singe macaque et le dauphin manifestent des capacités métacognitives remarquables alors qu'ils ne possèdent ni la théorie de l'esprit ni le mindreading. Dans les différents tests auxquels ces animaux ont été soumis, lorsqu'ils sont en face de deux tâches A et B, ils choisissent la tâche la plus facile qu'ils sont capables d'effectuer; lorsqu'ils sont incapables de réaliser les tâches, ils refusent tout simplement de poursuivre les tests. Le macaque et le dauphin ont réussi les tests de discrimination visuelle et du rappel de mémoire, établis comme des critères démontrant des aptitudes à évaluer leur incertitude, donc qu'ils possèdent des capacités métacognitives. Pourtant ils échouent à des expériences mentalisatrices quand il s'agit de prédire le comportement d'un agent à partir de ses croyances et intentions. Ces animaux non humains parviennent à faire le contrôle et le suivi de leur cognition sans pour autant disposer d'une théorie de l'esprit ou du mindreading.

De ces expériences, on peut retenir la remarque suivante : la théorie de l'esprit n'est pas nécessaire pour la capacité métacognitive ; un organisme peut alors évaluer ses aptitudes cognitives sans posséder le mindreading. Mais, selon Carruthers, ce qui est considéré ici comme preuve métacognitive est tout simplement une cognition du premier ordre. Car la métacognition nécessite la métareprésentation et un organisme qui ne possède pas la pensée d'ordre supérieur (HOT) serait incapable de faire des inférences métacognitives.

Ensuite, des découvertes récentes faites en phylogénie et en neurophysiologie ont permis de distinguer les mécanismes de fonctionnement de la métacognition et du mindreading<sup>38</sup>. En neurophysiologie<sup>39</sup>, il a été démontré chez l'homme que les aires cérébrales qui entrent en jeu dans la métacognition et le mindreading sont distinctes : pour la métacognition, ce sont les aires sensorielles, le cortex préfrontal dorsal, le cortex préfrontal ventro-médian et le cortex cingulaire antérieur, etc. ; pour le mindreading, c'est le cortex préfrontal antéro-médian, la zone temporo-pariétale, etc. En phylogénie il y a une distinction entre le mécanisme de fonctionnement de la métacognition procédurale et la métacognition analytique<sup>40</sup> : «the first type relies on a variety of mechanisms to detect and control error ; the second is a distinct adaptation, which enables agents to understand error as false belief»<sup>41</sup>.

Enfin, l'argument décisif sur lequel la conception auto-évaluative s'appuie est conceptuel. En effet, les concepts de métacognition, de métareprésentation ou de mindreading peuvent être considérés comme des sous-ensembles distincts et complémentaires, mais ils ne sont pas nécessairement dépendants<sup>42</sup>. Il y a une hypothèse logique : le mindreading implique la métareprésentation et, souvent, la métacognition ; mais, une activité métacognitive peut être réalisée sans pour autant impliquer la métareprésentation. Les schizophrènes, par exemple, sont atteints de différentes sortes de perturbations psychologiques ; ils attribuent leurs actions à des agents extérieurs. Mais, c'est ce qui est remarquable, ces patients sont capables de rapporter et d'évaluer leurs états mentaux qu'ils attribuent aux autres. Ils délirent, mais ils rapportent parfaitement le contenu de leurs délires : ils sont d'ailleurs hyper réflexifs et rapportent même

les détails de leurs délires. En ce sens, la capacité métacognitive qui consiste à évaluer une action cognitive est distincte de la capacité métareprésentationnelle et du mindreading qui permettent de conceptualiser et d'inférer le contenu des représentations cognitives.

La métacognition est, généralement, directe et évaluative alors que le mindreading et la métareprésentation se font de façon indirecte et par inférence. Lorsqu'on me demande si je connais la nationalité du philosophe Kierkegaard<sup>43</sup>, avant de répondre à cette question, le mécanisme métacognitif me permet de savoir si je dispose des informations à propos de ce nom : voici une évaluation métacognitive. Le résultat de l'évaluation déterminera, par ailleurs, si je dois me souvenir au cas où je sais que je possède des informations sur cette personne ; à défaut, je répondrai que je n'ai aucune connaissance de cette personne.

Pour dépasser l'opposition entre ces deux principales conceptions, à savoir la SAV et la SEV, nous proposons d'examiner les différentes formes de métacognition dans lesquelles chacune de ces approches peut trouver un écho de ses arguments.

### *3.2 Les formes de métacognition*

La métacognition définie comme une cognition sur la cognition est sans doute difficile à vérifier chez les animaux non humains. Robert Hampton<sup>44</sup> a proposé quatre critères pour opérationnaliser le concept de métacognition :

- a) il faut nécessairement un comportement primaire identifiable ;
- b) ensuite, une incertitude du résultat : l'agent doit faire face à une situation d'incertitude à partir de laquelle il doit chercher une solution métacognitive à partir de ses capacités cognitives ;
- c) puis, il faut un comportement secondaire qui permet de réguler le comportement primaire : suivi et évaluation des capacités pour l'exécution ou non du comportement primaire ;
- d) enfin, le résultat du comportement secondaire est nécessairement évalué.

En appliquant ces critères dans la cognition animale, Hampton a réussi à distinguer deux formes de métacognition : *la métacognition publique et la métacognition privée*. La première fait référence aux mécanismes métacognitifs publiquement évaluables. Par exemple, un enseignant pose une question à un élève sur le nombre de planètes, la réponse de l'élève sera non seulement basée sur ses connaissances acquises, mais le reste de la classe peut évaluer la conformité de la réponse de leur camarade par rapport aux connaissances acquises durant la leçon sur les planètes. Il est également possible d'évaluer les aptitudes métacognitives de l'animal non humain dans des tests publiquement observables. Dans la métacognition publique, les mécanismes métacognitifs se traduisent dans le comportement observable (hésitation dans le test, refus, acceptation, etc.) La seconde forme permet de surveiller et de contrôler ses processus cognitifs non observables publiquement. En fait, dans la métacognition privée il s'agit d'une évaluation personnelle sur une situation cognitive particulière dont on ne peut observer les mécanismes. L'introspection est un moyen qui permet d'expérimenter et de vérifier les mécanismes de la métacognition privée.

Proust<sup>45</sup> propose également une distinction importante entre deux types de métacognition. Il s'agit de la *métacognition théorique et la métacognition procédurale*. Les mécanismes de la métacognition théorique dépendent de la métareprésentation, par exemple, une représentation de la croyance et de son contenu. La métacognition théorique est ce qu'on appelle également métacognition analytique. Ce type de métacognition nécessite une théorie de l'esprit avec le mindreading c'est-à-dire la capacité à lire le contenu des croyances. La métacognition théorique ou analytique se fait alors par inférence par le moyen de la métareprésentation.

La métacognition procédurale s'articule principalement sur l'évaluation des aptitudes cognitives en fonction des objectifs cognitifs visés, elle ne nécessite pas une métareprésentation parce qu'elle accède directement à l'information sans faire des inférences. Proust note deux fonctions importantes de cette forme de métacognition : une fonction instrumentale et une fonction exécutive. La fonction instrumentale est orientée vers l'obtention du

résultat ou la récompense visée par l'agent. Par exemple, l'animal dans son environnement est capable de discriminer le type de proie qui l'intéresse le mieux et choisir le moyen (la technique de chasse) qu'il doit utiliser pour atteindre son objectif. La fonction exécutive quant à elle consiste à organiser et à allouer les ressources cognitives indispensables au succès de l'événement métacognitif. Dans l'organisation et l'utilisation des ressources, il y a nécessairement l'évaluation et la surveillance de la cognition par rapport à la norme ou l'objectif métacognitif visé. On peut dire que la fonction instrumentale de la métacognition se rend possible par le concours de la fonction exécutive.

On peut noter, à la lumière de ces formes de métacognition, que la conception auto-attributive ne prend en charge que la métacognition analytique ou théorique. Car ces formes de métacognition nécessitent l'acquisition d'une théorie de l'esprit, le mindreading et la métareprésentation. Le mécanisme de ces formes de métacognition est indirect parce qu'il se réalise par inférence. La métacognition procédurale, essentiellement épistémique, est prise en charge par l'orientation de la SEV. Le mécanisme de cette forme de métacognition est direct et permet d'évaluer les aptitudes en fonction des situations précises. En somme, l'usage du concept de métacognition doit être déterminé par les formes de métacognition. On peut donc affirmer que l'animal non humain a des capacités métacognitives en se référant à la métacognition procédurale.

Mais la métacognition procédurale essentiellement épistémique est une forme métacognitive pratique et limitée. Cela représente un problème dans l'approche de la SEV et donne souvent raison aux partisans de la SAV. En effet, là où l'homme est capable de s'attribuer des connaissances métacognitives qui passent par la métareprésentation; l'animal non humain, par la métacognition procédurale, n'agit que de façon directe. En ce sens, il serait difficile à l'animal de faire un retour sur ses propres états mentaux, savoir que c'est lui qui est l'auteur de ses pensées. Le problème pour la SEV est finalement d'aller au-delà de cette forme restrictive de la métacognition procédurale et de voir si les animaux non humains possèdent une théorie de l'esprit qui représente l'élément central

pour l'acquisition de toutes les formes de métacognition. Des études faites sur les singes sont prometteuses, mais elles sont confrontées, pour le moment, à des problèmes méthodologiques que les partisans de la SEV doivent surmonter.

Le dernier élément à explorer dans ce travail est le suivant : à quoi sert l'attribution de la capacité métacognitive à l'animal non humain ? En analysant la perspective des travaux de Joëlle Proust, nous trouvons des réponses à cette question.

### *3.3 Cognition, métacognition et statut de l'animal non humain*

Le débat sur la rationalité de l'animal non humain a opposé différentes thèses à propos de l'attribution ou non de la raison aux animaux non humains. Contrairement à certains philosophes comme Descartes<sup>46</sup> qui estiment que la raison est une propriété strictement humaine ; d'autres comme Davidson<sup>47</sup> qui pensent que le langage est un critère de la représentation mentale ; d'autres comme Bermudez et Proust ont essayé de démontrer les aptitudes rationnelles chez les animaux non humains. Proust<sup>48</sup> définit la raison comme un système de commande capable de s'adapter aux situations auxquelles l'agent est confronté et ce système de commande est déterminé par les aptitudes et connaissances dont dispose l'agent.

Il existe, en effet, une relation entre la rationalité et la métacognition. Mais le sens que l'on attribue au concept de rationalité détermine son rapport avec la métacognition. En général, la rationalité est définie comme la capacité dont dispose un agent à respecter les normes du raisonnement logique<sup>49</sup>. Chez les partisans de la SAV, la rationalité correspond à l'acquisition de la théorie de l'esprit et du mindreading. Être rationnel dans la perspective de la SAV, exige posséder un esprit, être capable de se représenter ses croyances et celles des autres. Cette forme de rationalité implique le mindreading, la métareprésentation et, du coup, ne prend en charge que la forme de la métacognition analytique ou théorique. La rationalité définie chez Proust comme une flexibilité ou une adaptabilité d'un système de commande est différente de la perspective de la SAV ; elle ne prend en compte que la forme de métacognition procédurale. Il est donc tout à fait possible qu'un animal soit rationnel et possède la capacité métacognitive,

dans la perspective de Proust, sans nécessairement posséder une théorie de l'esprit et le mindreading ; et sans qu'il soit nécessairement conscient. La rationalité, entendue comme adaptabilité et flexibilité du système de commande représente la capacité cognitive de l'agent. Cette capacité cognitive entretient alors un rapport étroit avec la métacognition procédurale qui permet, par exemple, d'évaluer la relation entre les ressources cognitives mobilisées par le système de commande d'un animal, et le but et les stratégies d'action à entreprendre.

Dans ce cadre, la métacognition procédurale permet aux animaux non humains d'évaluer leur cognition et le comportement des autres agents dans leur environnement de façon mentaliste. Autant que le sujet humain évalue sa pensée avant d'agir, autant l'animal non humain se sert de la métacognition procédurale avant d'engager une action en vue d'obtenir un résultat. Être rationnel ne nécessite pas chez un agent de posséder une théorie de l'esprit ou de représenter ses croyances par le langage ; c'est en fait la capacité cognitive de l'animal non humain. La métacognition procédurale permet à ce dernier d'avoir un comportement rationnel, en fonction des situations, en évaluant et en optimisant les ressources cognitives indispensables pour obtenir un résultat. Les capacités métacognitives sont en fait des preuves indubitables et élémentaires de la rationalité d'un animal non humain.

La propriété fondamentale de la métacognition est, écrit Proust, par définition, celle d'un système de contrôle opérant sur le système cognitif. La métacognition vise à assurer l'adéquation cognitive des moyens dont dispose le système et de ses buts cognitifs : il s'agit d'obtenir de l'information au niveau nécessaire et suffisant, ou de maintenir l'intensité de l'engagement ou du contrôle critique au niveau requis<sup>50</sup>.

La métacognition conduit également à une forme de connaissance et de contrôle de soi. Elle est, d'une part, une forme de contrôle de soi en ce sens qu'elle permet à l'agent de faire le suivi et l'évaluation de sa cognition ; d'autre part, elle permet à l'agent d'avoir certaines formes supérieures de conscience : la conscience de soi. Les preuves

métacognitives retrouvées chez le singe macaque par Hampton indiquent que la métacognition est un moyen de conscience de soi. Si l'animal non humain parvient à contrôler et à évaluer ses cognitions, il peut également être conscient qu'il est l'auteur de ses activités cognitives. La connaissance de ses aptitudes cognitives est aussi un moyen de connaissance de soi. «C'est en effet, écrit Proust, grâce à cette structure de contrôle que l'on peut se reconnaître soi-même comme tel, par opposition au cas où l'on se perçoit «objectivement», c'est-à-dire sans se reconnaître comme étant soi-même<sup>51</sup>» La métacognition est alors une condition nécessaire à la connaissance de soi. Cette condition n'est, cependant, pas suffisante, car il n'est pas évident que l'animal non humain puisse se reconnaître comme étant toujours l'auteur de ses propres cognitions. Cela ne peut être possible qu'avec la forme de métacognition analytique qui passe par la métareprésentation. C'est d'ailleurs, nous le rappelons, la difficulté principale à laquelle la SEV est confrontée. La métacognition procédurale est certes utile dans l'action pratique de la vie des animaux non humains, mais elle est limitée puisqu'elle ne permet pas de prouver rigoureusement d'autres formes de pensées telles que la conscience phénoménale et la représentation des croyances. Seule la métacognition analytique implique ces formes de pensée. Par contre, cela ne signifie pas que les animaux ne possèdent pas ces formes supérieures de pensée. Gennaro<sup>52</sup>, un partisan de la SAV, accepte non seulement que les animaux non humains aient la capacité métacognitive, mais qu'ils possèdent la conscience phénoménale sans pour autant acquérir la HOT. Le problème de la métacognition chez l'animal non humain est à la fois un défi, pour prouver que les animaux non humains possèdent certaines capacités métacognitives et une perspective, dans les recherches sur la relation que cette capacité mentale peut entretenir ou non avec les autres formes de pensées complexes.

### *Conclusion*

Le problème de la métacognition chez l'animal non humain est, de prime abord, un problème théorique. Le problème est tout à fait lié à la définition de la métacognition selon les différentes conceptions.

La métacognition est définie, en effet, selon les préoccupations conceptuelles de chaque conception: dans la SAV elle sert à s'auto-attribuer des cognitions tandis que dans la SEV elle sert à auto-évaluer ses aptitudes cognitives. Toutefois, en opérationnalisant ce concept dans la perspective de Hampton, on se rend compte qu'il n'existe pas une seule forme de métacognition, mais plusieurs formes qui se retrouvent dans les usages des différentes théories. Cela permet, néanmoins, de valider l'option de l'approche auto-évaluative sans invalider entièrement les considérations conceptuelles de l'approche auto-attributive. La métacognition procédurale permet de reconnaître les capacités métacognitives à l'animal non humain.

Les découvertes en éthologie et en sciences cognitives accréditent davantage les arguments de la SEV. Le problème est de savoir si la théorie de l'esprit et le mindreading, chers aux partisans de la SAV, sont présents chez les animaux non humains. Les arguments ne manquent pas pour espérer cette découverte et l'organisation sociale complexe et dynamique des animaux ne cesse de dissiper les doutes et les préjugés sur le statut de l'esprit des animaux non humains. Les recherches interdisciplinaires allant de la sociologie, à la linguistique, la philosophie et les sciences cognitives peuvent encore nous surprendre. Mais, pour le moment, certains philosophes résistent avec leurs arguments et les théories s'opposent.

- 
1. René Descartes, *Méditations métaphysiques*, Paris, Flammarion, 2010 (1641).
  2. Sandra D. Mitchell, «Anthropomorphism and Cross-Species Modeling Thinking with Animals» dans Lorraine Daston and Gregg Mitman (eds.) *New Perspectives on Anthropomorphism*, New York, Columbia University Press, 2005, V.
  3. D. Davidson, «Rational Animals» dans *Dialectica*, 1982, vol. 36, n° 4.
  4. J-M. Schaeffer, *La fin de l'exception humaine*, Paris, Gallimard, 2007.
  5. Joëlle Proust, «Redéfinir l'humain», *Le Débat* 3, 2014, n ° 180, pp. 56-69.
  6. *Id.* «Rationality and metacognition in non-human animals» dans Susan Hurley & Matthew Nudds (eds.) *Rational Animals?* Oxford, Oxford University Press, 2006, pp. 247-274.

7. J-L. Bermúdez, «Rationality and Psychological Explanation without Language» dans José Luis Bermúdez & Alan Millar *Reason and Nature : Essays in the Theory of Rationality*, Oxford, Oxford University Press, 2002.
8. Joëlle Proust, «Metacognition» dans *Philosophy Compass*, 2010, 5, 11, pp. 989-998.
9. Le lecteur doit garder à l'esprit la signification donnée à ce concept. En plus, je l'utiliserai couramment dans le texte sans l'usage des guillemets à des fins de commodité de lecture.
10. Olivier Houdé, *Vocabulaire des sciences cognitives*, Paris, Quadrige, 2003, pp. 289-294.
11. *Ibid.* p. 290.
12. J. Proust, «Metacognition and mindreading: one or two functions?» dans M. Beran, J. Brandl, J. Perner & J. Proust (Eds.) *The Foundations of Metacognition*, Oxford, Oxford University Press, 2012, pp. 234-251.
13. Les philosophes tels que Platon, Descartes et Kant considèrent l'émotion et les passions de façon générale comme un obstacle à la raison.
14. E. Kant écrit à ce sujet: «L'émotion est le sentiment d'un plaisir ou d'un déplaisir actuel qui ne laisse pas le sujet parvenir à la réflexion. Dans l'émotion, l'esprit surpris par l'impression perd l'empire sur lui-même» cf. E. Kant, *Anthropologie du point de vue pragmatique*, Paris, Flammarion, 1997 (1863).
15. R. de Sousa , *The rationality of emotion*, MIT, 1990 et Jonh Elster, *Alchemies of the Mind: Rationality and the Emotions*, Cambridge University Press, 1999.
16. J. Proust, «Metacognition and mindreading: one or two functions?» dans M. Beran, J. Brandl, J. Perner & J. Proust (Eds.), *The Foundations of Metacognition*, Oxford, Oxford University Press, 2012, pp. 246-247.
17. Nous soulignons.
18. P. Carruthers, «Meta-cognition in animals : A skeptical look» dans *Mind and Language*, 2008, 23, pp. 58-89.
19. J. Proust, *op.cit.* pp. 234-251.
20. Peter Carruthers, «How we know our own minds: The relationship between mindreading and metacognition» dans *Behavioral and Brain Sciences*, 2009, 32, pp. 121-138.
21. A. Koriat et al., «The intricate relationships between monitoring and control in metacognition: Lessons for the cause-and-effect relation between subjective experience and behavior» dans *Journal of Experimental Psychology*, General, 2006, 135, 1, pp. 36-69.

22. T. O. Nelson and L. Narens, «Metamemory : a theoretical framework and new findings» dans T. O. Nelson (ed.) *Metacognition, Core Readings*, 1992, pp. 117-130.
23. J. H. Flavell, «Metacognition and cognitive monitoring: A new area of cognitive developmental inquiry» dans *American Psychologist*, 1979, 34, pp. 906-911.
24. J. Perner, T. Ruffman, «Episodic memory and automatic consciousness: Developmental evidence and a theory of childhood amnesia. Special Issue : Early memory» dans *Journal of Experimental Child Psychology*, 1995, 59, 3, pp. 516-548.
25. Onishi et Baillargeon ont essayé de prouver l'existence d'un mindreading biologique chez le bébé de 15 mois, et probablement chez les singes macaques puisque le singe macaque adulte a la même performance cognitive qu'un enfant de 3 ans. Ils estiment qu'il existe des formes précoces de mindreading, disons le proto-mindreading. La capacité à lire le contenu des croyances n'est pas strictement, selon ces auteurs, une acquisition sociale ; elle est d'abord innée. cf. K. H. Onishi, & R. Baillargeon, «Do 15-month-old infants understand false beliefs ?» in *Science*, 2005, 308, pp. 255-258.
26. Peter Carruthers, «Meta-cognition in animals: A skeptical look» dans *Mind and Language*, 2008, 23, pp. 58-89.
27. La pensée d'ordre supérieur est une pensée qui se situe au-delà de la simple cognition. Elle est par exemple rattachée aux croyances, à la conscience de soi et à la conscience phénoménale. Ces formes de pensées sont nécessairement liées à l'acquisition d'une théorie de l'esprit et du mindreading. La pensée d'ordre supérieur fait référence à la théorie de l'esprit chez Carruthers.
28. Pour le singe rhésus, cf. R. R. Hampton, «Rhesus monkeys know when they remember.» dans *Proceedings of the National Academy of Sciences U.S.A.*, 2001, 98, pp. 5359-5362 ; pour le dauphin cf. J. D. Smith et al. «The uncertain response in the bottle nosed dolphin» dans *Tursiops truncatus. Journal of Experimental Psychology: General*, 1995, 124, pp. 391-408. Ces deux auteurs démontrent par des expériences empiriques que les animaux non humains ont la capacité métacognitive.
29. Joëlle Proust, «Can non-human primates read minds?» dans *Philosophical Topics*, 2000, 27, 1, pp. 203-232.
30. Nous soulignons.
31. *Idem*.
32. Peter Carruthers, «Meta-cognition in animals : A skeptical look.» dans *Mind and Language*, 2008, 23, pp. 58-59.

33. Nous reviendrons sur certains de ces exemples dans les sections suivantes.
34. Il vise principalement l'article de Joëlle Proust, «Does metacognition necessarily involve metarepresentation?» dans *Behavior and Brain Sciences*, 2003, 26, 3, p. 352.
35. Parmi ces auteurs, on peut citer Koriat et al. (2006) Proust (2012).
36. R. R. Hampton, «Multiple demonstrations of metacognition in nonhumans: Converging Evidence or multiple mechanisms?» dans *Comparative Cognition and Behavior Reviews*, 2009, 4, pp. 17-28.
37. J.D Smith et al., «The comparative psychology of uncertainty monitoring and metacognition.» dans *Behavioral and Brain Sciences*, 2003, 26, pp. 317-373.
38. Joëlle Proust, «Metacognition and mindreading : one or two functions ?» dans M. Beran, J. Brandl, J. Perner & J. Proust (Eds.) *The Foundations of Metacognition*, Oxford, Oxford University Press, 2012, p. 248.
39. R. Kiani, R. & M. N. Shadlen, «Representation of Confidence associated with a decision by neurons in the parietal cortex» dans *Science*, 2009, 324, 5928, pp. 759-764.
40. Il existe différentes formes de métacognition, nous reviendrons sur ces deux formes de métacognition dans la section suivante. Cela évitera de confondre l'usage du terme «métacognition» chez Proust et chez Carruthers, puisqu'ils ne traitent pas de la même forme de métacognition.
41. Joëlle Proust, «Metacognition and mindreading : one or two functions ?» dans M. Beran, J. Brandl, J. Perner & J. Proust (Eds.) *The Foundations of Metacognition*, Oxford, Oxford University Press, 2012, p. 247.
42. *Id.* «La métacognition comme contrôle de soi» dans *Psychiatrie, Sciences Humaines, Neurosciences*, 2008, 6, 1, pp. 31-37.
43. Søren Kierkegaard (1813-1855) est un philosophe danois.
44. R. R. Hampton, «Multiple demonstrations of metacognition in nonhumans: Converging evidence or multiple mechanisms?» dans *Comparative Cognition and Behavior Reviews*, 2009, 4, pp. 17-28.
45. Joëlle Proust, «Metacognition and mindreading : one or two functions ?» dans M. Beran, J. Brandl, J. Perner & J. Proust (Eds.), *The Foundations of Metacognition*, Oxford, Oxford University Press, 2012, pp. 242-244
46. René Descartes, *op.cit.*
47. Donald Davidson, *op.cit.*
48. Joëlle Proust (2006) «Rationality and metacognition in non-human animals» dans Susan Hurley & Matthew Nudds (eds.) *Rational Animals ?*, Oxford, Oxford University Press, 2006, pp. 247-274.
49. O. Houdé *Op.cit.*, pp. 372-374.

50. J.Proust, «Les conditions de la connaissance de soi» dans *Philosophiques*, 2008, 27, 1, pp. 161-186.
51. *Ibid.*
52. R.Gennaro, «Animals, consciousness and I-thoughts Metacognition and mindreading : one or two functions?» dans Lurz R. (ed.) *Philosophy of Animal Minds*, Cambridge University Press, 2009, pp. 184-200.



# L'empathie chez les animaux : le modèle des poupées russes

FÉLIX AUBÉ BEAUDOIN, *Université Laval*

## *Introduction*

Darwin n'hésitait pas à prêter aux « animaux inférieurs » un répertoire émotionnel aussi riche que celui des humains. Cela lui paraissait si bien établi qu'il ne jugeait pas nécessaire, dans *The descent of man in relation to sex*, « d'ennuyer le lecteur avec beaucoup de détails<sup>1</sup> ». La confiance dont il fait preuve peut sembler naïve au lecteur contemporain. Cela dit, l'idée d'une continuité entre les animaux non humains et les humains en ce qui concerne les états mentaux, y compris les émotions, n'est pas dénuée de plausibilité. Des chercheurs de divers horizons — éthologues, primatologues, philosophes, etc. — l'acceptent jusqu'à un certain point<sup>2</sup>. Je soutiendrai dans ce qui suit que Darwin était sur la bonne voie, à tout le moins en ce qui concerne l'empathie<sup>3</sup>. Cependant, reconnaissant que son approche anecdotique est nettement insuffisante, je me permettrai, n'en déplaise au lecteur, de fournir « beaucoup de détails » afin de supporter cette thèse.

Dans un premier temps, je tâcherai de définir ce qu'est l'empathie. Je présenterai ensuite le modèle des « poupées russes » développé par Frans de Waal. Comme son nom l'indique, ce modèle suppose que l'empathie comporte plusieurs « couches » (*multi-layered*) et que les formes d'empathie les plus complexes se développent à partir des formes les plus simples. Enfin, je tenterai de répondre à deux objections importantes auxquelles fait face cette approche, soit 1) qu'elle surestime le rôle des mécanismes d'« ordre inférieur », en particulier chez les humains et 2) qu'elle postule des états mentaux inutilement complexes afin de rendre compte des comportements altruistes de certains animaux.

### *Une définition inclusive de l'empathie*

«Empathie» est un mot que l'on emploie assez couramment. Il relève de ce que l'on appelle parfois la psychologie populaire. On reproche, par exemple, à celui qui est indifférent aux malheurs d'autrui de manquer d'empathie. Inversement, on louange la personne qui y est sensible en disant qu'elle en fait preuve. Derrière cette apparente simplicité se cache un phénomène complexe et très difficile à caractériser. C. Daniel Batson recense pas moins de huit usages différents du terme dans la littérature. Il note que les choses seraient nettement plus simples si «l'empathie référerait à un seul objet et que tout le monde s'accordait sur ce qu'est cet objet», mais que, comme cela se produit souvent avec les termes psychologiques, ce n'est pas le cas<sup>4</sup>. La leçon qu'il en tire est la suivante : on ne peut que reconnaître cette complexité, préciser l'usage que l'on entend faire des termes et être cohérent ensuite dans l'application que l'on en fait<sup>5</sup>. C'est cet objectif modeste que je poursuis dans la présente section.

Certains auteurs définissent l'empathie de manière restrictive en mettant à l'avant-plan la capacité à adopter la perspective d'autrui. Cela requiert des capacités cognitives sophistiquées, une «théorie de l'esprit» ou l'imagination par exemple, qui ne sont sans doute pas largement répandues dans le règne animal. Le modèle défendu par de Waal reconnaît l'existence et l'importance d'une telle forme d'empathie. Il l'envisage cependant comme un développement tardif, rendu possible par des mécanismes plus primitifs qu'une définition adéquate devrait permettre d'inclure. De ce point de vue, le problème avec les définitions restrictives est qu'elles «sont si *top-down* qu'elles déconnectent l'empathie de ses possibles précurseurs<sup>6</sup>». Afin d'éviter la confusion, j'emploierai le terme «empathie cognitive» pour désigner la forme d'empathie à laquelle renvoient ces définitions. Le terme «empathie», lorsqu'employé seul, servira à désigner à la fois l'empathie cognitive et des réponses affectives plus primitives dont le dénominateur commun est qu'elles dénotent une certaine «sensibilité émotionnelle envers autrui<sup>7</sup>». Enfin, lorsque je référerai spécifiquement à la forme d'empathie basée sur ces réponses primitives, j'emploierai le terme «empathie primale<sup>8</sup>».

On pourrait objecter que la conception de l'empathie proposée ici est trop vague ou trop inclusive. Or, une définition adéquate est souvent l'aboutissement de la recherche plutôt que son point de départ<sup>9</sup>. S'il s'avérait que les mécanismes primitifs qui sont au cœur du modèle des poupées russes ne jouent pas le rôle qui leur est attribué par de Waal, la pertinence d'une définition inclusive pourrait être remise en cause. Autrement dit, il n'y aurait plus nécessairement d'utilité, d'un point de vue théorique, à regrouper ces phénomènes sous un même terme. Cela suggère que l'enjeu principal n'est pas de définir l'empathie a priori, mais plutôt d'évaluer le modèle à la lumière des données empiriques. Avant de procéder à son évaluation, il convient toutefois de le présenter plus en détail.

### *Le modèle des poupées russes*

De Waal reproche à certains théoriciens, des psychologues notamment, leurs tendances « saltationnistes ». Certaines facultés humaines telles que le langage sont décrites par ceux-ci comme de tous nouveaux modules, comme des sauts brusques dans l'évolution. La métaphore des poupées russes est censée rendre compte du fait que l'évolution est plutôt un processus graduel et, pourrait-on dire, économe. Plutôt que de jeter pour ensuite recommencer à neuf, elle étend ou modifie des structures déjà en place. Les nageoires avant des poissons, par exemple, sont devenues avec le temps des pattes, des ailes et des mains<sup>10</sup>. De manière générale, un récit plausible des origines d'un trait donné devrait, selon lui, avoir ce caractère progressif. Comme il le souligne, les biologistes préfèrent « les explications *bottom-up* aux explications *top-down*<sup>11</sup> ». C'est dans cette perspective que s'inscrit son modèle de l'empathie.

La forme d'empathie la plus simple, l'empathie primale, se manifeste notamment par la contagion émotionnelle. Elle représente la première couche de la poupée russe. Un bébé qui se met à pleurer parce qu'il entend un autre bébé pleurer constitue un bon exemple de contagion émotionnelle. Ce type de réaction serait généré par un mécanisme perception-action (*perception-action mechanism* ou PAM) permettant en quelque sorte à un sujet d'accéder « aux états subjectifs d'autrui (l'objet) via ses propres représentations neurales et

corporelles<sup>12</sup>». Les réponses générées par le PAM sont programmées génétiquement (*hard-wired*) et elles sont largement involontaires. Ce même mécanisme permettrait aussi d'expliquer l'imitation motrice ou « effet caméléon<sup>13</sup> ». Il s'agit de la tendance chez les humains et d'autres animaux à imiter inconsciemment les postures, expressions faciales et autres comportements de leurs congénères<sup>14</sup>. Le fait que les bébés humains commencent à imiter les expressions faciales des adultes à peine une heure après leur naissance illustre assez bien le caractère automatique et inconscient de l'imitation<sup>15</sup>. Il existe de nombreuses études explorant les relations entre l'imitation motrice et les formes d'empathie plus « cognitives ». Celles portant sur le bâillement — dont on sait qu'il est souvent « contagieux » — suggèrent l'existence d'un lien étroit entre ces deux phénomènes. J'y reviendrai plus en détail au moment d'évaluer si le modèle résiste à l'épreuve des faits.

Les deux autres formes d'empathie postulée par le modèle, soit le souci compatissant (*sympathetic concern*) et l'adoption empathique de la perspective d'autrui (*empathic perspective-taking*), sont celles qui relèvent de l'empathie cognitive. En effet, elles supposent toutes deux une certaine compréhension par le sujet de la situation dans laquelle se trouve l'objet, voire des états mentaux (croyances, désirs, buts, etc.) de celui-ci.

Le souci compatissant constitue la couche intermédiaire de la poupée russe. À ce stade, les réponses automatiques générées par le PAM s'accompagnent d'une évaluation plus ou moins sophistiquée de la situation d'autrui. Dans un cas de simple contagion émotionnelle, le sujet peut, à la limite, ne faire aucune distinction entre lui-même et autrui. Le bébé qui se met à pleurer parce qu'un autre bébé pleure près de lui ressent sans doute des émotions négatives. Cependant, il ne reconnaît pas qu'elles ont une source extérieure et il ne saurait prendre des mesures appropriées afin d'y remédier. Par contraste, un enfant plus âgé cherchera à consoler une personne triste par divers moyens, en l'enlaçant par exemple. Ce genre de comportement, qui se manifeste très tôt chez les humains, est aussi observé chez quelques primates non humains et chez certains singes<sup>16</sup>. Les comportements de consolation sont l'un des meilleurs indicateurs, selon de Waal, de

la présence de «souci compatissant<sup>17</sup>». Le fait de venir en aide à un congénère qui est victime d'une agression et le partage de nourriture sont d'autres exemples de comportements observés chez les primates non humains qui pourraient témoigner de la présence de cette forme d'empathie.

La forme d'empathie la plus complexe postulée par le modèle, qui représente la couche extérieure de la poupée, est l'adoption empathique de la perspective d'autrui (*empathic perspective-taking*). L'origine de la capacité à «se mettre à la place des autres» n'est pas forcément liée à l'empathie. On peut assez aisément imaginer des exemples de situations où l'on adopte le point de vue d'autrui sans ressentir quelque émotion que ce soit. Cela dit, il semble aussi y avoir des cas où les deux se combinent. C'est-à-dire que l'adoption de la perspective d'autrui est parfois accompagnée d'un degré plus ou moins fort d'engagement émotionnel. On parlera alors d'adoption empathique de la perspective d'autrui.

Les comportements dits d'aide ciblée (*targeted helping*) seraient le principal indicateur de la présence de cette forme d'empathie<sup>18</sup>. L'aide ciblée se définit comme l'aide et les soins «basés sur l'appréciation cognitive du besoin ou de la situation d'autrui<sup>19</sup>». Cette capacité serait présente chez de rares espèces, notamment les primates, les cétacés et les éléphants. On en trouve de nombreux exemples dans la littérature. Yamamoto et ses collègues ont mené une étude dans laquelle un chimpanzé devait venir en aide à un autre individu en lui fournissant, parmi un choix d'outils, celui qui convenait à la tâche qu'il devait accomplir. Le succès des chimpanzés dans cette tâche indique qu'ils sont en mesure de comprendre la situation, les objectifs (*goals*) poursuivis par leurs partenaires et d'ajuster leur comportement d'aide en fonction de cela<sup>20</sup>. Un exemple plus spectaculaire, bien qu'anecdotique, est celui de Binti-Jua, une femelle gorille ayant porté assistance à un enfant de trois ans tombé dans son enclos. Elle l'a pris dans ses bras et l'a consolé, a éloigné les autres gorilles qui cherchaient à s'approcher, avant de le remettre sain et sauf aux responsables du zoo<sup>21</sup>.

### *Deux objections*

Il n'existe aucun consensus quant à la validité du modèle des poupées russes. Cela n'a rien d'étonnant si l'on considère que l'on peine à s'entendre ne serait-ce que sur une définition adéquate de l'empathie. Il y a au moins deux objections importantes auxquelles j'estime qu'un défenseur du modèle devrait apporter une réponse.

On peut d'abord se demander s'il est possible de réunir tous les phénomènes plus ou moins disparates qui sont décrits par le modèle de manière à fournir une explication unifiée de l'empathie. Plus précisément, on peut se demander si l'empathie primale et l'empathie cognitive ne seraient pas des phénomènes entièrement distincts qu'il conviendrait d'étudier séparément.

On peut aussi douter que les explications de divers comportements altruistes qui sont proposées — en termes de *compréhension* de la situation, d'attribution d'états mentaux à autrui et ainsi de suite — soient appropriées. N'existe-t-il pas des explications moins « mentalistes<sup>22</sup> » de ces comportements ? Je tenterai de répondre tour à tour à ces objections dans ce qui suit.

### *Empathie primale et empathie cognitive : des phénomènes distincts ?*

Considérons d'abord la première objection. Tous n'admettent pas que les phénomènes qui sont évoqués par de Waal puissent être réunis de manière à fournir une explication unifiée de l'empathie. Batson, par exemple, affirme que la thèse selon laquelle la contagion émotionnelle ou l'imitation motrice<sup>23</sup> sont « la source unificatrice de tous les sentiments empathiques semble surestimer leur rôle, en particulier chez les humains<sup>24</sup> ». Cette objection doit être prise au sérieux. Notons cependant que la critique de Batson, prise littéralement, suggère un critère de preuve qu'aucune théorie sur l'empathie, à ma connaissance, ne saurait satisfaire. En effet, il est peu probable que l'on parvienne à démontrer que *tous* les sentiments empathiques ont une même source, quelle qu'elle soit. Un critère plus raisonnable me semble être le suivant : il faut montrer que les mécanismes qui sont à l'origine de l'empathie primale jouent un rôle inéliminable dans une explication adéquate de l'empathie cognitive, y compris chez les humains. Autrement dit, une théorie qui ne leur

accorderait aucune place aurait des lacunes importantes sur le plan explicatif. Peut-on satisfaire ce critère ?

J'ai évoqué précédemment les études portant sur le bâillement contagieux. Ces études suggèrent l'existence d'une relation étroite entre le mécanisme perception-action (associé à l'empathie primale) et l'empathie cognitive. Chez les humains, cela est notamment illustré par le fait que les enfants souffrant de troubles du spectre autistique, dont les difficultés sur le plan de la « théorie de l'esprit » et de l'empathie cognitive sont bien documentées<sup>25</sup>, démontrent aussi une faible réaction au bâillement d'autrui<sup>26</sup>. Selon de Waal, le peu de susceptibilité au bâillement contagieux, de même que les déficits sur le plan de l'empathie cognitive, pourraient découler d'une dysfonction du PAM<sup>27</sup>. Cette dysfonction pourrait aussi expliquer le fait que les enfants autistes aient des difficultés motrices importantes, un symptôme dont on a longtemps cru qu'il n'était que « périphérique<sup>28</sup> ».

Platak et ses collègues, dans un article datant de 2003<sup>29</sup>, proposent plutôt que le bâillement contagieux est une *conséquence* de la théorie de l'esprit. Si cette hypothèse était vérifiée, le fait que les enfants autistes y soient moins susceptibles que les enfants au développement typique n'offrirait aucun support au modèle défendu par de Waal. L'absence de bâillement, de ce point de vue, découlerait directement des déficits relatifs à la théorie de l'esprit. Au moment de la publication de l'article de Platak et ses collègues, le phénomène du bâillement contagieux n'avait été étudié que chez certains primates non humains et chez les humains, des espèces qui sont en mesure de réussir des tâches d'auto-reconnaissance et d'attribution d'états mentaux. Cela offrait un certain support à leur hypothèse. Ils concluaient sur cette supposition : seules les espèces qui sont en mesure d'accomplir de telles tâches peuvent faire preuve de bâillement contagieux<sup>30</sup>.

Des travaux plus récents permettent de douter que ce soit le cas. Des chercheurs se sont tournés depuis vers d'autres espèces et il semble que certaines d'entre elles, les chiens<sup>31</sup> et les loups<sup>32</sup> notamment, soient aussi sensibles au bâillement d'autrui. Or, on peut douter que les chiens ou les loups aient une « théorie de

l'esprit<sup>33</sup>». Si ces résultats sont exacts<sup>34</sup>, l'hypothèse avancée par de Waal pourrait bien être la plus plausible : une défaillance dans les mécanismes formant la couche intérieure de la poupée russe peut avoir des conséquences importantes sur les mécanismes des couches extérieures. Si l'empathie primale et l'empathie cognitive étaient des phénomènes complètement distincts, on voit mal pourquoi il en irait ainsi.

D'autres découvertes au sujet du bâillement contagieux parlent en faveur du modèle des poupées russes. Le degré de proximité sociale entre les loups permet de prédire le degré de contagion du bâillement : plus les individus sont intimement liés, plus le bâillement est contagieux<sup>35</sup>. Des résultats analogues ont été obtenus dans le cadre d'une étude portant sur des babouins gélada<sup>36</sup>. Campbell et de Waal, ont pour leur part évalué si le degré de contagion du bâillement variait en fonction de l'appartenance au groupe chez les chimpanzés. Ils ont découvert que les chimpanzés bâillent davantage lorsqu'ils voient bâiller des membres de leur propre groupe que des membres d'autres groupes<sup>37</sup>. Ces résultats suggèrent que le phénomène du bâillement contagieux comporte une importante dimension affective. Les auteurs de chacune de ces trois études notent d'ailleurs explicitement la compatibilité de leurs découvertes avec un modèle « multicouche » de l'empathie. Malheureusement, aucune étude n'a été menée à ce jour afin de déterminer si le degré de proximité sociale ou l'appartenance au groupe ont des effets similaires chez les humains. On sait cependant que les humains démontrent une réponse empathique plus grande à l'observation de douleur infligée à un membre de leur groupe qu'à un membre d'un autre groupe<sup>38</sup>. Il ne serait pas étonnant que l'on puisse observer un effet similaire en ce qui concerne le bâillement.

### *Le problème des explications « mentalistes »*

Considérons maintenant la seconde objection. L'existence des divers comportements d'aide décrits par de Waal n'est pas tellement sujette à controverse. Toutefois, l'interprétation que l'on peut en faire l'est davantage. Les comportements qui sont censés témoigner de la présence d'empathie cognitive (consolation, aide ciblée, etc.)

sont des comportements que l'on peut qualifier d'«altruistes» au sens évolutionnaire. C'est-à-dire qu'ils bénéficient à autrui tout en imposant un coût, plus ou moins important selon le cas, à celui qui agit. De Waal semble cependant franchir un pas supplémentaire en présumant que ces comportements témoignent aussi d'altruisme au sens psychologique. Un comportement est dit altruiste en ce sens s'il est motivé par un désir pour le bien-être d'autrui<sup>39</sup>. Comment s'assurer que c'est ce genre de motivations qui expliquent l'action? Ne pourrait-on pas rendre compte de ces comportements en ne postulant que des motifs égoïstes? De plus, ne serait-il pas possible de les expliquer par des mécanismes moins exigeants cognitivement que l'adoption de la perspective d'autrui, le simple conditionnement par exemple?

Joan B. Silk, commentant le cas du gorille Binti-Jua, dont il a été question précédemment, note la difficulté qu'il y a à tirer des conclusions à partir d'exemples anecdotiques. Preston et de Waal n'hésitent pas à évoquer l'empathie cognitive, un souci pour le bien-être du petit garçon, afin d'expliquer les gestes posés par Binti-Jua<sup>40</sup>. Or, d'autres interprétations sont possibles, voire préférables. Il s'avère que Binti-Jua a été élevée par des humains, sa mère l'ayant rejetée très tôt. Soucieux qu'elle devienne elle-même une mère négligente éventuellement, le personnel du zoo a tenté de l'entraîner à dispenser des soins maternels. L'une des tâches qui lui étaient assignées consistait à retrouver une poupée et à la rapporter aux entraîneurs. La similarité entre la tâche en question et les comportements observés lors du sauvetage du garçon est évidente. Il n'est pas impossible, par conséquent, que les effets de l'entraînement, plutôt qu'un souci pour le bien-être de l'enfant, rendent compte des gestes posés par Binti-Jua<sup>41</sup>.

J'estime qu'il y a là une leçon importante à retenir, soit qu'il faut être prudent avant de tirer des conclusions. Plusieurs explications, des plus strictement «béhavioristes» aux plus «cognitivement sophistiquées», peuvent potentiellement expliquer un comportement donné. Il s'agit d'un problème très général que l'on rencontre dans les débats concernant la pertinence d'attribuer toutes sortes d'états mentaux aux animaux, qu'il s'agisse des concepts, de la

métacognition, de la conscience phénoménale, ou dans le cas qui nous occupe, de l'empathie. Par ailleurs, une approche trop anecdotique, je l'ai mentionné dès l'introduction, est insuffisante.

Cela dit, on peut admettre que le cas de Binti-Jua ne soit pas des plus concluants tout en maintenant que les « anecdotes » pertinentes, prises dans leur ensemble, suggèrent la présence d'empathie cognitive chez les primates non humains. Autrement dit, il est loin d'être évident que *tous* les incidents rapportés se prêtent à une explication alternative aussi convaincante que dans le cas particulier de Binti-Jua. O'Connell, par exemple, a procédé à une méta-analyse de plus de deux mille anecdotes recensées chez des chimpanzés et conclut que certains cas, en particulier des cas d'assistance à autrui dans une situation dangereuse, suggèrent la présence d'empathie cognitive<sup>42</sup>.

Silk maintient que les compilations d'anecdotes sont elles aussi sujettes à divers biais. En particulier, elle souligne que les observateurs sont davantage susceptibles de remarquer et de se rappeler d'incidents où les animaux semblent faire preuve d'empathie que des situations où ils sont indifférents aux malheurs de leurs congénères<sup>43</sup>. Cet argument se résume, ultimement, à noter que les comportements suggérant la présence d'empathie cognitive sont plutôt rares — à tout le moins plus rares que l'indifférence — chez les animaux non humains. En quoi s'agit-il d'un problème pour un défenseur du modèle des poupées russes ?

Imaginons un observateur du comportement humain. Nul doute qu'il observerait quantité de situations où les humains font preuve d'indifférence envers les besoins d'autrui. Les comportements empathiques pourraient même être l'exception et l'indifférence la norme. Dans une expérience bien connue, Darley et Batson ont montré que des éléments contextuels assez peu importants, tels que le fait d'être en retard à un rendez-vous, peuvent conduire des sujets humains (même des séminaristes à qui l'on vient tout juste de lire la parabole du bon Samaritain !) à ignorer la détresse d'une autre personne. Pas moins de 90 % des sujets qui étaient dans la condition « très pressé » prévue par le test n'ont apporté aucune forme d'aide à une personne simulant la détresse sur leur chemin<sup>44</sup>. Supposons que

notre observateur ne dispose d'aucune autre source d'information sur le comportement humain. Que devrait-il conclure ? Que les quelques rares sujets qui se sont arrêtés ne sont que des cas anecdotiques, des anomalies statistiques peu représentatives des « capacités empathiques » de l'espèce humaine ? J'estime que la rareté des comportements en question n'offre aucune raison de ne pas les considérer sérieusement. Par ailleurs, le fait que l'on parvienne, avec le temps, à cumuler des centaines, voire des milliers d'anecdotes pertinentes n'indique-t-il pas qu'il y a bel et bien un phénomène sous-jacent qui réclame une explication ?

La provenance de toutes ces anecdotes est elle aussi significative. Comme le remarque de Waal, les scientifiques qui travaillent avec des dauphins, des éléphants et des primates non humains ont quantité d'histoires d'aide ciblée à raconter, tandis que les scientifiques qui travaillent avec d'autres animaux n'en ont pas. Est-il plausible qu'il s'agisse d'une simple coïncidence<sup>45</sup> ? Il y a fort à parier que les capacités cognitives particulières de ces animaux y soient pour quelque chose, ce qui parle en faveur des explications plus « mentalistes ».

Il est important de noter aussi que les cas recensés d'aide ciblée ne proviennent pas *exclusivement* d'anecdotes. Les travaux de Yamamoto et ses collègues, dont il a été question précédemment, ont été menés dans un environnement contrôlé. Warnaken et ses collègues ont aussi mené récemment une série d'études où des chimpanzés sont venus en aide à d'autres chimpanzés et à des humains sans qu'il y ait possibilité pour eux d'obtenir une récompense<sup>46</sup>. Il est possible de répliquer les conditions de ces tests et d'isoler diverses variables afin d'exclure d'autres explications possibles (le conditionnement par exemple), ce qui n'est évidemment pas possible pour les cas anecdotiques. En contrepartie, un avantage important des anecdotes est qu'elles offrent souvent des exemples de comportements empathiques « extrêmes » (p.ex. : risquer sa vie pour sauver autrui) qu'il serait difficile, et sans doute contraire à l'éthique, de reproduire en laboratoire<sup>47</sup>. On peut en conclure que les deux sources de données sont complémentaires. Dans leur ensemble, les observations provenant de ces deux sources sont certainement

compatibles avec les explications « mentalistes » des comportements d'aide. Il ne me semble pas déraisonnable de croire en plus, à la lumière de ce qui précède, que ce sont aussi les explications les plus plausibles.

Évidemment, rien de ce qui précède ne permet de conclure hors de tout doute que les gestes posés par les animaux émanent d'un véritable souci pour le bien-être d'autrui. Cela dit, il ne faudrait pas exiger un degré de preuve qu'il est impossible de satisfaire. Même chez les humains, qui sont en mesure de rapporter verbalement les motifs qui les ont incités à poser tel ou tel geste, il demeure des sceptiques quant à la possibilité de *l'altruisme psychologique*. *L'égoïsme psychologique*, la thèse selon laquelle le bien-être des autres ne nous intéresse que dans la mesure où notre propre bien-être en dépend, a encore des défenseurs. Même le comportement d'un soldat qui se couche sur une grenade afin d'éviter la mort à ses camarades, un exemple classique dans la littérature sur le sujet, admet une explication en terme de motifs égoïstes. Il existe de nombreux arguments de part et d'autre — invoquant la parcimonie, le sens commun, etc. — dont aucun à ce jour n'est décisif<sup>48</sup>.

Je n'entrerai pas dans les détails de ces débats, cela exigerait de trop longs développements. Je me contenterai de noter mon accord avec la position défendue par Sober et Wilson, soit qu'il est tout à fait possible que l'évolution ait sélectionné des mécanismes proximaux basés sur le souci envers autrui afin de favoriser des comportements qui sont avantageux sur le plan de la survie et de la reproduction. Les soins parentaux offrent un bon exemple. Si l'« objectif » poursuivi est la survie des petits, la sélection d'un mécanisme qui incite les parents à se préoccuper de leur progéniture semble être la manière la plus directe de susciter un comportement adéquat<sup>49</sup>. Pourquoi, demande Sober, la sélection naturelle aurait-elle incité les organismes à se soucier uniquement de leur propre bien-être, ce qui ne concerne qu'indirectement leur progéniture, plutôt que de les inciter à « garder un œil sur l'objectif » (*set their eyes on the prize*)<sup>50</sup>? Les comportements qui relèvent de l'empathie cognitive admettent, je crois, une explication similaire. Si l'on suppose qu'ils favorisent en quelque manière la survie et la reproduction des animaux qui en sont

capables, la manière la plus directe de les susciter est probablement un véritable souci pour le bien-être d'autrui.

### *Conclusion*

Il faudrait en dire bien davantage afin de valider le modèle des poupées russes. Plusieurs études dont il a été question s'inscrivent dans de vastes chantiers qui sont toujours en cours. C'est le cas notamment des études visant à déterminer quelles espèces sont sujettes au bâillement contagieux et des études sur les causes des troubles du spectre autistique notamment. Il n'est pas impossible que des résultats contredisant ce qui a été présenté ici émergent éventuellement. En ce sens, l'avenir du modèle est incertain. Cela dit, j'espère avoir montré les limites de deux objections importantes qui sont souvent invoquées à l'encontre de celui-ci et, plus généralement, des conceptions «bottom-up» de l'empathie. L'empathie primale et l'empathie cognitive sont suffisamment «imbriquées» pour qu'il soit justifié de les inclure, au moins provisoirement, au sein de notre définition de l'empathie. En ce qui concerne l'empathie cognitive, il y a de bonnes raisons de croire que certains animaux non humains en sont capables. Le fait que cela exige des capacités cognitives complexes et que l'on doive parfois s'en remettre à des anecdotes n'est pas un motif suffisant pour en douter.

- 
1. «The fact that lower animals are excited by the same emotions as ourselves is so well established, that it will not be necessary to weary the reader by many details.» Charles Darwin, *The Descent of man in relation to sex*, Princeton, Princeton University Press, 1981, p. 39.
  2. Voir par exemple: Collin Allen et Marc Bekoff, *Species of mind: the philosophy and biology of cognitive ethology*, Cambridge, The MIT Press, 1999, p. xv; Frans de Waal, «Putting altruism back into altruism: the evolution of empathy», *Annual Review of Psychology*, n° 59, 2008 : p. 282.
  3. Notons que l'empathie ne fait pas partie des émotions que Darwin attribuait aux animaux, le terme n'étant pas encore en usage à son époque. Il ne fait son apparition qu'au début du 20<sup>e</sup> siècle. Voir Lauren Wispé, «History of the Concept of Empathy» dans Nancy Eisenberg

- et Janet Strayer (éd.), *Empathy and Its Development*, Cambridge, Cambridge University Press, 1987, pp. 17-37.
4. C. Daniel Batson, «These things called empathy» dans Jean Decety et William Ickes (éd.), *The social neuroscience of empathy*, Cambridge, The MIT Press, 2009, p. 8.
  5. *Ibid.*
  6. Frans de Waal, *loc. cit.*, p. 282.
  7. Frans de Waal, *loc. cit.*, p. 282.
  8. Jaak Panksepp et Jules B. Panksepp, «Toward a cross species understanding of empathy», *Trends in neurosciences*, vol.36, n°8, 2013, p. 490.
  9. Allen et Bekoff avancent le même argument pour défendre la définition imprécise qu'ils proposent du jeu social (*social play*) et de la conscience. Voir Allen et Bekoff, *op. cit.*, pp.90-91 et pp. 143-144.
  10. Frans de Waal, «The Russian doll model of empathy» dans Stein Braten (éd.), *On being moved: from neuron mirrors to empathy*, Amsterdam, John Benjamin's Publishing Company, 2009, p. 49.
  11. *Ibid.*
  12. Frans de Waal, «Putting altruism back into altruism», *loc. cit.*, p. 286.
  13. Tanya L. Chartrand et John A. Bargh, «The chameleon effect: the perception-behavior link and social interaction», *Journal of Personality and Social Psychology*, vol. 76, n° 6, 1999, p. 893.
  14. *Ibid.*
  15. Marc D. Hauser, *Moral minds : how nature designed our universal sense of right and wrong*, New York, Harper Collins, 2006, p. 28.
  16. Elisabetta Palagi et al., «Exploring the evolutionary foundations of empathy: consolation in monkeys», *Evolution and human behavior*, n° 35, 2014, p. 341.
  17. Frans de Waal, «Putting altruism back into altruism», *loc. cit.*, p. 284.
  18. Frans de Waal, «Putting altruism back into altruism», *loc. cit.*, p. 286.
  19. Shinya Yamamoto et al., «Chimpanzees' flexible targeted helping based on an understanding of conspecifics' goals», *Proceedings of the National Academy of Sciences*, vol. 109, n°9, 2012, p. 3588.
  20. *Ibid.*
  21. Joan B. Silk, «Social preferences in primates» dans Paul W. Glimcher et al. (éd.), *Neuroeconomics : decision making and the brain*, London, Academic Press, 2008, p. 276.; Stephanie D. Preston et Frans de Waal, «Empathy: its ultimate and proximate basis», *Behavioral and Brain Sciences*, n° 25, 2002, p. 19.

22. Elske van der Vaart et Charlotte K. Hemelrijk, «Theory of mind in animals: ways to make progress», *Synthèse*, vol. 191, n° 3, 2014, p. 336.
23. Ce que j'ai appelé aussi l'«effet caméléon».
24. Daniel Batson, *op. cit.*, p. 5.
25. R. J. R. Blair, «Responding to the emotions of others: dissociating forms of empathy through the study of typical and psychiatric populations», *Consciousness and cognition*, n° 14, 2005, p. 706. Dans cet article, Blair traite «théorie de l'esprit» et «empathie cognitive» comme des synonymes. J'ai suggéré, suivant de Waal, que le fait que l'on puisse adopter la perspective d'autrui sans engagement émotionnel est une raison de les distinguer. Quoi qu'il en soit, rien de décisif pour l'argument de la présente section n'en dépend. Qu'il s'agisse d'une même capacité ou de deux capacités différentes, les enfants souffrant d'autisme ont un déficit important.
26. Atushi Senju et al., «Absence of contagious yawning in children with autism spectrum disorder», *Biology Letters*, vol. 3, n° 6, 2007, p. 706.
27. Frans de Waal et Evan Thompson, «Primates, monks and the mind», *Journal of Consciousness Studies*, vol. 12, n° 7, 2005, p. 40.
28. Vittorio Gallese et al., «The mirror mechanism and its potential role in autism spectrum disorder», *Developmental Medicine & Child Neurology*, vol. 55, n° 1, p. 17.
29. Steven M. Platek et al., «Contagious yawning: the role of self-awareness and mental state attribution», *Cognitive Brain Research*, n° 17, 2003, pp. 223-227.
30. Steven M. Platek et al., *loc. cit.*, p. 226.
31. Ramiro M. Joly-Mascheroni et al., «Dogs catch human yawn», *Biology Letters*, vol. 4, n° 5, 2008, pp. 446-448.
32. Teresa Romero et al., «Social modulation of contagious yawning in wolves», *Plos One*, vol. 9, n° 8, 2014, pp. 1-6.
33. Cela dit, certains suggèrent que la théorie de l'esprit n'est pas une capacité unitaire, qu'elle admet des degrés. Les chiens et les loups démontrent peut-être, selon cette conception, une «théorie de l'esprit rudimentaire». Voir Alexandra Horowitz, «Theory of mind in dogs?: examining method and concepts», *Learning & Behavior*, vol. 14, n° 2, 2011, p. 316.
34. O'Hara et Reeve font état de la difficulté à répliquer les résultats chez les chiens. Voir Sean O'Hara et Amy V. Reeve, «A test of the yawning contagion and emotional connectedness hypothesis in dogs, *Canis familiaris*», *Animal Behaviour*, n° 81, 2011, p. 335-340.
35. Teresa Romero et al., *loc. cit.*, p. 6.

36. Elisabetta Palagi et al., «Contagious yawning in gelada baboons, as a possible expression of empathy», *Proceedings of the National Academy of Sciences*, vol. 106, n° 46, 2009, pp. 19262-19267.
37. Matthew W. Campbell et Frans de Waal, «Ingroup-outgroup bias in contagious yawning by chimpanzees supports link to empathy», *PLoS One*, vol. 6, n° 4, 2011.
38. *Ibid.*, p. 3.
39. Voir Stephen Stich, «Evolution, altruism and cognitive architecture : a critique of Sober and Wilson's argument for psychological altruism», *Philosophy and biology*, n° 22, 2007, p. 267-268.
40. Stephanie D. Preston et Frans de Waal, *loc. cit.*, p. 19.
41. Joan B. Silk, *op. cit.*, p. 276.
42. Stephanie D. Preston et Frans de Waal, *loc. cit.*, p. 19 ; S. M. O'Connell, «Empathy in chimpanzees: evidence for theory of mind?», *Primates*, vol. 36, n° 3, 1995, pp. 397-410. La terminologie employée est cependant différente de celle que j'ai adoptée. O'Connell distingue quatre types d'empathie : visuelle, émotionnelle, de concordance et étendue.
43. Joan B. Silk, *op. cit.*, p. 276.
44. John M. Doris, *Lack of character*, Cambridge, Cambridge University Press, 2002, p. 34 ; John M. Darley et Daniel Batson, «From Jerusalem to Jericho: a study of dispositional variables in helping behaviour», *Journal of Personality and Social Psychology*, vol. 27, n° 1, 1973, pp. 100-108.
45. Frans de Waal, «The Russian doll model of empathy and imitation», *op. cit.*, p. 56.
46. Dans Joshua Greene, *Moral tribes: emotions, reason and the gap between us and them*, New York, The Penguin Press, 2013, pp. 38-39.
47. Frans de Waal, «Putting altruism back into altruism», *loc. cit.*, p. 290.
48. Elliott Sober, «Psychological egoism», H. LaFollette et I. Persson (éd.), *The Blackwell Guide to Ethical Theory*, Malden, Blackwell Publishing, 2013, chap. 7.
49. *Ibid.* ; Elliott Sober et David Sloan Wilson, *Unto others: the evolution and psychology of unselfish behavior*, Cambridge, Harvard University Press, 1998, partie 2.
50. Elliott Sober, *op. cit.*, p. 147. Pour une critique des arguments de Sober et Wilson, voir Stephen Stich, *loc. cit.*





# *Commentaires*



# Contexte de découverte et contexte de justification : une aporie dans l'empirisme logique ?

PIER-ALEXANDRE TARDIF, *Université Laval*

## *1. Introduction*

Dans un article publié en 1969, *Epistemology Naturalized*, Willard Van Orman Quine aborde le problème épistémologique du fondement de la science tel qu'il s'est posé dans la première partie du XX<sup>e</sup> siècle pour l'empirisme logique. Le problème du fondement de la science concerne les mathématiques et ce que Quine appelle la connaissance naturelle ou empirique. Du côté des mathématiques, la fondation consiste à clarifier les lois et concepts mathématiques en les explicitant dans les termes de la logique des prédicats du premier ordre. Les vérités mathématiques seraient ainsi réduites à des vérités logiques évidentes, ou à tout le moins potentiellement évidentes moyennant leur dérivation à partir de vérités et d'étapes évidentes. Du côté de la connaissance empirique, la fondation consiste cette fois à fonder les connaissances et vérités naturelles sur l'expérience sensible en définissant, par exemple, la notion de «corps<sup>1</sup>» en termes sensoriels et en justifiant les connaissances empiriques également en termes sensoriels. Quine fait cependant remarquer que, à la suite d'échecs répétés, l'abandon du projet fondationnaliste, qui consiste à fonder la science sur quelque chose de plus assuré qu'elle, a mené au «délogement de l'épistémologie hors de son vieux statut de philosophie première [qui] a entraîné une vague de scepticisme épistémologique<sup>2,3</sup>». Non certes sans une pointe d'ironie, Quine précise que, après Carnap qui qualifiait péjorativement la métaphysique comme étant sans signification, Wittgenstein et ses disciples à Oxford ont trouvé une nouvelle vocation à la philosophie sous la forme d'une thérapie : guérir les philosophes de l'illusion qu'il y a une telle chose que des problèmes

épistémologiques<sup>4</sup>. L'objectif du présent article ne consiste ni à suivre Carnap ou Wittgenstein dans leurs dénonciations ni à faire l'apologie de l'épistémologie fondationnaliste, voire naturaliste, mais à reconstruire la signification de l'expression «scepticisme épistémologique» qui se trouve dans l'article de Quine. Pour ce faire, nous utiliserons, à la façon d'un algorithme, la distinction entre «contexte de découverte» et «contexte de justification» — ci-après dénommée «distinction DJ» — introduite par l'un des fondateurs de l'empirisme logique, Hans Reichenbach. Nous exposerons d'abord deux formes de cette distinction — celle de Reichenbach puis celle de Popper — en faisant ressortir quelques implications et difficultés les caractérisant. Nous brosserons ensuite un portrait schématique de l'empirisme logique en nous référant à la reconstitution proposée par Radnitzky. En conclusion, nous montrerons comment l'application de la distinction DJ à l'empirisme logique permet de faire ressortir l'impasse dans laquelle se trouve ce dernier, ce qui clarifiera en retour la signification de l'expression «scepticisme épistémologique» telle qu'employée par Quine.

## *2. Contexte de découverte et contexte de justification*

### *2.1. Hans Reichenbach et Wesley Salmon*

Dans *Experience and Prediction*<sup>5</sup>, Hans Reichenbach introduit une première distinction entre «relations internes» et «relations externes» afin de distinguer les champs d'investigation respectifs de l'épistémologie et de la sociologie. Les relations internes concernent le contenu de la connaissance, tandis que les relations externes combinent la connaissance avec des énonciations d'un autre type qui ne concerne pas ce contenu. Par exemple, un sociologue pourra rapporter non seulement que les astronomes construisent des télescopes pour observer les étoiles, mais également que les astronomes sont fréquemment des musiciens ou qu'ils appartiennent généralement à la classe bourgeoise de la société. Alors que la sociologie s'intéresse à la connaissance avant tout comme un fait social, l'épistémologie se concentre sur la structure logique interne (les relations internes<sup>6</sup>) des connaissances scientifiques. Cette première distinction n'est cependant pas suffisante, selon

Reichenbach, pour distinguer l'épistémologie de la psychologie. Contrairement à la psychologie, qui a pour tâche de décrire les processus réels de la pensée, l'épistémologie entend reconstruire les processus de pensées tels qu'ils devraient faire occurrence dans un système consistant, c'est-à-dire dénué de contradictions. En ce sens, l'épistémologie substitue aux processus réels de la pensée des reconstructions logiques formulées à l'aide d'un langage artificiel idéal (par exemple la logique des prédicats du premier ordre), ce que l'on désigne sous l'expression «reconstruction rationnelle», c'est-à-dire la reconstitution sous forme déductive de la structure logique de l'ensemble des énoncés d'une théorie scientifique. Reichenbach est ainsi amené à distinguer entre la façon dont un mathématicien découvre un théorème (le contexte de découverte) et la façon dont ce mathématicien présente la démonstration de ce théorème (le contexte de justification).

Dans son ouvrage *Logic*, Wesley Salmon apporte quelques précisions concernant la distinction DJ. Salmon fait référence à l'histoire selon laquelle Sir Isaac Newton aurait eu une intuition brillante au moment où, étant assis dans un jardin près d'un arbre, il a vu une pomme tombée sur le sol : les planètes dans leurs courses, les objets tombant sur terre, les marées, tout, bref, est gouverné par la loi universelle de la gravitation. Il s'agit d'une charmante anecdote concernant la découverte d'une théorie, mais elle n'a rien à voir avec sa justification : la question de la justification, précise Salmon, ne peut être considérée qu'en termes d'observations, d'expérimentations et d'arguments — «en deux mots, la justification de la théorie dépend de l'évidence qui la supporte et non des facteurs psychologiques qui ont donné à Newton l'idée de cette théorie<sup>7</sup>». Nous avons vu que Reichenbach propose la distinction DJ pour marquer une différence entre les champs d'investigations de la psychologie et de l'épistémologie ; Salmon précise que cette distinction est étroitement reliée à celle entre une inférence et un argument. L'activité psychologique de faire une inférence est un processus de découverte, mais la personne doit ensuite découvrir la conclusion — parfois, c'est même la conclusion qui est découverte en premier et l'on doit se mettre à la recherche d'évidences qui la

supportent. Pour bien marquer la différence, Salmon fait référence au perspicace détective Sherlock Holmes qui voudrait bien enseigner à Watson ses méthodes d'investigations, mais qui doit bien se rendre à l'évidence qu'il n'y a pas de règles qu'il peut communiquer à Watson afin que ce dernier puisse parvenir à réaliser ses prouesses. Les habiletés de Holmes font intervenir des facteurs tels que sa curiosité affûtée, son intelligence, son imagination fertile, la richesse de sa culture générale, etc. La difficulté est qu'il n'y a pas un ensemble de règles qui peut se substituer à ces habiletés. Si tel était le cas, ces règles constitueraient les règles de la découverte. Salmon de conclure que les gens peuvent certes améliorer leur raisonnement par l'éducation et la pratique, mais que ceci n'a rien à voir avec le fait d'adopter un ensemble de règles logiques. Ce que la logique nous apprend n'est pas à faire naturellement des inférences, mais à déterminer lesquelles nous pouvons légitimement accepter. Étant donné que le contexte de découverte fait intervenir l'imagination et des circonstances fortuites dans le processus réel de la pensée, l'imposition à ce processus réel de règles logiques constituerait plutôt selon Salmon une camisole de force pouvant inhiber la créativité du chercheur. Sans emprise conséquente ou positive sur le contexte de découverte, la logique (ou l'épistémologie) concerne alors uniquement le contexte de justification. Puisque la justification d'un énoncé est un argument, la justification comprend deux aspects indépendants l'un de l'autre : la vérité de ses prémisses et la rectitude logique de l'argument (sa validité). En montrant que les prémisses sont fausses ou douteuses, de même qu'en montrant que l'argument est logiquement invalide, nous montrons certes que la justification est insatisfaisante, mais cela ne suffit pas à montrer que la conclusion est fausse. Par ailleurs, Salmon souligne que de considérer que les éléments du contexte de découverte — par exemple les particularités historiques, nationales, religieuses ou autres entrant dans la genèse d'une théorie — sont nécessairement pertinents à la question de la valeur de vérité est une erreur. À titre d'exemple de cette erreur, à savoir le sophisme génétique (*genetic fallacy*), Salmon rappelle la condamnation par les nazis de la théorie de la relativité d'Einstein

sous prétexte que celui-ci était juif. Attardons-nous maintenant à la seconde forme de la distinction.

## *2.2. Karl Popper*

Sir Karl Popper introduit la distinction DJ dans son ouvrage *La logique de la découverte scientifique*<sup>8</sup>. Il fait remonter cette distinction à celle posée par Kant entre les questions de fait (*quid facti?*) et les questions de justification ou de validité (*quid juris?*). D'aucuns ont souligné à juste titre que « [p]our Popper, il ne faut pas confondre le processus d'invention d'un énoncé et son examen logique<sup>9</sup> ». Alors que la psychologie traite des faits empiriques, la logique de la connaissance traite uniquement des relations logiques entre les énoncés. Popper distingue ainsi le processus de conception d'une nouvelle idée en spécifiant que la question de « savoir comment cette idée peut naître dans l'esprit d'un homme — qu'il s'agisse d'un thème musical, d'un conflit dramatique ou d'une théorie scientifique — peut être d'un grand intérêt pour la psychologie empirique, mais cette question ne relève pas de l'analyse logique de la connaissance scientifique<sup>10</sup> ». Comme pour Reichenbach et Salmon, l'acte de concevoir ou d'inventer ne peut faire l'objet d'une analyse logique selon Popper. L'analyse méthodologique que celui-ci préconise est une sorte de reconstruction rationnelle qui consiste à faire ressortir la charpente (forme) logique de la procédure de testabilité (*testing*). Dans ces conditions, malgré une certaine coïncidence du contexte de découverte entre Reichenbach et Popper, nous devons néanmoins conclure que leurs distinctions DJ sont différentes. Car même si Popper discute de justification en parlant du contexte scientifique propre à l'analyse épistémologique, il considère le vérificationnisme comme une erreur et le processus d'évaluation des théories scientifiques pour lequel il opte diffère de celui de Reichenbach. On ne peut donc pas réduire l'une à l'autre les deux distinctions entre « contexte de découverte » et « contexte de justification<sup>11</sup> ». Alors que Reichenbach préconise l'induction et une théorie vérificationniste selon laquelle les énoncés universels sont vérifiés par l'expérience sensible (c'est-à-dire par des énoncés singuliers), Popper opte pour la déduction et la théorie falsificationniste. Selon la méthodologie

falsificationniste que Popper expose dans *La logique de la découverte scientifique*, une hypothèse théorique sera dite testable et scientifique si elle est falsifiable en principe — sinon, il s'agit d'une hypothèse métaphysique. Une théorie est falsifiable en principe si la classe de ses falsificateurs potentiels (ou logiques, par opposition à actuels ou réels) n'est pas vide. Un falsificateur potentiel est un énoncé de base décrivant un événement dont la possibilité est logiquement interdite par la théorie. Si cet événement fait occurrence, alors la théorie sera contredite. La différence méthodologique cruciale entre les conceptions de Reichenbach et de Popper peut ainsi être explicitée à l'aide de leurs différents critères de démarcation. Comme le précise Popper, le critère de démarcation de la logique inductive consiste à dire que «l'on doit pouvoir décider de manière définitive de la vérité et de la fausseté de tous les énoncés de la science empirique (ou encore tous les énoncés "pourvus de sens")<sup>12</sup>». Or, bien qu'un énoncé strictement universel<sup>13</sup> soit falsifiable en principe, puisqu'il suffit de trouver un seul cas particulier le contredisant pour le rendre faux, il n'est pas vérifiable (une variante du problème de l'induction soulevé par Hume). C'est pourquoi Popper adopte la falsifiabilité plutôt que la vérifiabilité d'un système comme critère de démarcation. Étant donné que cette différence entre les contextes de justification de Popper et de Reichenbach touche à un élément crucial du programme de recherche de l'empirisme logique, il importe de s'attarder succinctement à leurs contextes de justification avant de présenter schématiquement ce dernier.

### 2.3. Critique de l'induction et de la théorie vérificationniste

Dans un ouvrage postérieur à *Experience and Prediction* intitulé *The Rise of Scientific Philosophy*, Hans Reichenbach précise que l'acte de découvrir échappe à l'analyse logique, puisqu'il n'y a aucun ensemble de règles à partir desquelles une machine à découvertes pourrait être construite et se substituer à la fonction créatrice du génie. Il poursuit en notant que la philosophie scientifique s'intéresse uniquement au contexte de justification et que «la justification en termes de données observables est le sujet de la théorie de l'induction<sup>14</sup>». À cet égard, le principe d'induction, censé justifier

les inférences inductives, détermine selon Reichenbach la vérité des théories scientifiques. Soulignons que ce principe est fondamental à la fois pour la science et pour l'épistémologie de l'empirisme logique. Éliminer le principe d'induction de la science « ne signifierait rien de moins que priver celle-ci de son pouvoir de décider de la vérité et de la fausseté de ses théories. Il est clair que sans lui la science ne garderait plus longtemps le droit de distinguer ses théories des créations fantasques et arbitraires de l'esprit du poète<sup>15</sup> ». Dans la mesure où l'empirisme logique conçoit l'épistémologie comme une science ayant notamment pour tâche de justifier les connaissances scientifiques, le principe d'induction s'avère être un élément de première importance. Dans le premier chapitre de *La logique de la découverte scientifique*, Popper mène précisément une critique à l'endroit de la logique inductive de Reichenbach et du vérificationnisme défendu par les membres du néopositivisme auxquels il fait référence, notamment Carnap, Schlick, Wittgenstein et Waismann. De l'avis de Popper, du point de vue du problème de la démarcation entre science et non-science, la logique inductive et le vérificationnisme coïncident et constituent une erreur.

Une inférence inductive consiste à passer des énoncés singuliers (comme des comptes rendus d'observations ou d'expériences) à des énoncés universels (des hypothèses et des théories). La difficulté de savoir si ces inférences inductives — c'est-à-dire l'inférence d'énoncés universels à partir d'énoncés singuliers — sont justifiées ou à quelles conditions elles le sont est ce que l'on a appelé, à la suite de Hume, le problème de l'induction : à partir du comportement des choses dans le passé, nous inférons qu'il en sera de même dans le futur en vertu du principe de l'uniformité de la nature. Hume défend la thèse selon laquelle ce principe et cette inférence, laquelle contient plus d'information dans sa conclusion qu'il y en a dans ses prémisses, ne sauraient se justifier rationnellement ou logiquement. Dans la perspective qu'il adopte en effet, il n'y a que deux façons dont nous pourrions justifier un tel principe : soit i) comme une relation entre idées, soit ii) par une réduction à des faits. D'un côté (i), la voie de la démonstration déductive *a priori* est interdite, puisque prétendre que le futur ne sera pas conforme au passé n'implique pas

de contradiction. De l'autre (ii), on ne saurait justifier ce principe par un recours aux faits sous peine de circularité ; car, pour ce faire, il faudrait avoir recours à l'induction et donc présupposer le principe que nous voulons justifier. Le seul argument supplémentaire qu'ajoute Popper dans sa critique du principe d'induction entériné par Reichenbach est que, pour justifier ce principe, nous devrions supposer un principe d'induction d'un ordre supérieur et ainsi de suite — ce qui conduit à une régression à l'infini<sup>16</sup>.

Comme le rapporte Salmon<sup>17</sup>, Reichenbach croyait avoir résolu le problème de l'induction posé par Hume. La clef de cette réussite selon Salmon est la reconnaissance, par Reichenbach, que ce qui est demandé n'est pas de prouver la proposition factuelle de l'uniformité de la nature comme le suppose Hume, mais de justifier une règle. À la suite de Herbert Feigl, Salmon nous invite à distinguer entre deux sortes de justification (*justification*) : d'une part la validation (*validation*), qui consiste à dériver une règle ou une proposition à partir d'un principe plus fondamental ; d'autre part, la défense (*vindication*), qui consiste à montrer que l'usage d'une règle ou d'une proposition est bien adapté à nos objectifs. Salmon note ensuite que la justification que Reichenbach propose de l'induction n'est pas une validation, mais bien une défense. Brièvement, l'argument de Reichenbach consiste à dire que si nous sommes chanceux et que la nature se trouve être uniforme, alors l'usage de l'induction répondra à nos attentes. Comme le souligne Salmon, la structure de l'argument de Reichenbach est similaire au pari de Pascal : il ne consiste pas à justifier la croyance en une proposition, mais plutôt à justifier ou à démontrer l'utilité d'une pratique (la méthode inductive), ce pour quoi Reichenbach caractérise sa justification de pragmatiste.

Manifestement, ces justifications pragmatistes ne sont pas suffisantes aux yeux de Popper, pour qui le principe d'induction est à la fois superflu et à la base d'incohérences logiques. Sa critique est de nature logique et consiste à souligner que l'inférence inductive n'est pas valide : la conclusion d'un argument inductif, qui porte sur des cas passés, présents et futurs, contient toujours plus d'informations (pouvant modifier sa valeur de vérité) que les prémisses, qui ne portent que sur les cas passés et présents. Un problème similaire

se pose dans le cas du vérificationnisme, qui suppose que la vérité et la fausseté d'un énoncé dépendent de sa méthode de vérification. Si l'énoncé que l'on confronte à l'expérience est vérifié, alors il est vrai ; s'il est réfuté, alors il est faux. La difficulté est que la forme de raisonnement implicite derrière la théorie vérificationniste fait appel à une conception pour le moins difficilement défendable. Le raisonnement consiste à affirmer le conséquent d'une implication pour en déduire l'antécédent : si P alors Q, Q, donc P. On se retrouve avec une forme de raisonnement qui est à la croisée du *modus ponens* (si P alors Q, P, donc Q) et du *modus tollens* (la contraposition : si P alors Q, non Q, donc non P). La difficulté est que, dans cette conception, la causalité est identifiée au connecteur logique de l'implication matérielle. Or, l'implication logique se définit par rapport à la table de vérité selon laquelle si Q est vrai, alors peu importe la valeur de vérité de P, l'implication est toujours vraie. Reformulée en termes de causalité, cette conception du connecteur logique de l'implication matérielle implique que peu importe la présence ou l'absence de la cause, l'effet se produira. On a voulu palier cette difficulté en proposant une implication stricte, qui peut se formuler de la façon suivante en logique modale :  $\neg\Diamond(P \wedge \neg Q)$ , c'est-à-dire « il n'est pas possible que (P et non-Q) ». De l'avis de Mario Bunge<sup>18</sup>, cette seconde formulation devait également s'avérer inadéquate.

Il ne s'agit pas d'insinuer que les défenseurs de l'induction et du vérificationnisme ne reconnaissent pas ces difficultés logiques, au contraire. À cet égard, l'on ne s'étonnera pas que des auteurs comme Reichenbach et Carnap se soient tournés vers le problème du fondement de la logique des probabilités et aient modifié le principe d'induction en conséquence. Du point de vue probabiliste, les énoncés scientifiques ne seraient plus strictement vrais ou faux, mais approximativement vrais ou approximativement faux. Au diapason avec l'aphorisme de l'évêque Joseph Butler selon lequel la probabilité est le guide même de la vie, l'objectif fondamental de la défense (*vindication*) de l'induction est d'acquérir la connaissance des probabilités. Reichenbach est ainsi amené à proposer une règle d'induction : « Si une section observée initiale comprenant  $n$  membres d'une séquence de plusieurs A contient  $m$  éléments avec l'attribut

B, stipulons que la limite de la fréquence relative de B dans A est contenue à l'intérieur de l'intervalle  $m/n \pm \delta^{19}$ . La règle d'induction constitue ainsi un instrument pour trouver des probabilités : lorsque l'événement A fait occurrence, la règle d'induction nous fait anticiper l'événement B avec la probabilité  $m/n \pm \delta$ . La valeur de l'écart-type  $\delta$  est quant à elle choisie par des aspects pragmatiques du contexte, à savoir le degré de précision requis dans une situation particulière. Reichenbach étend la notion d'« implication » de la logique classique du premier ordre à celle d'« implication probabiliste », notion qui exprime la mesure dans laquelle l'énoncé « e » confirme l'hypothèse « h » :  $x \in A \supset_p x \in B$ , où « p » est une quantité entre zéro et un inclusivement. Si cette adaptation pragmatiste justifie (constitue une défense de) l'applicabilité du principe d'induction, c'est néanmoins au détriment d'une perte de validité logique, puisqu'on ne peut plus parler de véritable dérivation logique dans le calcul des prédicats du premier ordre<sup>20</sup>. Par ailleurs, aux yeux de Popper, ce principe d'induction tombe sous le coup de sa précédente critique et doit être justifié à son tour, ce qui conduit à nouveau soit à une régression à l'infini, soit à l'apriorisme. Ces diverses difficultés attachées à la logique inductive sont à son avis insurmontables.

### 3. L'empirisme logique

Notre présentation schématique de l'empirisme logique est tirée d'un ouvrage de Gerard Radnitzky intitulé *Contemporary schools of metascience*<sup>21</sup>. Radnitzky distingue l'épistémologie traditionnelle (la philosophie des sciences) de la métascience dont il se réclame. L'épistémologie traditionnelle n'est pas une discipline scientifique, mais une branche de la philosophie qui produit des connaissances philosophiques sous la forme de croyances, de jugements de valeur et d'images idéales de la science au moyen d'une méthode d'argumentation critique. Au contraire, la métascience est une discipline scientifique qui produit des systèmes de savoir à propos de son objet, la science. La métascience est organisée comme un système cybernétique et procède par accumulation du savoir par la vérification empirique de ce savoir (sans jamais le démontrer définitivement) et par l'organisation de ce savoir en systèmes. La

méthode que préconise Radnitzky est la praxéologie, laquelle constitue une branche de la métascience dont le but est de formuler une théorie de la croissance du savoir qui ne soit ni psychologique ni sociologique, c'est-à-dire d'énoncer des critères objectifs de la croissance du savoir. De façon générale, la praxéologie<sup>22</sup> est une discipline qui se donne pour objet l'analyse de la «*praxis*», c'est-à-dire de la pratique ou de l'action humaine. Chez Radnitzky, la théorie praxéologique élabore des modèles de production du savoir et est considérée comme une science humaine de la décision et de l'action rationnelle. Sa praxéologie conçoit l'entreprise de recherche comme un système d'actions, de stratagèmes, de mouvements, etc<sup>23</sup>. À cet égard, la théorie praxéologique entend permettre au chercheur, qui doit décider de la manière la plus rationnelle de faire une recherche, d'accroître sa liberté de décision individuelle à l'aide de critères rendus explicites par la théorie de la pratique<sup>24</sup>. La perspective praxéologique de Radnitzky présente une entreprise de recherche — dans le cas qui nous intéresse, l'empirisme logique — comme un système d'actions constitué non d'un ensemble de personnes, mais d'un ensemble de publications. Chaque entreprise de recherche comprend un aspect synchronique (ou programmatique, l'école) et un aspect historique (la tradition). Si l'ensemble de publications est étudié dans une perspective systémiste ou synchronique, il porte le nom d'«*école de pensée*» avec des publications relativement contemporaines. Si le même ensemble de publications est étudié dans une perspective historique ou diachronique, il porte cette fois le nom de «*tradition intellectuelle*». La praxéologie de Radnitzky ne vise pas à décrire ni à expliquer historiquement une tradition, mais à la reconstruire rationnellement. En appliquant son modèle d'une entreprise de recherche aux publications des représentants de la tradition de l'empirisme logique, Radnitzky tente d'identifier un cadre, des outils intellectuels, un programme de recherche, etc. qui leur sont communs.

Distinguons d'abord deux grands blocs : le cadre et l'ensemble de la recherche autour du programme de recherche. Le cadre est une sélection qu'opère l'entreprise de recherche parmi les conceptions du monde, de la connaissance, des valeurs, etc., qui circulent

dans leur milieu intellectuel. Ce cadre se divise en deux parties principales : une partie cognitive et une partie axiologique. En premier lieu, nous pouvons considérer le milieu intellectuel et la partie cognitive du cadre de l'entreprise de recherche de l'empirisme logique, qui comprennent chacun six éléments, chaque élément de l'un allant de pair avec un élément de l'autre. Dans le cas des deux premières paires, il s'agit de réactions à l'encontre du milieu intellectuel. Ce milieu était dominé, d'une part par l'idéalisme allemand, c'est-à-dire le transcendantalisme qui régnait à l'époque du Cercle de Vienne; d'autre part par les discussions sur la théorie des jugements synthétiques *a priori* chez Kant. À ces deux éléments du milieu intellectuel correspondent deux éléments de la partie cognitive, d'une part l'opposition farouche de l'empirisme logique à toute forme de métaphysique; d'autre part le rejet par l'empirisme logique de toute proposition synthétique *a priori*. Les quatre autres paires constituent non des rejets, mais plutôt des emprunts au milieu intellectuel. Du côté du milieu intellectuel, nous avons l'entreprise de Whitehead et Russell dans les *Principia Mathematica*; la théorie du langage idéal de Descartes et Leibniz à Peano, Frege et Russell; le conventionnalisme de l'époque et finalement la tradition des empiristes anglo-saxons (dont Mill et Hume). Du côté de la partie cognitive, nous avons la théorie de la dichotomie entre analytique (sciences formelles) et synthétique (sciences empiriques); l'idée de l'isomorphisme entre le langage idéal et la forme logique du monde; le principe selon lequel toute orientation dans le monde est plus ou moins conventionnelle et finalement le principe de l'empiricité des théories.

Le milieu intellectuel fournit également la partie axiologique du cadre de l'empirisme logique, c'est-à-dire les idéaux que les représentants de cette entreprise de recherche partagent et que nous pouvons séparer en trois aspects. Le premier élément correspond à l'idéal du savoir : le savoir doit i) être certain comme le savoir formel exemplifié par les vérités logiques, ii) définitif et éternel comme le savoir en mathématique et iii) unifié, c'est-à-dire présenté sous une forme axiomatisée. De cet idéal du savoir dérive le second élément, l'idéal de la science qui correspond, pour les sciences formelles,

à la métamathématique, et pour les sciences empiriques, à la physique. Finalement, cela donne lieu au troisième élément, l'idéal de l'épistémologie, c'est-à-dire la conception générale de ce que les chercheurs voudraient que soit leur discipline. Cet idéal correspond à la métamathématique et consiste à imiter les *Principia Mathematica* de Whitehead et Russell.

Le programme de recherche est le fruit d'une première prospection des tâches et est gouverné par la partie axiologique du cadre de l'entreprise de recherche. Le programme de recherche de l'empirisme logique est double. Il consiste à constituer une philosophie scientifique et à clarifier et préciser l'idéal de la science unifiée.

La « philosophie scientifique » dont il est question comporte deux aspects : a. l'étude des systèmes formels en métamathématique et en logique et b. l'étude de la structure formelle et du caractère empirique du langage de la science. L'idéal de la science unifiée consiste à réduire le langage des différentes sciences à un langage commun. Comme on le sait, la planification de la recherche de l'empirisme logique est guidée par une « stratégie » qui consiste à imiter la métamathématique<sup>25</sup>.

Le plan général de la recherche consiste à formaliser et axiomatiser les concepts fondamentaux de la science idéale. Les projets particuliers de l'empirisme logique peuvent être regroupés sous trois grands titres : le problème de la signification empirique, le problème de la confirmation et le problème de l'explication. Le problème de la signification empirique consiste à développer des critères d'adéquation permettant d'introduire des termes et des concepts non-logiques dans le schéma de la science unifiée visée par l'empirisme logique. Autrement dit, il s'agit de spécifier un algorithme (des règles ou des conditions nécessaires et suffisantes<sup>26</sup>) permettant de décider si un énoncé (ou un texte) contient un contenu, à savoir s'il fait référence au monde extérieur. Par exemple, ce qui distingue une explication (ou une hypothèse scientifique) d'un théorème mathématique est qu'une théorie physique possède un contenu représentationnel (fait référence au monde), alors que les formules tautologiques de la logique formelle n'en ont aucun.

Le problème de la signification empirique implique ainsi deux démarcations, à savoir celle entre physique et mathématiques, puis celle entre science empirique et métaphysique. Le problème de la confirmation des théories (ou connaissances) scientifiques ne donne pas lieu, pour sa part, à une théorie de la confirmation à proprement parler, mais plutôt, comme nous l'avons vu avec Carnap, à une tentative de fondation de la logique inductive guidée par l'objectif de répondre à la question « qu'est-ce que signifie le fait que l'énoncé  $e$  confirme l'hypothèse  $h$  au degré  $r$ ? ». Finalement, le problème de l'explication, qui s'inscrit dans le cadre métamathématique guidant les recherches de l'empirisme logique, prend les traits d'un problème concernant la théorie de la démonstration ou théorie de la preuve (*proof theory*). L'explication est ainsi conçue comme un argument dont tout l'intérêt consiste à montrer que les prémisses sont liées à la conclusion d'une manière formellement acceptable. Le paradigme le plus simple d'une telle explication E est «  $\{C\} \vdash E$  », où C représente une cause quelconque (des conditions ou circonstances initiales).

Selon Radnitzky, l'étude de la mise en œuvre de ces trois projets particuliers met en évidence que la partie du programme de recherche consistant à constituer une philosophie scientifique a été complètement supplantée par l'autre, à savoir clarifier et préciser l'idéal de la science. Les trois projets particuliers ne sont alors ni plus ni moins que des sous-tâches de la clarification de l'idéal de la science unifiée. À mesure que l'empirisme logique est de plus en plus guidé par cet idéal, le programme initial est relégué à l'arrière-plan. Les objectifs fondamentaux consistent alors simplement à élaborer, dans les termes de Radnitzky, une métascience ressemblant le plus possible à la métamathématique. Dans ces conditions, l'empirisme logique n'est pas guidé, au niveau de ses projets particuliers, par un véritable programme de recherche pouvant éventuellement être amélioré, mais est plutôt aveuglément assujéti à un idéal de la science et à un idéal de l'épistémologie qu'on ne peut jamais remettre en question<sup>27</sup>. Quels sont alors les résultats de la recherche de l'empirisme logique selon Radnitzky? Le résultat le plus intéressant consiste à avoir élaboré des outils intellectuels d'une grande précision, diverses techniques de l'analyse logique et principalement de l'axiomatisation

des théories scientifiques. Mais le seul contact qu'entretient cette entreprise de recherche avec son milieu intellectuel est celui issu de son intérêt pour l'explication scientifique. Hormis ce contact restreint, l'empirisme logique constitue un cercle fermé dont les représentants ne se soucient guère de l'intérêt que suscitent leurs travaux hors de l'entreprise de recherche. Les trois principaux projets particuliers ont abouti à des impasses sans issue. De ce fait, les représentants de l'empirisme logique se sont vus contraints de faire des concessions mineures — par exemple le tournant probabiliste que nous avons vu avec le problème de l'induction —, mais sans jamais réajuster leurs présupposés, leurs idéaux et leur programme de recherche. Notons à cet égard que lorsque vient le temps de spécifier qui sont les penseurs inclus et exclus de l'empirisme logique à partir de la définition radnitzkienne, le cas le plus problématique est celui de Quine. Au début de sa présentation de l'empirisme logique, Radnitzky considère Quine comme un de ses représentants. Un peu plus loin, Radnitzky se ravise et considère que Quine devrait être étudié à part puisqu'il rejette une conception fondamentale de l'empirisme logique, à savoir la distinction « analytique-synthétique ». Radnitzky considère en effet la dualité « formel-analytique » et « empirique-synthétique » comme étant caractéristique du fondement philosophique de l'empirisme logique<sup>28</sup>. Selon cette distinction, les énoncés analytiques sont des énoncés dont la vérité est fondée seulement sur la signification des mots et indépendamment des faits, alors que les énoncés synthétiques sont des énoncés dont la vérité repose sur des questions de fait et de ce qui se déroule dans le monde. Revenons maintenant à notre objectif qui consiste à clarifier l'expression « scepticisme épistémologique » chez Quine.

#### *4. Une impasse cognitive*

Sur la base de la distinction « analytique-synthétique », l'empirisme logique renvoie les questions concernant le contexte de découverte à la science empirique (synthétique *a posteriori*) et les questions concernant le contexte de justification à la logique (analytique *a priori*)<sup>29</sup>. Pour l'empirisme logique, la question épistémique, qui concerne la façon dont on acquiert une théorie, a en effet perdu toute

pertinence, puisqu'elle ne peut être caractérisée par un ensemble de règles logiques, c'est-à-dire un algorithme conduisant, en un nombre fini d'étapes, à un résultat valide, en l'occurrence une découverte, ce qui présuppose que la question épistémique ne peut recevoir une réponse adéquate que dans une perspective formaliste ou logico-déductive. L'empirisme logique se détourne donc du contexte de découverte et se consacre exclusivement au contexte de justification, qui concerne la question fondationnelle : la façon dont on justifie les connaissances. Le programme de recherche initial de l'empirisme logique était double : d'une part, constituer une philosophie scientifique, d'autre part clarifier et préciser l'idéal de la science unifiée. Comme nous l'avons dit en introduction, ces deux aspects vont de pair. Clarifier les concepts scientifiques en les définissant dans les termes de la logique des prédicats du premier ordre et en les ordonnant déductivement revient à fonder logiquement la théorie scientifique faisant l'objet de la traduction ou de la reconstruction rationnelle. Carnap entendait réaliser cette fondation à l'aide d'énoncés analytiques *a priori*. Du côté des sciences formelles, la difficulté que souligne Quine est qu'historiquement, on a montré que les mathématiques ne se réduisent pas strictement à la logique, mais uniquement à la logique symbolique additionnée de la notion de «classes» empruntée à la théorie axiomatique des ensembles. Du côté des sciences empiriques, la difficulté est qu'il n'est tout simplement pas possible de rendre compte du monde extérieur uniquement à l'aide d'une suite de définitions purement *a priori* comme entendait le faire Carnap dans son ouvrage *La construction logique du monde*<sup>30,31</sup>. Nous avons vu non seulement que, des deux tâches principales du programme de recherche de l'empirisme logique, la première (constituer une philosophie scientifique) a été remplacée par la seconde (clarifier et préciser l'idéal de la science unifiée), mais également que les trois projets particuliers associés à la seconde tâche (les problèmes de la signification empirique, de la confirmation et de l'explication) ont abouti à des impasses. La critique de la théorie vérificationniste de la signification prend ainsi toute son importance : s'il n'est pas possible de confirmer ni de réfuter une théorie ou un énoncé théorique, c'est-à-dire s'il n'est pas possible de dire s'ils sont vrais ou faux,

alors comment pourrait-on en connaître la signification ? Comme mentionné en introduction, les échecs répétés du projet visant à conférer à la science un fondement irréfutable ont alimenté, selon Quine, une vague de scepticisme épistémologique grandissante. Ces échecs donnent lieu à l'idée qu'il n'y a pas, contrairement à l'idéal de la philosophie scientifique de l'empirisme logique, de tel fondement ultime et donc qu'il y a seulement plusieurs fondements possibles pour chaque discipline scientifique particulière — et les mieux placés pour s'en occuper sont les scientifiques eux-mêmes. Non sans une certaine ironie, c'est sur la base de la distinction DJ que l'empirisme logique entendait spécifier le champ d'investigation dans lequel la philosophie prétendait contribuer à l'avancement des connaissances. Or, ce que montrent les échecs répétés du projet fondationnaliste (par exemple la fondation logique des probabilités) est que le contexte de justification, et non seulement le contexte de découverte, sont du ressort des sciences. Il y a lieu de parler d'une impasse cognitive pour l'épistémologie telle que conçue par l'empirisme logique, puisque celui-ci s'est non seulement résigné à remplacer la science au niveau du contexte de découverte, mais s'est également révélé incapable de la fonder. Explicitée à l'aide de la distinction DJ, l'expression «scepticisme épistémologique» chez Quine signifie que le contexte de justification n'est pas du ressort de l'épistémologie telle qu'elle est conçue par l'empirisme logique. Cette conséquence, loin de nous éloigner de l'épistémologie, nous enjoint plutôt — à l'instar de Quine qui préconise l'épistémologie naturalisée comme solution de rechange — à repenser l'épistémologie afin d'en élaborer une conception qui puisse échapper à un tel cul-de-sac.

- 
1. Conformément à l'usage répandu notamment en philosophie analytique, des guillemets («») seront utilisés lorsqu'il sera fait mention plutôt qu'usage d'un terme.
  2. Willard Van Orman Quine, *Ontological relativity: and other essays*, New York, Columbia University Press, 1969, p. 87 (nous traduisons).
  3. Quine emploie l'expression «*epistemological nihilism*». Lors de la traduction, le terme «scepticisme» a été préféré au terme «nihilisme» étant donné qu'il rend aussi bien compte de la signification véhiculée

par l'expression de Quine et qu'il est plus usité en matière de théorie de la connaissance (par exemple le scepticisme inhérent à tout empirisme, une tradition dont Quine se réclame).

4. Il importe d'éviter une ambiguïté dans l'emploi des termes «épistémique» et «épistémologique». Il est de coutume, dans la tradition philosophique anglo-saxonne, de distinguer entre les thèses philosophiques qui s'intéressent à la façon dont l'être humain connaît naturellement et spontanément, l'*epistemology*, et le champ de recherche en philosophie contemporaine qui prend pour objet la science moderne (et notamment sa justification), la *philosophy of science*. Nous traduirons (ou utiliserons) en conséquence et autant que possible «*epistemology*» par (pour) «théorie de la connaissance» et «*philosophy of science*» par (pour) «épistémologie». La difficulté est que Quine, qui s'intéresse d'abord et avant tout à l'*epistemology*, n'admet aucune distinction tranchée entre ces deux champs de recherche, puisqu'à son avis la science n'est qu'une extension, plus raffinée et plus abstraite, de la connaissance naturelle, de sorte qu'à son avis les connaissances naturelles sont sur le même continuum que les mathématiques.
5. Hans Reichenbach, *Experience and prediction*, Chicago, III. : University of Chicago Press, 1938 (1952), pp. 5-8 pour le reste du paragraphe.
6. Reichenbach fait appel à la distinction entre «relations externes» et «relations internes» comme une heuristique afin de relever les champs distinctifs de l'épistémologie et de la sociologie, mais n'entérine pas nécessairement les présupposés relatifs à cette distinction (notamment l'essentialisme défendu par les idéalistes du 19<sup>e</sup> siècle). Les relations internes sont censées être des relations intrinsèques (à ne pas confondre avec des qualités intrinsèques) à la nature des choses, énoncés, ou faits mis en relation. Par exemple, l'arc d'un cercle est relié de façon interne (intrinsèque) au centre de ce cercle au sens où il ne pourrait être cet arc sans entretenir une relation avec ce centre.
7. Wesley Salmon, *Logic*, Englewood Cliffs, N.J. Prentice-Hall, 1973, p. 10.
8. Notons que ce titre génère une confusion puisque Popper s'intéresse au contexte de justification et non au contexte de découverte dans cet ouvrage. Dès la première page de celui-ci, Popper précise que sa thèse est que la tâche de la logique de la connaissance scientifique consiste à fournir une analyse logique des méthodes des sciences empiriques ; or, à son avis, le processus de découverte ne saurait faire l'objet d'une telle analyse logique.
9. Christian Hallé, *Thomas Kuhn et l'approche logico-normative*, Québec, Université Laval, Accès via Proquest (thèse de maîtrise), p. 13.

10. Karl Raimund Popper, *La logique de la découverte scientifique*, traduit de l'anglais par Nicole Thyssen-Rutten et Philippe Devaux, préface de Jacques Monod, Paris, Payot, 1982, p. 27.
11. Christian Hallé, *op. cit.*, p. 14.
12. Popper, *op. cit.*, p. 36.
13. Un énoncé strictement universel est un énoncé portant sur un nombre infini de cas particuliers sans aucune restriction quant au temps ou à l'espace et se distingue donc de l'énoncé numériquement universel, qui porte sur tous les cas particuliers spécifiques à un lieu et un temps donnés ; par exemple, «tous les enfants» versus «tous les enfants présents dans le local 501 de l'école Sacré-Cœur le 21 janvier 1991».
14. Hans Reichenbach, *The Rise of Scientific Philosophy*, Berkeley, University of California Press, 1963, p. 231.
15. Hans Reichenbach, «Kausalität und Wahrscheinlichkeit» dans *Erkenntnis*, vol. 1, 1930, pp. 158-188, p. 186 ; cité dans Popper, *op. cit.*, p. 24.
16. Popper, *op. cit.*, pp. 23-25.
17. Wesley C. Salmon, «Hans Reichenbach's Vindication of Induction» dans *Erkenntnis* (1975-), vol. 35, n° 1/3, Special Volume in Honor of Rudolf Carnap and Hans Reichenbach (Jul., 1991), pp. 99-122, pp. 99-100.
18. Mario Bunge, *Causality in Modern Science* (3<sup>e</sup> ed.), New York, Dover Publications Inc., 1979, p. 244. Disponible aussi sous Mario Bunge, *Causality : the place of causal principle in modern science*, Cleveland, World Co., 1963 (2<sup>e</sup> ed.).
19. Salmon, «Hans Reichenbach's Vindication of Induction», *loc. cit.*, pp. 101-102 (nous traduisons).
20. Après avoir précisé que la procédure inductive constitue une méthode d'anticipation sous la forme d'essais et erreurs, Reichenbach soutient que «*When we use the logical conception of probability, the rule of induction must be regarded as a rule of derivation, belonging in the metalanguage*» (c'est l'auteur qui souligne). Mais il admet implicitement aussitôt que le concept de «*dérivation inductive*» n'est pas strictement synonyme avec celui de «*dérivation déductive*» : «*It [la règle de dérivation inductive] is comparable to the rule of inference of deductive logic, but differs from it in that the conclusion is not tautologically implied by the premises*», cf. Hans Reichenbach, *The theory of probability*, University of California Press, 1949, p. 448 pour les deux citations.
21. Gerard Radnitzky, *Contemporary schools of metascience*, Göteborg, Akademiförlaget, 1968.

22. D'abord revendiquée par Ludwig von Mises dans son ouvrage *L'Action Humaine* en 1949.
23. Gerard Radnitzky, «Towards a theory of reserach which is neither logical reconstruction nor psychology or sociology of science», dans *Quality and Quantity*, vol. 6 (2), 1972 pp. 193-238.
24. Gerard Radnitzky, «La philosophie de la recherche scientifique», dans *Archives Philosophiques*, 1974, pp. 55-56.
25. François Tournier, *Le problème de la détermination sociale en science*, Université du Québec à Trois-Rivière, 1983, p. 98.
26. Radnitzky donne cet exemple : «*A sentence S of a language L is ES [empirically significant] if and only if: (1) S belongs to L (is wellformed in L) and is not analytic in L; (2) there is a finite set of sentences {O1, O2, ..., On} such that: (a) O1, O2, ..., On are observation sentences, b) the conjunction of O1, O2, ..., On is consistent in L; (3) S is derivable from the set {O1, O2, ..., On}*», cf. Radnitzky, *Contemporary schools of metascience*, p. 121.
27. François Tournier, *op. cit.*, p. 99.
28. Radnitzky, *Contemporary schools of metascience*, p. 58.
29. Si une théorie scientifique est confrontée au moyen d'expérimentations dites empiriques, l'ensemble d'énoncés en jeu contiendra à la fois des énoncés empiriques et des énoncés théoriques. Mais, comme nous l'avons vu, la justification prend pour l'empirisme logique la forme d'un argument, dont la validité ne concerne que la forme logique, c'est pourquoi elle est *a priori*.
30. Rudolf Carnap, *La construction logique du monde*, Paris, Librairie Philosophique J. Vrin, 2002.
31. Quine, *op. cit.*, p. 70 et 76.





# L'influence décisive de Nietzsche sur le jeune Camus

OLIVIER MASSÉ, *Université de Montréal*

## *Introduction*

Que l'œuvre de Camus ait été influencée par Nietzsche, cela semble de plus en plus évident pour les lecteurs attentifs. Alors que les biographies, à l'époque, faisaient généralement état d'une première découverte par Camus des écrits de Nietzsche au milieu des années 1930<sup>1</sup>, les sources plus récentes laissent plutôt croire que le philosophe franco-algérien a côtoyé le philosophe allemand dès ses premières lectures<sup>2</sup>. Il suffit par exemple de penser au *Manuel de philosophie* d'Armand Cullivier, lu par la très grande majorité des lycéens de l'époque après sa parution en 1927, qui lui fit assurément découvrir Nietzsche<sup>3</sup>. Mais surtout, avec la parution des œuvres complètes de Camus, nous avons désormais entre nos mains un article sur la musique que Camus a publié à 19 ans, en 1932, tout simplement intitulé *Essai sur la musique*, dont le point de départ était les réflexions esthétiques de Schopenhauer et de Nietzsche<sup>4</sup>. À ses débuts, donc, Camus lisait Nietzsche, mais aussi à la fin de sa vie lorsqu'un exemplaire du *Gai savoir* a été retrouvé dans le funestement connu accident de voiture qui a causé sa mort<sup>5</sup>. De cet article du lycée sur la musique jusqu'à sa mort, Camus n'a eu de cesse de lire et de relire Nietzsche, sans jamais reprendre simplement le contenu de ses essais. À eux seuls, ses *Carnets* — qui regroupent des réflexions philosophiques et politiques, des plans en vue de l'élaboration de ses œuvres, des citations de philosophes et d'écrivains divers, ou encore des notes intimes — font référence au philosophe allemand pas moins de 65 fois<sup>6</sup> ! Ce qui surprend d'autant plus dans les références de Camus à Nietzsche, c'est qu'elles ne se limitent pas simplement à ses idées, mais qu'elles concernent aussi sa vie. Il suffit de citer ce passage des *Carnets* pour montrer à quel point le philosophe franco-algérien s'identifie à l'homme lui-même et non pas seulement

au philosophe: «Nietzsche a travaillé puis sombré dans la folie. Je n'ai *jamais pu lire sans pleurer* le récit de l'arrivée d'Overbeck, son entrée dans la pièce où Nietzsche fou délire, puis le mouvement de celui-ci qui se jette dans les bras d'Overbeck en pleurant. Devant cette maison j'essaie de penser à lui *que j'ai toujours aimé d'affection autant que d'admiration*, mais en vain<sup>7</sup>».

Ses *Carnets*, empreints de tels rapprochements, invitent à mettre en lumière l'incidence du philosophe allemand sur la pensée de Camus<sup>8</sup>. En effet, l'œuvre philosophique de Camus, relativement courte, ne cesse de faire allusion explicitement ou implicitement aux écrits de Nietzsche. Tout au long du présent article, il sera question de soutenir la thèse selon laquelle la philosophie de Camus, sans s'y réduire, est d'inspiration nietzschéenne. Pour ce faire, il a été décidé de se concentrer exclusivement sur l'influence décisive de Nietzsche sur le jeune Camus, sur son entrée en philosophie, ce qui écartera de notre étude les deux écrits plus proprement philosophiques du penseur franco-algérien, *Le mythe de Sisyphe* et *L'homme révolté*. Il faudra bien plutôt se référer à son *Essai sur la musique* (1932), à *L'envers et l'endroit* (1937) et à *Noces* (1939), qui nous permettront de considérer l'influence de Nietzsche sur le jeune Camus à travers trois thèmes : l'esthétisme, l'*amor fati* et la critique du christianisme.

Cet article répondra à trois défis. D'abord, indirectement, je tenterai de déconstruire un mythe largement répandu selon lequel Camus serait, au mieux un philosophe pour classes terminales ou un piètre moraliste<sup>9</sup>, au pire un écrivain dont la pensée ne saurait s'élever à la philosophie<sup>10</sup>. Il est espéré par l'auteur de cet article que le rapprochement entre Nietzsche et le jeune Camus pourra montrer l'étendue et la complexité des questionnements au fondement de ses essais de jeunesse, notamment rassemblés dans *L'envers et l'endroit* et *Noces*, ce qui permettra de le considérer à juste titre, dès sa jeunesse, comme un philosophe. Puis, il sera bien entendu question de prouver une bonne fois pour toutes que la philosophie de Camus s'inspire bel et bien de Nietzsche. Il est d'usage dans les études portant sur Camus de citer au passage un aphorisme du célèbre philosophe allemand ou de sous-entendre certains rapprochements entre ces deux penseurs au tournant d'un paragraphe ou d'une phrase<sup>11</sup>. Mais rares sont

les recherches qui s'attardent véritablement sur de tels liens, et il faudra remédier à ce manque. Finalement, le présent article vise une réhabilitation de la pertinence des œuvres de jeunesse du philosophe franco-algérien. Très souvent, l'étude de la philosophie de Camus est limitée au *Mythe de Sisyphe* et à *L'homme révolté*. Un tel choix n'est pas sans fondement, étant donné que ces deux essais constituent les œuvres « idéologiques » (au sens philosophique du terme, pour reprendre l'expression de Camus) au fondement de ses deux cycles de création, celui sur l'absurde et celui sur la révolte. Toutefois, cela ne légitime pas que l'on écarte souvent du revers de la main *L'envers et l'endroit* et *Noces*, plutôt considérés comme de simples productions littéraires sans contenu philosophique.

*1. L'entrée en philosophie : un questionnement esthétique sous l'ombre de Nietzsche*

L'année 1932 constitue un moment important de la vie du jeune Camus : pour la première fois, alors âgé de 19 ans, il publie dans une revue (*Sud*) des réflexions sur des auteurs et des sujets de son choix. D'abord, en mars, il publie un article intitulé *Un nouveau Verlaine*, où il tente de réhabiliter la réputation du poète maudit. Puis, le numéro suivant, en mai, il publie un article intitulé *Jean Rictus, le poète de la misère*, où il tente à nouveau la réhabilitation d'un poète, Jean Rictus, aujourd'hui peu connu, en montrant que son poème *Soliloques du pauvre* constitue l'expression sincère de l'état d'âme du « misérable », du « pauvre ». Cependant, c'est en mai 1932 que la contribution de Camus à la revue *Sud* devient intéressante pour mon propos, puisqu'il fait ce que l'on pourrait nommer son entrée en philosophie, publiant du même coup *La philosophie du siècle*, courte critique de l'ouvrage de Bergson *Les Deux Sources de la morale et de la religion*, et *Essai sur la musique*, déjà mentionné dans l'introduction. C'est cet article qui fera l'objet de la présente réflexion, d'abord parce qu'il livre les toutes premières réflexions philosophiques personnelles de Camus, puis parce que Nietzsche y joue un rôle des plus importants.

Le ton de l'*Essai sur la musique*, caractéristique de la jeunesse, est ambitieux, catégorique et péremptoire. Camus y propose l'esquisse

d'une esthétique qui tente de dépasser le dualisme entre les conceptions réalistes et idéalistes de l'art. Selon lui, l'art ne doit pas être une simple copie « neutre » du réel, pas plus qu'il ne doit « embellir » le réel au point de proposer un exemple de la conduite morale par excellence. En fait, le *telos* de l'art consiste à créer « un monde de Rêve<sup>12</sup> », le monde idéal, qui voilerait le monde réel avec toutes ses souffrances et ses horreurs<sup>13</sup>. Comme on peut déjà le constater, cette réflexion esthétique s'ouvre sur un constat éminemment pessimiste. Après sa courte introduction, Camus livre ses réflexions sur la théorie esthétique de Schopenhauer, où il retrouve la conception de l'art qu'il veut défendre. Chez Schopenhauer, et Camus le voit bien, le but de l'existence humaine est de « réaliser la volonté », et c'est bel et bien la musique qui l'incarne le mieux parmi les arts<sup>14</sup>. Mais comment une telle incarnation de la volonté dans la musique peut-elle mener à cette contemplation du beau et du monde idéal dont parlait Camus dans son introduction ? Là, en fait, est tout le problème : l'*Essai sur la musique* identifie explicitement volonté et vie — alors même qu'une lecture rigoureuse de la traduction de Burdeau utilisée à l'époque de *Le monde comme volonté et représentation* montre que la volonté renvoie aussi bien au règne organique qu'au règne inorganique et qu'elle n'en est pas l'équivalent<sup>15</sup> —, et Camus considère donc que, réalisant la volonté, la musique rend compte de l'essence même de la vie, ce qui transporterait selon lui le spectateur hors de ce monde de souffrances et d'horreurs. Camus a tout à fait raison de noter que la musique transporte le spectateur hors des souffrances de ce monde en le ramenant à ce qu'il a de plus pur en lui. Toutefois, cette essence du monde n'est pas à chercher, du moins selon Schopenhauer, dans la vie, mais plutôt, encore plus fondamentalement, dans la volonté. Ce « voyage » dans l'essence du monde, c'est-à-dire dans la volonté, crée selon Camus une illusion qu'il associe au rêve, bien que Schopenhauer ne décrive pas la musique en ces termes.

Après ce court examen de la théorie esthétique de Schopenhauer, l'*Essai sur la musique* passe au rapport de Nietzsche à la musique, ce dernier constituant le véritable centre d'intérêt de Camus dans cet article<sup>16</sup>, d'autant plus que Camus considère Nietzsche à la fois comme le continuateur et le pourfendeur de Schopenhauer<sup>17</sup>. Selon

Camus, il y a deux Nietzsche du point de vue esthétique : le Nietzsche de *La Naissance de la tragédie* et le Nietzsche du petit essai *Sur la musique et la parole*. Le premier reste très attaché à l'esthétique de son « maître » et constitue une référence pour le philosophe franco-algérien. C'est une certaine lecture, assez cohérente avec Nietzsche, de la théorie de l'apollinisme et du dyonysisme qui permettra à Camus de retrouver encore une fois l'esthétique qu'il veut substituer aux concepts réalistes et idéalistes de l'art. *L'Essai sur la musique* prend acte du fait que la pensée grecque était éminemment pessimiste et que son art était donc une tentative de « fuir une vie trop douloureuse<sup>18</sup> ». Toute la conception grecque de l'art partirait donc d'une tentative inconsciente, celle de fuir la douleur, qui expliquerait aussi bien le versant apollinien que dionysiaque de l'art grec. Pour Nietzsche, rappelons-le, l'apollinisme consiste à créer une illusion dont le but est d'élever l'individu dans un règne de beauté qui nous fera oublier le monde, tandis que le dyonysisme consiste à transporter l'homme dans un état d'ivresse tel que, en oubliant sa propre individualité, il se réconciliera avec le monde en entier<sup>19</sup>. Or, encore selon Nietzsche et selon Camus qui s'en inspire, c'est la tragédie grecque, s'incarnant le plus parfaitement dans la musique, qui a su unir ces deux tendances naturelles. La thèse de *L'Essai sur la musique*, ici plutôt simple, soutient que les deux versants de la tragédie grecque, Apollon et Dionysos, tels que considérés par cette première esthétique nietzschéenne procèdent en fait du même *telos*, celui de fuir le monde par le rêve. Cette thèse s'accorde parfaitement, évidemment, avec le versant apollinien de la tragédie grecque, de même qu'avec son versant dionysiaque, même si cela apparaît plutôt contradictoire, car l'ivresse dionysiaque partage fondamentalement le même but que la métamorphose apollinienne, qui est en dernière instance la formation d'une illusion que Camus associe au rêve et qui délivrera l'homme de sa douleur. Cela mène d'ailleurs Camus à répudier le Nietzsche qui se détourne de sa première œuvre et qui s'abaisse, faute de trop connaître l'homme, à ne plus considérer la beauté parfaite et tragique de l'œuvre wagnérienne<sup>20</sup>.

Commentant *L'Essai sur la musique*, Roger Grenier a certainement raison de douter que Camus se soit pleinement intéressé à la musique

d'un point de vue philosophique durant sa vie<sup>21</sup>. En effet, après cette œuvre de jeunesse, il ne publiera pas de réflexion sur la musique, et ses *Carnets*, dont on connaît l'importance, font peu mention d'une réelle influence de la musique sur son œuvre. Toutefois, comme tous ceux qui se risquent à commenter cet article, Grenier fait l'erreur de ne pas soulever à quel point l'*Essai sur la musique* — certes qui postule plus qu'il ne démontre — peut être important pour comprendre la genèse du corpus camusien. D'abord, si, comme le veut Camus dans son *Discours de Suède*, sa vie a été influencée par l'idée qu'il se faisait de son art, il est impossible de ne pas porter attention à sa toute première réflexion esthétique<sup>22</sup>. Bien qu'il n'ait jamais publié une théorie esthétique en tant que telle, il est possible d'observer des réflexions sur l'art qui réagissent à cette première esthétique, notamment dans *Le mythe de Sisyphe*, mais surtout dans *L'homme révolté*. Puis, c'est le premier texte qui aborde de près la philosophie de Nietzsche, ce qui n'est pas un détail pour le lecteur attentif à l'œuvre de Camus. Cela montre hors de tout doute que Nietzsche a joué un rôle clé dans l'éducation philosophique du jeune Camus. Détourné, lors de la rédaction de *L'envers et l'endroit*, des questionnements esthétiques, il continuera néanmoins de subir l'influence de Nietzsche.

## 2. Le spectre de Nietzsche dans *L'envers et l'endroit*

Si l'article sur la musique a le mérite d'être l'un des premiers textes publiés par le jeune Camus dans une revue, *L'envers et l'endroit* a la caractéristique intéressante d'être le premier recueil « d'essais » qu'il publie dans une maison d'édition (en 1937). Les guillemets sont à utiliser avec prudence : bien qu'il ne soit pas de mon intention d'ouvrir un débat purement nominaliste et que Camus lui-même les considérerait comme des essais, il faut avouer que les textes publiés dans ce recueil, loin des discours réflexifs auxquels le lecteur pourrait s'attendre, sont à la limite de la description narrative<sup>23</sup>. Mais pourquoi, alors, doit-on s'intéresser à pareil recueil ? La réponse se trouve dans la fameuse préface à *L'envers et l'endroit*, écrite en 1958 : « Pour moi, je sais que ma source est dans *L'envers et l'endroit*, dans ce monde de pauvreté et de lumière où j'ai longtemps vécu et

dont le souvenir me préserve encore des deux dangers contraires qui menacent tout artiste, le ressentiment et la satisfaction<sup>24</sup>».

Lorsqu'un auteur indique que sa «source» se trouve dans une œuvre en particulier, il me semble nécessaire de porter attention à cette dernière. Écrire sur la musique dans la perspective de Nietzsche, cela ne fait pas encore d'un penseur un philosophe d'inspiration nietzschéenne; encore faut-il poursuivre le dialogue dans ses productions subséquentes. Or, *L'envers et l'endroit* — qui regroupe cinq essais : *L'ironie*, *Entre oui et non*, *La mort dans l'âme*, *Amour de vivre*, *L'envers et l'endroit* — ne fait aucunement mention de Nietzsche. Camus se serait-il donc détourné du philosophe allemand? En fait, le lecteur attentif assiste à tout le contraire. Dans ce recueil, le philosophe franco-algérien reprend subtilement le thème nietzschéen de l'*amor fati* (que Nietzsche a lui-même emprunté aux Stoïciens<sup>25</sup>) et l'applique à des situations concrètes. De manière préliminaire, qu'est-ce que l'*amor fati*? Pour Nietzsche, c'est le *Ja sagen*, le grand «oui», le fait d'accepter et d'aimer son destin. Ce qui est valorisé, c'est la vie elle-même, aussi bien dans ce qu'elle a de «positif» que de «négatif», à laquelle il faut consentir. Le *Gai savoir* est explicite sur ce point : «Je veux apprendre toujours plus à voir dans la nécessité des choses le beau : je serais ainsi l'un de ceux qui embellissent les choses. *Amor fati* : que ce soit dorénavant mon amour ! [...] Je veux même, *en toutes circonstances*, n'être plus qu'un homme qui dit oui!<sup>26</sup>» Dans ce passage, Nietzsche affirme très clairement qu'il faut dire «oui» à toutes les circonstances, à tout ce qui se présente à nous, bref à notre destin. C'est une citation de Nietzsche notée dans ses *Carnets* quelques années après la parution de *L'envers et l'endroit* qui, pour la première fois, m'a mis sur la piste de la possibilité d'une reprise par Camus de la thèse de l'*amor fati*. Le philosophe franco-algérien recopie le fameux «ce qui ne me fait pas mourir me rend plus fort» du *Crépuscule des idoles*<sup>27</sup>. Or, comme s'en réclame Nietzsche, cet essai a pour tâche d'«ausculter les idoles», de détruire à coups de marteau les illusions qui ont recouvert la réalité<sup>28</sup>. Ce qui se met en branle dans *L'envers et l'endroit* — même si l'avant-propos n'en rend pas compte en ces termes —, c'est une affirmation de l'importance de l'*amor fati*,

puisqu'il faut, pour aimer notre destin, retourner au monde lui-même et ne plus se laisser bernier par les idoles. Il est impossible de savoir si Camus a développé une telle interprétation du *Crépuscule des idoles* ou encore s'il a même lu cet essai ou tout autre essai abordant le thème de l'*amor fati* à l'époque de la rédaction de *L'envers et l'endroit*, mais cette hypothèse, qui n'est pas invraisemblable compte tenu de l'affinité manifeste entre les deux penseurs, permet de mieux comprendre d'un point de vue philosophique le but de ce recueil.

Il faut passer à l'analyse de *L'envers et l'endroit*. D'abord et avant tout, ce qui joue en faveur d'un rapprochement entre Camus et Nietzsche sur le thème de l'*amor fati*, c'est la forme même des essais de *L'envers et l'endroit*. En effet, dans les essais de *L'envers et l'endroit* comme de *Noces*, Camus propose de tirer des règles d'action, de conduite, de sagesse à partir de la manière dont le monde, dans sa quotidienneté, nous apparaît. Or, lorsque l'on dit «oui» au destin, n'est-il pas tout à faire normal, si ce n'est souhaitable, d'appuyer nos conclusions normatives sur un discours descriptif? Cela expliquerait donc le choix de Camus pour une écriture à la fois descriptive et normative, au sens où tout le propos de *L'envers et l'endroit*, comme de *Noces*, sera de célébrer notre destin, celui qui nous lie au monde, en prenant conscience de sa beauté. Il faut aussi remarquer que les conclusions des essais en présence abordent presque toutes l'acceptation et l'amour du destin. La dernière ligne d'*Entre oui et non* affirme qu'«il y a des gens qui préfèrent regarder leur destin dans les yeux<sup>29</sup>», ce qui dénote une acceptation et une lucidité face à notre *fati*. Dans *L'ironie*, face à trois destinées différentes, Camus se pose sobrement la question suivante: «Qu'est-ce que ça fait, si on accepte tout?<sup>30</sup>» Dans *Amour de vivre*, il commence son dernier paragraphe en écrivant ceci: «Je sais bien que j'ai tort, qu'il y a des limites à se donner. À cette condition, l'on crée. Mais il n'y a pas de limites pour aimer et que m'importe de mal êtreindre si je peux tout embrasser<sup>31</sup>». Ce dernier extrait, le plus intéressant de tous, affirme très clairement que l'amour ne doit pas avoir de limites et que l'objet de cet amour est le «tout» — toutes les circonstances, tout ce qui se présente à nous, tout ce qui est dans le monde —, ce qui n'est pas sans rappeler l'extrait de Nietzsche plus haut. Il faut donc aussi

bien accepter l'endroit que l'envers de l'existence humaine. Sur ce point, François Chavanes n'a pas tort de décrire le jeune Camus de ce recueil en termes «de fidélité à une réalité dont il ne veut rien exclure<sup>32</sup>». Les conclusions de ces trois essais, loin de la simple coïncidence, mettent en fait le lecteur sur la piste de l'*amor fati*, au fondement de *L'envers et l'endroit*.

Dans *Ecce homo*, Nietzsche traite à un certain moment d'«une formule d'acquiescement supérieur, née de la plénitude et de la surabondance, *un oui dit sans réserve à la vie*, et même à la douleur, et même à la faute, à tout ce qu'il y a de déroutant et de problématique dans la vie<sup>33</sup>». Il est impossible de lire cet extrait sans voir une marque de l'*amor fati*. Il est d'autant plus significatif pour notre propos que Camus considère *L'envers et l'endroit* comme la manifestation de son «appétit désordonné de vivre<sup>34</sup>», ce qui n'est pas sans rappeler la citation tirée d'*Ecce homo*. Pour le philosophe franco-algérien, le grand «oui» au destin est un grand «oui» à la vie, de laquelle il faut jouir de toutes ses forces. Quel est donc le destin dont fait état *L'envers et l'endroit* et duquel nous ne pouvons pas nous séparer? C'est le monde, bien entendu, qui est inchangeable, qui me restreint et auquel je dois dire «oui». Il est possible de retrouver cette idée dans la préface de *L'envers et l'endroit*, lorsque Camus soutient ceci : «Changer la vie, oui, mais non le monde dont je faisais ma divinité<sup>35</sup>». Changer la vie, cela signifie qu'il faut changer notre attitude face au monde. Encore une fois, même s'il ne s'exprime pas en ces termes, Camus soutient la thèse de l'*amor fati*. L'idée que le monde est inchangeable et qu'il faut y consentir est entre autres illustrée dans *Amour de vivre*, où le philosophe franco-algérien affirme que «ce qui [le] frappait alors ce n'était pas un monde fait à la mesure de l'homme — mais *qui se refermait sur l'homme*<sup>36</sup>». L'homme, donc, doit se soumettre au monde. Cela n'est pas une malédiction.

### 3. *L'amor fati* : la thèse des Noces

Cette jouissance de vivre que laisse entrevoir *L'envers et l'endroit* est exemplifiée de manière encore plus frappante dans le recueil *Noces*, publié en 1939, qui comprend *Noces à Tipasa*, *Le vent à Djémila*, *L'été à Alger* et *Le désert*. Encore une fois, Nietzsche n'est

pas nommé explicitement, mais il n'en reste pas moins que *Noces* s'inscrit dans la lignée de *L'envers et l'endroit* et du thème de l'*amor fati*, du grand «oui» au destin. Dans ce recueil, la notion d'*amor fati* se précise: le destin de l'homme est de fusionner avec le monde, la nature. À ce titre, *Noces à Tipasa* est certainement le texte le plus évocateur: «Non, ce n'était pas moi qui comptais, ni le monde, mais seulement l'accord et le silence qui de lui à moi faisait naître l'amour<sup>37</sup>». Cet accord est exemplifié en tant que «joie de vivre<sup>38</sup>», en tant qu'«amour sans mesure<sup>39</sup>». Dans ce recueil, la tâche de Camus consiste ni plus ni moins à nous montrer toute la joie de vivre qui résulte d'une fusion avec le monde, d'un changement radical de notre rapport au monde, du fait que nous ne devons plus être au monde, mais que nous devons assumer d'être une partie du monde. Suivant l'enseignement des *Noces*, l'homme doit donc se considérer comme une partie du tout que constitue le monde. Sur ce point, *Noces* poursuit l'objectif de *L'envers et l'endroit*, qui faisait du monde une «divinité» et qui considérait que c'était le rapport au monde qu'il était possible de changer. Encore une fois, la poétique des essais a pour but de conforter la thèse d'une fusion avec le monde. En effet, comme l'ont noté plusieurs commentateurs, le lyrisme de *Noces* a pour tâche, d'abord de témoigner de cette expérience d'une fusion avec le tout, puis de faire adhérer le lecteur à une telle attitude en la lui faisant revivre<sup>40</sup>. Il faut aussi compter sur le fait que la nature est constamment personnifiée et l'homme chosifié, ce qui tend à donner l'impression qu'ils ne se distinguent pas<sup>41</sup>. Ainsi, c'est une thèse profondément philosophique — le «oui» à la nature — qui explique le choix de Camus pour un discours lyrique et non réflexif, le meilleur moyen de convaincre le lecteur du bien-fondé d'une telle attitude existentielle étant de témoigner de sa joie de vivre et de se restreindre au sensible. Dans *Noces*, il apparaît clairement que l'*amor fati* est un «oui» à la vie, puisque vivre consiste ouvertement selon lui à s'offrir au monde.

Mais comment est-il possible d'expliquer que *Le vent à Djémila*, le deuxième essai de *Noces*, soit aussi empreint de malheur et de peur face à la mort, face au fait qu'un jour le bonheur se terminera? Cela ne reviendrait-il pas à renoncer à une partie de son destin, à ne pas

souscrire totalement à l'*amor fati*? Alors que, dans *Noces à Tipasa*, la vie était célébrée sans aucune mesure, *Le vent à Djémila* plonge le lecteur dans une méditation lyrique sur la mort. Il ne faut pas croire que Camus refuse la condition mortelle de l'homme. Cependant, dès lors que l'on jouit de la vie et que l'on place tout son amour en elle, il n'y a qu'une posture possible face à sa mort : la terreur. Mais une telle peur à l'endroit de la mort ne constitue que l'implication logique d'un homme qui se donne tout entier à son destin, au monde, à la nature, et qui ne peut pas accepter sans peur qu'il se détachera un jour du présent duquel il jouit. Sur ce point, peut-être n'est-il pas inutile de citer *Par-delà bien et mal* : « [Celui-là aura peut-être ouvert les yeux sur l'idéal] de l'homme le plus généreux, le plus vivant et le plus affirmateur, qui ne se contente pas d'admettre et d'apprendre à supporter la réalité telle qu'elle fut et telle qu'elle est, mais qui veut la revoir *telle qu'elle fut et telle qu'elle est*, pour toute l'éternité, criant insatiablement *da capo*<sup>42</sup> ». Ce qu'il faut surtout remarquer dans cet extrait, c'est que Nietzsche, qui recycle encore un point doctrinal du stoïcisme, affirme très clairement que le grand « oui » à la vie est une acceptation de l'éternel retour<sup>43</sup>. L'éternel retour rend possible cet amour du destin, puisque cette doctrine consiste à vivre chaque moment de notre vie comme si nous allions la revivre une infinité de fois. Même si Nietzsche ne s'exprime pas ainsi, dès lors que l'homme est incapable de porter le poids de l'éternel retour, la preuve est faite qu'il ne souscrit pas à l'*amor fati*, car celui qui aime son destin ne peut que vouloir qu'il se répète inlassablement. À partir de ce lien, il est possible de mieux comprendre la reprise opérée par Camus de la thèse de l'*amor fati*. Pour Camus, aimer le destin consiste à se donner tout entier au présent, ce qui explique pourquoi la première réaction face à la mort est empreinte de terreur. « Il faut vivre maintenant », pour reprendre le titre du fameux essai de Chavanes. Toutefois, il n'affirme jamais que, confronté à la mort, il la refusera. Au contraire, Camus veut porter la « lucidité jusqu'au bout<sup>44</sup> ». Il exprime donc par cette peur que, dans le présent, lorsqu'il jouit du monde, il ne peut vouloir la mort, car ce serait le nier. Autrement dit, dans la perspective de *Noces*, l'*amor fati* est une acceptation absolue du présent. Il est bien entendu possible de retrouver cette idée dans *Le*

vent à *Djémila* lorsque le philosophe franco-algérien affirme ceci : « Si je refuse obstinément tous les “plus tard” du monde, c’est qu’il s’agit aussi bien de ne pas renoncer à ma richesse présente<sup>45</sup> ». Un peu plus loin, il affirme même que « tout ce qu’on [lui] propose s’efforce de décharger l’homme du poids de sa propre vie<sup>46</sup> ». Bien que Camus ne traite pas explicitement de l’éternel retour, il est difficile de ne pas tenter de comparer cette dernière proposition au plus célèbre passage du *Gai savoir* concernant cette doctrine :

Si cette pensée s’emparait de toi [l’éternel retour], elle te métamorphoserait, toi, tel que tu es, et, peut-être, *t’écraserait* ; la question, posée à propos de tout et de chaque chose, « veux-tu ceci encore une fois et encore d’innombrables fois ? » ferait peser sur ton agir *le poids le plus lourd* ! Oui, combien te faudrait-il aimer et toi-même et la vie pour ne plus aspirer à rien d’autre qu’à donner cette approbation et apposer ce sceau ultime et éternel<sup>47</sup> ?

Le fait que cet aphorisme s’intitule *Le poids le plus lourd* n’est pas sans rappeler la dernière proposition de Camus, selon laquelle ce qui tend à détourner l’homme de sa « richesse présente » le déchargera du *poids* de sa vie. Chez Nietzsche, comme on le constate dans le présent extrait, l’éternel retour est le poids le plus lourd, car il est la marque suprême de l’acceptation de la vie. Bien que nous n’ayons aucune preuve absolue que Camus ait lu le *Gai savoir* à l’époque de la rédaction de *Noces*, il faut tout de même admettre qu’il conçoit l’*amor fati* comme une acceptation absolue du présent, qui est en fait une acceptation du poids de sa vie. Camus n’exprime pas l’*amor fati* en les termes de l’éternel retour, mais il est impossible de ne pas considérer le fait que l’acceptation camusienne du présent remplit la même fonction que la doctrine nietzschéenne de l’éternel retour.

#### 4. La critique du christianisme

De *L’envers et l’endroit* jusqu’à *Noces*, Camus poursuit donc un projet inspiré de l’*amor fati*. Mais encore, ces deux recueils d’essais initient plusieurs critiques implicites adressées au christianisme, ce qui constitue la deuxième filiation possible entre les philosophies de

Nietzsche et du jeune Camus. Il faut préciser trois choses. D'abord, ces deux recueils ne présentent pas de critiques systématiques du christianisme, ce qui jurerait avec le but des recueils, plus personnels, loin de l'ordre du discours réflexif et systématique. Il faut plutôt les rechercher au détour d'une page, d'un aparté, d'une phrase, d'une allusion. Puis, de telles critiques s'inscrivent dans la droite lignée d'une défense de l'*amor fati* et, quoique cela soit extrêmement implicite, d'une philosophie à coups de marteau dont le but est le dénigrement des multiples dénis de la vie opérés par le christianisme. Enfin, le rapport entre Camus et le christianisme étant extrêmement complexe, il ne faut pas considérer que les quelques points exposés ici épuisent entièrement cette question difficile. Seulement est-il espéré que le traitement du christianisme par le jeune Camus soit ici quelque peu éclairé.

La première des critiques adressées au christianisme concerne son déni du monde. À la lecture de *L'envers et l'endroit*, mais surtout de *Noces à Tipasa*, il est clair que le jeune Camus enseigne la «fidélité à la terre», pour reprendre l'expression d'Arnaud Corbic<sup>48</sup>. De fait, ses écrits de jeunesse célèbrent le monde tel qu'il est, avec ses bonheurs comme avec ses malheurs. Il n'est pas inutile, sur ce point, de citer Jean Grenier, qui fut le professeur de Camus : «C'est à cette réalité [la réalité sensible] qu'il veut toujours revenir, c'est cette réalité que jamais il ne veut perdre de vue<sup>49</sup>». Il suffit de penser à *Noces à Tipasa*, où Camus soutient que «jamais [il] ne s'approcher [a] assez du monde<sup>50</sup>». Cette phrase, en apparence anodine, exprime cette fidélité à la terre dont il était question plus haut. Cela le mène à adresser plusieurs critiques contre toute pensée — notamment la doctrine chrétienne — qui réduit le monde à n'être que l'expression visible d'un monde «plus vrai». Il est possible de penser à cette proposition pleine de sous-entendus : «À Tipasa, je vois équivaut à je crois, et je ne m'obstine pas à nier ce que ma main peut toucher et mes lèvres caresser<sup>51</sup>». Il est clairement affirmé par Camus que la croyance, loin d'être dirigée vers une réalité divine et extra-sensible, la réalité de l'au-delà, doit porter sur les messages de nos sens. Si l'on ne fait plus confiance à ce que notre main et nos lèvres touchent, il n'y a plus rien. Ce n'est pas sans rappeler l'aphorisme § 124 du *Gai*

*Savoir*, où Nietzsche, par sa brillante métaphore de l'infini, élabore sa fameuse et connue critique des « arrières-mondes » et s'exclame « il n'y a plus de terre ! »

Cela m'amène à traiter de la deuxième des critiques, partie liée à la première, qui concerne le déni de la mort. Toute sa vie durant, Camus a refusé obstinément l'un des pans les plus importants du christianisme, à savoir la croyance en une vie éternelle. Pour le philosophe franco-algérien, l'espoir en une vie après la mort est la marque d'une faiblesse. Il faut notamment penser à *L'ironie*, où l'espoir en Dieu est décrit comme un besoin pour les personnes âgées dans le but de surmonter leurs angoisses liées à la maladie et à la mort<sup>52</sup>. Il n'est pas inutile de rappeler que, pour le jeune Camus, il faut obstinément refuser « les “plus tard” du monde » par attachement au présent et à notre destin. *L'été à Alger* est des plus clairs sur la question de l'espoir : « Car l'espoir, au contraire de ce qu'on croit, équivaut à la résignation. Et vivre, c'est ne pas se résigner<sup>53</sup> ». Il est clair que l'espoir en une vie éternelle équivaut à la résignation suprême, à la résignation qui se détourne complètement de la vie. Il n'est pas inutile de rappeler que, dans *Noces à Tipasa*, il voulait être lucide jusqu'au bout et affronter la mort. Pour Camus, le plus grand des malheurs consiste à espérer une autre vie, car un tel espoir est en fait un refus de notre destin, du présent, du monde.

Il faut s'intéresser à la dernière critique adressée au christianisme par le jeune Camus. En particulier dans *Noces*, le philosophe franco-algérien s'oppose à de nombreux dogmes de la morale chrétienne. D'abord, il s'oppose à la dévalorisation chrétienne du corps. En cela, il est l'héritier du Nietzsche qui voit en l'ascèse chrétienne un déni de soi<sup>54</sup>. Il faut se référer à Camus : « Mais aujourd'hui l'imbécile est roi, et j'appelle imbécile celui qui a peur de jouir<sup>55</sup> ». Or, *Noces à Tipasa* célèbre un rapport fusionnel et sensuel à la terre et en appelle à une jouissance du moment présent. Il est évidemment clair qu'une telle jouissance charnelle — pensons seulement au moment où Camus affirme que le plus grand amour consiste à « êtreindre un corps de femme<sup>56</sup> » — se dresse totalement contre les dogmes chrétiens, par voie de retour aux Grecs<sup>57</sup>. Puis, encore plus fondamentalement, il refuse les notions fondatrices de

la morale chrétienne, celles de péché et de commandement. Dans *L'été à Alger*, Camus affirme n'avoir jamais bien compris le mot « péché<sup>58</sup> ». Le lecteur attentif remarquera que, dans les deux recueils de jeunesse du philosophe franco-algérien, le seul péché est celui qui nie le corps et la vie, ce qui procède d'un retournement de la pensée chrétienne. D'ailleurs, n'est-ce pas Camus, dans *Noces à Tipasa*, qui appelait à jouir orgueilleusement de la vie, donc à vivre conformément au péché de Satan<sup>59</sup>? Ce qui l'intéresse, c'est un peuple capable de porter le poids de la vie : ce peuple, il l'a trouvé en Méditerranée, à Alger, où les hommes ne connaissent ni la vertu ni l'enfer, mais où ils sont capables de la plus grande justice et du plus grand don de soi, celui de jouir de l'existence humaine en divinisant le monde.

### *Conclusion*

Après avoir retracé les multiples influences de Nietzsche sur la philosophie du jeune Camus, il faut maintenant conclure. Lors du présent article, plutôt que d'affirmer des généralités vides de sens sur les rapports entre ces deux penseurs, il a été question de lire très minutieusement les textes de jeunesse de Camus et de se référer le plus souvent possible à de courts extraits, souvent anodins en apparence, qui démontrent hors de tout doute un nietzschéisme latent. S'il faut tenter de penser à partir de Nietzsche, j'espère que j'aurai pu montrer que le jeune Camus est bel et bien un philosophe à tendance nietzschéenne. Devant la profusion des références et des envois implicites à Nietzsche, il est impossible de considérer que Camus n'ait pas lu le célèbre philosophe allemand au début et au milieu des années 1930. D'ailleurs, la philosophie de Camus n'a jamais cessé de constituer une réponse à diverses problématiques de la pensée de Nietzsche. En effet, *Le mythe de Sisyphe* et *L'homme révolté* constituent des réponses évidentes au problème du nihilisme. D'abord, dans *Le mythe de Sisyphe*, il tente de montrer que la vie vaut la peine d'être vécue même si — et même d'autant plus si — l'existence humaine est vouée au nihilisme, au non-sens. Comme le soulève Ashley Woodward, la reprise du fameux mythe de Sisyphe par Camus rappelle sans aucun doute la doctrine de l'éternel retour et

l'acceptation d'une vision tragique de l'existence<sup>60</sup>. Même si Camus semble se détourner de la thèse de l'*amor fati* dans le *Mythe*, la conclusion de cet essai, bien qu'elle semble contredire les chapitres précédents<sup>61</sup>, renoue avec une pensée qui dit « oui », ce qui constitue le seul moyen d'atteindre le bonheur, et ce qui s'inscrit au final dans la lignée de ses œuvres de jeunesse. Cela dit, dans *L'homme révolté*, Camus se détourne plus ouvertement de Nietzsche, surtout pour des raisons politiques, car l'*amor fati* de Nietzsche ne peut pas dire « non » aux camps de concentration, au totalitarisme, à la violence. Pourtant, cet essai a aussi la particularité de renouer explicitement avec des questionnements esthétiques, dans le sillage de Nietzsche, comme le montre William E. Duvall<sup>62</sup>. Peut-être est-il de propos de terminer cet article sur une citation de Camus, tirée d'une préface qu'il a écrite pour un ouvrage de Konrad Bieber : « Je dois à Nietzsche *une partie de ce que je suis*<sup>63</sup> ».

- 
1. Je pense surtout à la biographie de Jean-Claude Brisville, publiée en 1959, qui soutenait que Camus avait découvert Nietzsche pour la première fois en 1937. Cf. J.-C. Brisville, *Camus*, Paris, Gallimard, 1959, p. 24.
  2. O. Todd, *Albert Camus. Une vie*, Paris, Gallimard, 1996, p. 69.
  3. H. R. Lottman, *Albert Camus*, Paris, Éditions du Seuil, 1978, p. 65.
  4. Cet article a paru dans la revue *Sud* en juin 1932. Cf. A. Camus, *Essai sur la musique* dans *Œuvres complètes*, Tome I, Paris, Gallimard, Bibliothèque de la Pléiade, pp. 522-540. *L'envers et l'endroit* et *Noces* seront cités également à partir de ce volume.
  5. H. R. Lottman, *op. cit.*, p. 674.
  6. Cf. A. Camus, *Carnets 1935-1948* dans *Œuvres complètes*, Tome II, pp. 795-1125 & *Carnets 1949-1959* dans *Œuvres complètes*, Tome IV, pp. 999-1307.
  7. A. Camus, *Carnets 1949-1959*, p. 1200 (je souligne).
  8. Maurice Weyembergh a noté quelques éléments de la « communauté de destin », pour reprendre son expression, qui unit Camus et Nietzsche. Cf. M. Weyembergh, « Camus et Nietzsche : évolution d'une affinité » dans R. Gay-Crosier (éd.), *Albert Camus 1980 : second international conference*, Gainesville, University Press of Florida, 1980, pp. 221-222.

9. J.-J. Brochier, *Camus, philosophe pour classes terminales*, Paris, Balland, 1979, pp. 161-167. La conclusion de ce pamphlet, sans grandes nuances, est assez claire à ce sujet.
10. É. Barilier, *Albert Camus. Philosophie et littérature*, Lausanne, L'Âge d'homme, 1977, pp. 146-147. Pour Barilier, Camus ne peut prétendre à la philosophie, de par sa nature profondément systématique. C'est là l'aspect problématique de sa thèse: toute philosophie n'est pas systématique, à commencer par celle de Nietzsche et de Camus.
11. À ce propos, il est marquant que, dans un article consacré à l'influence de la philosophie allemande sur Camus, Pierre Vaydat se concentre presque exclusivement sur la critique de Hegel à laquelle procède le philosophe franco-algérien. Évidemment, il fait quelques fois référence à Nietzsche, mais il se limite à la critique de *L'homme révolté*, sans même expliciter minimalement le rapport «souterrain» entre les deux philosophes. Cf. P. Vaydat, «Albert Camus et la philosophie allemande» dans J. Quillien (éd.), *La Réception de la philosophie allemande en France aux XIXe et XXe siècles*, Lille, Presses universitaires de Lille, 1994, pp. 219-240.
12. A. Camus, *Essai sur la musique*, p. 524.
13. Il faut remarquer que Camus s'est distancé par la suite de cette première réflexion esthétique, notamment dans la troisième partie du *Mythe de Sisyphe*.
14. Sur l'étude de la musique chez Schopenhauer, cf. L. Ferrara, «Music as the embodiment of the Will» dans D. Jacquette (éd.), *Schopenhauer; philosophy, and the arts*, Cambridge, Cambridge University Press, 1996, pp. 183-199.
15. A. Schopenhauer, *Le monde comme volonté et représentation*, trad. A. Burdeau, revue par R. Ross, Paris, PUF, 1966, p. 1011. Mon interprétation de cet élément de la théorie de Schopenhauer se trouve d'ailleurs confirmée dans son essai *Philosophie et science de la nature*. Cf. A. Schopenhauer, *Philosophie et science de la nature*, trad. A. Dietrich, Paris, Alcan, p. 107. Camus s'est certainement référé à un commentaire inadéquat ou a procédé à une lecture rapide de la page 350 de *Le monde comme volonté et représentation* lorsque Schopenhauer soutient que «c'est une même chose et un simple pléonasmе de dire la volonté de vivre que de dire la volonté tout court».
16. «Nous accorderons d'ailleurs une plus grande place à ce dernier [Nietzsche], tout d'abord parce qu'il a donné à l'Art une grande partie de son œuvre et ensuite parce que la personnalité étrange de ce

- poète-philosophe est trop attirante pour ne pas la mettre au premier plan.» Cf. A. Camus, *Essai sur la musique*, p. 524.
17. «Celui-ci [Nietzsche], en effet, élève *très respectueux* [de Schopenhauer], fut aussi un disciple *très indépendant et très indocile*.» Cf. *ibid.*, p. 528 (je souligne).
  18. *Ibid.*, p. 530.
  19. F. Nietzsche, *La naissance de la tragédie*, trad. M. Haar, P. Lacoue-Labarthe & J.-L. Nancy, Paris, Gallimard, 1977, § 1, pp. 27-32.
  20. A. Camus, *Essai sur la musique*, p. 539.
  21. R. Grenier, *Albert Camus. Soleil et ombre*, Paris, Gallimard, 1987, p. 21.
  22. A. Camus, *Le discours de Suède* dans *Œuvres complètes*, Tome IV, p. 239.
  23. Par «essai», il faut aussi entendre «tentative». Il est évident que Camus joue sur les sens du terme «essai».
  24. A. Camus, *L'envers et l'endroit*, p. 32.
  25. À ce titre, il est difficile de dire si la présence du thème de l'*amor fati* dans les écrits du jeune Camus est aussi tributaire d'une certaine lecture du stoïcisme. Les sources à ce sujet sont des plus incomplètes. Il faudrait une autre étude plus approfondie.
  26. F. Nietzsche, *Le Gai savoir*, trad. P. Wotling, Paris, GF-Flammarion, 2007, § 276, pp. 225-226 (je souligne).
  27. En 1939, Camus lisait vraisemblablement le *Crépuscule des idoles* ou un commentaire assez précis. Cf. A. Camus, *Carnets 1935-1948*, p. 889. Pour la citation dans le corpus de Nietzsche, cf. F. Nietzsche, *Crépuscule des idoles*, trad. J.-C. Hémerly, Paris, Gallimard, § 8, p. 12. Hémerly traduit plutôt cette phrase par «ce qui ne me tue pas me fortifie». Camus fait référence à la traduction d'Henri Albert.
  28. *Ibid.*, p. 9.
  29. A. Camus, *L'envers et l'endroit*, p. 54.
  30. *Ibid.*, p. 46.
  31. *Ibid.*, p. 68.
  32. F. Chavanes, *Albert Camus: «Il faut vivre maintenant»*, Paris, Éditions du Cerf, 1990, p. 37.
  33. F. Nietzsche, «Naissance de la tragédie» dans *L'Antéchrist* suivi de *Ecce homo*, trad. J.-C. Hémerly, Paris, Gallimard, 1974, § 2, p. 141 (je souligne).
  34. A. Camus, *L'envers et l'endroit*, p. 36.
  35. *Ibid.*, p. 32.
  36. *Ibid.*, p. 67 (je souligne).
  37. A. Camus, *Noces*, p. 110 (je souligne).

38. *Ibid.*, p. 108.
39. *Ibid.*, p. 107.
40. Il est possible de penser à l'analyse de Marie-Louise Audin. Cf. M.-L. Audin «La condensation furieuse de l'image ou le double lyrisme camusien» dans J. Lévi-Valensi, J. & A. Spiquel (éd.), *Camus et le lyrisme*, Paris, SEDES, 1997, p. 27.
41. Sur ce point, je renvoie au fameux article de Michèle Monte. Cf. M. Monte, «Sobriété et profusion : une rhétorique du paysage dans *Noces* et *L'été* d'Albert Camus», *Babel*, Volume 7, 2003. Une présentation de Lilliam Hernández, intitulée *Vers une poétique de Noces*, tente de montrer comment Camus arrive à exprimer cette fusion de l'homme et du monde par l'utilisation de la métaphore. Cf. L. Hernández, «Vers une poétique de *Noces*» dans R. Gay-Crosier (éd.), *Albert Camus 1980: second international conference*, pp. 142-149.
42. F. Nietzsche, *Par-delà bien et mal*, trad. C. Haim, Paris, Gallimard, 1971, § 56, pp. 82-83.
43. Sur les liens entre l'*amor fati* et l'éternel retour, cf. B. Han-Pile, «Nietzsche and Amor Fati», *European Journal of Philosophy* vol. 19, n° 2, 2009.
44. A. Camus, *Noces*, p. 115.
45. *Ibid.*, p. 113.
46. *Idem.*
47. F. Nietzsche, *Le Gai savoir*, *op. cit.*, § 341, pp. 279-280.
48. A. Corbic, *Camus et l'homme sans Dieu*, Paris, Éditions du Cerf, 2007, p. 138.
49. Pour la source de la citation, cf. F. Chavanes, *Albert Camus: «Il faut vivre maintenant»*, p. 44.
50. A. Camus, *Noces*, p. 107.
51. *Ibid.*, p. 109.
52. *Ibid.*, p. 119.
53. *Ibid.*, p. 126.
54. F. Nietzsche, *Le Gai savoir*, *op. cit.*, § 131, pp. 181-182.
55. A. Camus, *Noces*, p. 108.
56. *Ibid.*, p. 107.
57. Pour Camus, nourri d'une certaine lecture de Nietzsche, le discrédit de la chair concorde avec la chute de la Grèce. Cf. *ibid.*, pp. 118-119.
58. *Ibid.*, p. 125.
59. *Ibid.*, pp. 108-109.
60. A. Woodward, «Camus and Nihilism» dans *Sophia*, vol. 50, No.4, 2011, p. 545.

*Commentaires*

61. A. Comte-Sponville, «L'absurde dans *Le Mythe de Sisyphe*», dans A.-M. Amiot & J.-F. Mattéi (dir.), *Albert Camus et la philosophie*, Paris, PUF, 1997.
62. W.-E. Duvall, «Camus Reading Nietzsche: Rebellion, Memory and Art» dans *History of European Ideas*, vol. 1, n° 2, 1999, pp. 43-49.
63. A. Camus, *Préface à «L'Allemagne vue par les écrivains de la résistance française» de Konrad Bieber* dans *Œuvres complètes*, Tome III, p. 937 (je souligne).





# Le Cas Nietzsche : sur la tentative d'une lecture métaphysique par Heidegger

EMMANUEL CHAPUT, *Université de Montréal*

Certains auteurs comme Thomas d'Aquin, Duns Scot, Malebranche, Spinoza, Leibniz ou Wolff inscrivent d'emblée leur réflexion dans le cadre de la tradition métaphysique occidentale, quand il ne s'agit pas pour eux de la fonder. D'autres, au contraire, situent leurs réflexions en opposition directe à cette tradition dont ils entendent bien se dégager. C'est selon cette seconde guise que l'on associe habituellement la pensée de Nietzsche à un dépassement de la métaphysique. Comme le rappelle Michel Haar :

Cette question [du dépassement de la métaphysique] nous ne l'imposons pas à Nietzsche de l'extérieur. Il l'a explicitement posée: «Dépasser la métaphysique [...] est une affaire qui exige la plus haute tension de la réflexion humaine...» Et elle marque sans aucun doute tout le développement de sa pensée depuis le «renversement du platonisme» jusqu'à l'affirmation de l'Éternel Retour<sup>1</sup>.

La question se pose cependant à savoir si l'on doit effectivement situer Nietzsche au-delà de toute métaphysique. N'est-il pas au contraire lui-même un simple métaphysicien critiquant le mode de concevoir la métaphysique ayant prévalu jusqu'à aujourd'hui? En d'autres mots, doit-on nécessairement souscrire à l'interprétation qu'un penseur a de son œuvre ou ne doit-on pas plutôt chercher dans son entreprise les signes d'une appartenance à ce dont il croit justement se défaire? C'est à partir de ce genre de réflexion qui, loin de souscrire à l'interprétation de Nietzsche comme penseur d'une anti- ou d'une non-métaphysique, que Heidegger, peut-être le premier, va mettre l'accent sur le caractère systématique et métaphysique de la pensée de Nietzsche: «Que Heidegger ait révélé chez Nietzsche une cohérence véritablement métaphysique, une

profonde systématique, qui n'avaient guère été remarquées jusqu'à lui, cela est hors de doute<sup>2</sup>». Ce faisant, une compréhension nouvelle de la pensée nietzschéenne s'ouvre à nous. Nous chercherons donc à mieux saisir quel est ce caractère métaphysique qui marque la pensée de Nietzsche — et ce, malgré son projet affirmé de dépasser la métaphysique — et comment Heidegger l'inscrit comme moment charnière<sup>3</sup> dans l'histoire de la métaphysique comme histoire de l'oubli de la question de l'être. Or, comme Haar le rappelle bien, le rapport entre Heidegger et Nietzsche est certainement des plus riches, mais aussi des plus complexes. En effet, jamais un philosophe n'aura autant écrit sur un prédécesseur<sup>4</sup>. D'autre part, l'interprétation heideggérienne de Nietzsche évolue et se développe de telle manière qu'il est difficile d'en rendre véritablement compte de manière synthétique: «En effet, depuis *l'Introduction à la Métaphysique* (1935) jusqu'à *Qu'appelle-t-on penser?* (1951-52), la position de Heidegger à l'égard de Nietzsche a considérablement évolué<sup>5</sup>».

Nous nous intéresserons donc à l'interprétation de Heidegger qui fait de Nietzsche celui qui incarne le plus véritablement le destin accompli de la métaphysique occidentale aboutissant au triomphe de la technique. Nous nous attarderons toutefois par la suite aux éléments qui, chez un Heidegger plus tardif, permettent une lecture plus généreuse de l'œuvre de Nietzsche comme préfigurant, sur certains points, la position que Heidegger adoptera lui-même en opposition à la philosophie, celle de la pensée à l'écoute de l'être.

### *Les cinq éléments-clés de toute métaphysique et leur forme nietzschéenne*

Avant de s'attarder sur la lecture heideggérienne de Nietzsche comme métaphysicien, il convient d'abord de définir brièvement ce que Heidegger entend par «métaphysique». Cette question elle-même étant d'une complexité qui dépasse de loin la visée de ce texte, contentons-nous de dire que pour Heidegger toute la métaphysique se caractérise essentiellement comme l'oubli de la question de l'être au profit de l'étant: «Il est vrai que la métaphysique représente l'étant dans son être et pense ainsi l'être de l'étant. Mais elle ne pense pas la différence de l'Être et de l'étant [...]. La métaphysique ne pose pas

la question de l'Être lui-même<sup>6</sup>). Ainsi, la métaphysique se réduit elle-même à n'être qu'un ensemble d'ontologies régionales (se posant la question « quel est l'être ou l'essence de l'étant  $x$  ou  $y$ ? ») plutôt qu'à fonder une véritable ontologie fondamentale portant sur l'être lui-même.

Dans ses leçons datant de 1940 et intitulées « La métaphysique de Nietzsche<sup>7</sup> », Heidegger résume l'essence de la métaphysique d'après cinq aspects formant un tout systématique. Il y a, d'une part, 1) ce que l'étant est, son *essentia*, ce que Heidegger appelle la propriété d'être de l'étant. D'autre part, il y a 2) la manière dont cet être de l'étant existe comme totalité, son *existentia*. Comme figures métaphysiques, ces deux modes de l'être de l'étant constituent sa vérité comme non-occultation de l'essence de l'étant. Autrement dit, la métaphysique implique d'emblée 3) une conception de la *vérité* qui rend possible la révélation de l'étant dans son essence et dans son mode d'existence. Or, cette vérité est toujours pensée historiquement par une humanité. C'est pourquoi toute métaphysique est 4) *anthropologie*, puisque la vérité de l'être de l'étant est à chaque fois « structurée, fondée, communiquée et ainsi conservée<sup>8</sup> » par l'humanité qui en a le souci. Or, cette humanité s'inscrit historiquement dans le monde et situe ainsi l'être de l'étant 5) dans une entreprise *historiale*. Comme le résume Heidegger :

La propriété d'être [*Seiendheit*]: ce que l'étant est en tant que tel et la totalité de l'étant; le *fait que* l'étant est dans sa totalité et *comment* il l'est, ensuite la sorte d'essence de la vérité et l'histoire de la vérité et enfin l'humanité transférée en elle pour la conservation de la vérité circonscrivent sous cinq aspects ce en quoi l'essence unitaire de la métaphysique se déploie et se ressaisit toujours<sup>9</sup>.

C'est donc dire que l'unité de toute métaphysique se manifeste autour de ces éléments fondamentaux qui, en s'intéressant à la propriété d'être de l'étant, occultent la question de l'être en lui-même. Heidegger ne veut pas dire pour autant que la métaphysique n'ait aucune part à la vérité, tout au contraire. Simplement, en se concentrant uniquement sur l'être de l'étant, le métaphysicien oublie que « pour toute

compréhension de l'étant dans son être, l'Être lui-même s'est déjà éclairé et s'est produit dans sa vérité<sup>10</sup> ». Or, c'est cet éclaircissement véritable de l'être qu'il importe d'écouter, bien avant de s'intéresser à la vérité de l'étant.

Chez Nietzsche, les cinq aspects de la métaphysique prendront le nom de « Volonté de puissance », « nihilisme », « Éternel Retour du Même », « Surhomme » et « Justice ». Le premier terme désigne l'être de l'étant comme tel, alors que le second détermine la dimension historique de cet étant. L'Éternel Retour signifie, quant à lui, le mode d'existence de l'étant dans sa totalité. La notion de Surhomme est, selon Heidegger, ce qui désigne l'humanité responsable de penser l'être de l'étant comme Volonté de puissance, alors que la notion de Justice désigne le mode de vérité de la Volonté de puissance<sup>11</sup>.

Une des particularités de la lecture de Heidegger — qui l'oppose en ce sens aux lectures de Baeumler, Jaspers et Löwith — est donc de penser l'interdépendance de l'Éternel retour du Même et de la Volonté de puissance, alors que ces autres commentateurs de Nietzsche y voyaient plutôt deux pensées irréconciliables<sup>12</sup>.

Pour comprendre en quoi selon Heidegger, la Volonté de puissance est le caractère fondamental de l'étant, il faut d'abord se rapporter à la notion nietzschéenne de vie, terme qui, comme le rappelle Heidegger, équivaut chez Nietzsche à celui d'être<sup>13</sup>. Or, il est clair que la vie est Volonté de puissance. Tout le vivant cherche à croître et à dépasser ce qu'il est. Nietzsche s'oppose ici au concept de *conatus* entendu comme conservation de soi. Toute vie qui ne se dépasse pas, mais ne fait que se maintenir est en fait en déclin : « Tout ce qui ne dépasse pas la conservation de la vie est déjà déclin de la vie. La puissance est l'impératif du surcroît de puissance<sup>14</sup> ». Mais cette volonté d'accroître la puissance implique aussi la volonté des moyens nécessaires à sa croissance. Ainsi, la Volonté de puissance se manifeste dans la création de valeurs permettant d'accroître la puissance<sup>15</sup>. Ce faisant, c'est également la volonté de détruire toute valeur qui, loin d'accroître la puissance — ce qui fut peut-être un jour le cas, mais qui, comme valeur réifiée ne joue plus qu'un rôle de conservation — l'affaiblit au contraire. La transvaluation de toute

valeur n'est donc que l'expression de la Volonté de puissance, mais en cela, cette volonté demeure nihiliste.

En effet, le nihilisme, tel que le comprend Heidegger, n'est rien d'autre que le phénomène historique selon lequel les valeurs suprêmes se dévalorisent jusqu'à perdre toute valeur : « La ruine de ces valeurs, c'est l'effondrement de la vérité jusqu'alors [énoncée] sur l'étant en tant que tel dans sa totalité<sup>16</sup> ». Or, la Volonté de puissance, en tant qu'elle veut s'accroître, participe de ce mouvement historial nihiliste qui consiste à se dépendre des valeurs suprêmes, dans la mesure où elles font obstacle à cette croissance<sup>17</sup>. Le nihilisme consiste alors dans un double mouvement caractéristique de toute la métaphysique occidentale. D'une part, instauration de valeurs suprêmes, d'autre part, dévalorisation de ces mêmes valeurs suprêmes au nom de nouvelles valeurs plus à même d'exprimer la Volonté de puissance. Ainsi, la dévalorisation ne se fait pas au nom du néant, mais au nom de nouvelles valeurs. C'est ce mouvement constant de la Volonté de puissance qui la lie indissociablement à l'Éternel retour. En effet, la transvaluation des valeurs suprêmes ne se fait pas en vue d'une fin, d'un *telos*, aboutissement de la volonté et de la puissance. Le devenir de la Volonté de puissance ne trouve un point d'équilibre que dans la répétition : « La Volonté de puissance, en tant qu'elle se surpasse elle-même, *retourne au-dedans d'elle-même* et ainsi elle donne à l'étant dans sa totalité [...] le caractère singulier du mouvement<sup>18</sup> ». C'est ce mouvement de surpassement retournant en soi qui constitue la forme de l'Éternel retour comme totalité de l'étant. Autrement dit, si l'on s'attarde à la totalité de l'étant et non plus uniquement à un étant particulier, nous observons le retour incessant de la puissance en elle-même, au travers son propre dépassement. C'est là l'Éternel retour tel que Heidegger semble le comprendre. Or, il faut bien voir que ce Retour confère au « devenir » de la volonté la stabilité de l'être<sup>19</sup>.

Par ce mouvement incessant de dépassement et d'instauration de valeur, la Volonté de puissance comme vérité prend la forme de la Justice : « La justice est le fait de penser au sens où la Volonté de puissance institue à elle seule des valeurs<sup>20</sup> ». Évidemment, il ne faut pas entendre justice dans un sens moral, mais véritablement

métaphysique : « La justice demeure sans doute ce qui est conforme à la “rectitude”, à ce qui est juste, “droit” [...] L’essence du droit, cependant, Nietzsche la circonscrit de la sorte : “Droit = la volonté d’éterniser à chaque fois un rapport de puissance” (XIII, 205) [...] Ce vouloir ne saurait être qu’en tant que Volonté de puissance<sup>21</sup> ». La vérité est ce qui dévoile l’être de l’étant. Comme justice, la vérité exprime l’être de l’étant en tant que Volonté de puissance. Est juste et vrai ce qui rend possible l’expression de la Volonté de puissance comme dépassement de soi, comme création de nouvelles valeurs. Or, comme le rappelle Heidegger : « La vérité de l’étant en tant que tel dans sa totalité est à chaque fois assumée, structurée et préservée par une humanité<sup>22</sup> ». Trop souvent, on oublie que si l’être de l’étant se donne à comprendre, c’est qu’il se donne à quelqu’un qui est capable de le comprendre. La métaphysique doit être supportée par une intelligence capable de la saisir. C’est pourquoi l’être de l’étant est indissociable, pour Heidegger, d’une humanité qui serait à son écoute<sup>23</sup>. La prise de conscience par la métaphysique de ce rôle de l’humanité dans la saisie de l’être de l’étant va prendre, avec la modernité, la forme exacerbée de la subjectivité. Chez Nietzsche, cette subjectivité prendra la forme du Surhomme<sup>24</sup>, expression de la Volonté de puissance capable de dépasser même la forme de l’homme : « [Le Surhomme] surpasse non seulement de manière nihiliste l’être humain qui a prévalu jusqu’alors, mais du même coup en tant que l’inversion de cet être par-delà lui-même en son inconditionnalité<sup>25</sup> ». Autrement dit, le Surhomme *est* Volonté de puissance à la fois comme dépassement de l’homme, mais aussi comme entité sans détermination, sans *telos*. C’est donc dire que le Surhomme n’a pas d’existence définie — même future — mais constitue l’horizon du dépassement incessant de l’être humain ; dépassement qui prend une forme nihiliste — comme dévalorisation de l’homme — et qui inscrit le Surhomme dans la logique de l’Éternel retour. On voit ainsi, sommairement, comment Heidegger articule ce qui pour lui constitue les cinq grands axes de la métaphysique nietzschéenne en un tout systématique. La position nietzschéenne ainsi exposée, en tant qu’elle répond à la question de l’être de l’étant dans sa totalité,

constitue pour Heidegger la *position métaphysique fondamentale* de la philosophie de Nietzsche<sup>26</sup>.

*La pensée de Nietzsche comme accomplissement de la métaphysique occidentale*

Maintenant que nous avons vu comment Nietzsche s'inscrit dans la métaphysique, autrement dit, comment sa philosophie se présente comme un tout systématique qui correspond aux cinq éléments clés de toute métaphysique telle que définie par Heidegger, il reste encore à savoir comment situer Nietzsche dans cette histoire métaphysique. Pour Heidegger, il est clair que le renversement du platonisme constitue en quelque sorte la conclusion de l'histoire de la métaphysique : « Si le platonisme est la figure initiale de la philosophie, et si Nietzsche définit *sa* philosophie comme en étant le retournement, quelle issue peut encore s'ouvrir à partir de là ?<sup>27</sup>. » C'est pourquoi Nietzsche est dans l'histoire métaphysique tracée par Heidegger, le dernier philosophe et partant, le dernier métaphysicien<sup>28</sup>. Il est en effet clair pour Heidegger que si Nietzsche opère effectivement un renversement du platonisme, celui-ci demeure interne à la métaphysique<sup>29</sup>. Comme le rappelle Jean Granier :

Nietzsche accomplit bien, sans doute, un renversement du platonisme, dans la mesure où il élimine le suprasensible comme lieu de fondation des anciens idéaux, mais cette opération s'effectue toujours selon l'esprit de la métaphysique elle-même, avec ce résultat que l'essence de la métaphysique, loin d'être amenée à découvert et référée à l'histoire de l'Être, trouve ici, la condition la plus favorable à l'épanouissement des menaces qu'elle recèle<sup>30</sup>.

En quoi consiste cette inversion du platonisme qui rend possible les pires menaces de la métaphysique ? Heidegger situe cette inversion au niveau de la relation entre le sensible et le suprasensible et plus précisément encore, au niveau du rapport entre l'art et la vérité. Comme le rappelle Heidegger : « Dans la hiérarchie des différents modes de la présence de l'étant, et partant, de l'Être, l'art, selon la métaphysique platonicienne, se situe fort au-dessous de la vérité<sup>31</sup>. »

De fait, dans la *République*, Platon associe l'art à la μίμησις (*mimésis*), à l'imitation, non pas de la vérité, mais de l'apparent<sup>32</sup>. C'est dire à quel point l'art est loin de fournir un accès à l'être véritable, l'art n'est que simulacre. La beauté qui nous subjugué dans l'œuvre d'art offre une voie à l'être de l'étant que dans la mesure où elle amène notre regard à s'intéresser à de beaux objets qui, eux-mêmes, ne sont qu'autant de manières de s'intéresser à l'idée du Beau et de la Beauté. C'est ainsi que, comme le rappelle Heidegger : « la Beauté, qui pénètre dans le sensible, subjugué à partir de celui-ci, pour ravir et transporter dans l'Être<sup>33</sup> ». Or, l'art, simple imitation du sensible, semble dès lors très loin de fournir une voie privilégiée à l'être. C'est pourquoi le platonisme privilégiera face à l'art — mais pas pour autant face à la Beauté qui demeure une voie également privilégiée du platonisme<sup>34</sup> — la vérité comme mode de dévoilement de l'être : « La vérité, se déifiant du sensible est, par rapport à celui-ci, le mode qui s'en écarte au préalable, mode le plus immédiat du dévoilement de l'Être dans la pensée de la philosophie<sup>35</sup> ».

Au contraire, Nietzsche aborde la vérité de l'être à l'aide d'une approche perspectiviste selon laquelle le vrai n'est qu'une apparence permettant d'assurer la stabilité de la Vie : « La vérité, c'est-à-dire le vrai en tant que ce qui est constant, stable, est une sorte de semblant qui se justifie en tant que condition nécessaire à l'affirmation de la vie<sup>36</sup> ». La vérité est, pour Nietzsche, une erreur que l'on maintient dans la mesure où elle assure le développement de la vie. Même comme justice et droit, la vérité reste une erreur. Le juste n'est juste qu'en tant qu'il permet l'expression de la Volonté de puissance. C'est pourquoi la vérité peut être elle-même dépassée si elle devient obstacle à la vie. La vérité est ainsi subordonnée au sensible, à la vie (saisie comme création, comme Volonté de puissance). C'est comme être vivants, sensibles, que l'on en vient à adopter telle ou telle vérité parmi un ensemble de perspectives. La transvaluation des valeurs qui, pour Nietzsche, doit nous permettre de dépasser le nihilisme — bien que comme l'a montré Heidegger, cette transvaluation demeure elle-même nihiliste<sup>37</sup> — implique aussi l'abandon de nos anciennes vérités nihilistes, platoniciennes<sup>38</sup>, pour créer de nouvelles vérités susceptibles de sauver le sensible. C'est en ce sens que

Nietzsche souligne, contre le nihilisme, l'importance de l'art comme expression de la volonté de puissance<sup>39</sup> : «Contre le nihilisme il s'agit de mettre en œuvre la vie créatrice, l'art avant tout : or, l'art crée à partir du sensible<sup>40</sup>». On voit ainsi ce qu'il faut entendre par l'inversion du platonisme : à la prédominance platonicienne du couple vérité/suprasensible sur le couple art/sensible, Nietzsche oppose la prédominance du couple art/sensible sur le couple vérité/suprasensible. C'est en effet par l'art que la Volonté de puissance est appelée à transvaluer les anciennes valeurs pour en créer de nouvelles et ainsi liquider le monde suprasensible des anciennes valeurs pour imposer à l'inverse le monde sensible comme lieu privilégié de l'affirmation de la vie entendue comme Volonté de puissance.

Or, pour Heidegger, malgré ce renversement, Nietzsche demeure prisonnier de la métaphysique platonicienne : «Après l'inversion — ce qui, du point de vue formel, est facile à évaluer — le sensible, le monde apparent, se situe *en haut*, le suprasensible, le monde vrai, *en bas* [...] Mais tant que ce *haut* et ce *bas* déterminent la structure de l'édifice platonicien, il subsiste dans son essence<sup>41</sup>».

En fait, en pensant la Volonté de puissance comme l'essence de l'être de l'étant, Nietzsche représente, selon Heidegger, l'apogée de la métaphysique de la volonté qui se développe à partir de Leibniz (avec son concept d'*impetus*) et dans la postérité de Kant chez des penseurs comme Fichte, Schelling, Hegel, mais également chez Schopenhauer<sup>42</sup>. Sans dépendre directement de ces penseurs, la pensée de Nietzsche s'inscrit néanmoins dans une lignée historique et métaphysique qui n'a rien d'arbitraire : «Cette référence aux ancêtres spirituels de Nietzsche relativement à la doctrine qui caractérise l'Être en tant que volonté ne prétend pas établir une quelconque dépendance, mais indique simplement que semblable doctrine n'en est pas une arbitraire, dans la métaphysique occidentale, mais peut-être même nécessaire<sup>43</sup>».

De fait, Heidegger conçoit le développement historique de la métaphysique depuis Platon comme une fatalité qui marque le cours de l'Europe occidentale et qui trouve son aboutissement dans l'époque de la technique. Ainsi, la fin de la métaphysique, comme achèvement, s'avère être en réalité sa réalisation pleine et entière.

Or, cette période qui achève la métaphysique, c'est la période durant laquelle l'être est conçu en termes de volonté. C'est ainsi que pour Heidegger : «L'achèvement de la métaphysique commence avec la métaphysique hégélienne du savoir absolu entendu comme esprit de la volonté<sup>44</sup>». L'être *comme volonté* se présente d'abord comme volonté d'*autre chose*, mais finit par se vouloir lui-même *en tant que volonté*. C'est ainsi qu'en définitive, la métaphysique de la volonté prend la forme de la volonté de la volonté. Or, cette volonté de la volonté nous dit Heidegger n'est rien d'autre que la technique, ultime oubli de l'être au profit de l'étant :

La forme fondamentale sous laquelle la volonté de volonté apparaît et, en calculant, s'installe elle-même dans la non-historicité du monde de la métaphysique achevée peut être appelée d'un mot : la «technique». Cette appellation englobe alors tous les domaines de l'étant, qui forment à chaque instant l'équipement du tout de l'étant : la nature objectivée, la culture maintenue en mouvement, la politique dirigée, les idéals surhaussés [...] Nous prenons ici «la technique» en un sens si essentiel qu'il équivaut à celui de «la métaphysique achevée»<sup>45</sup>.

Pour Heidegger, il est clair que Nietzsche joue un rôle déterminant dans l'accomplissement de la métaphysique comme technique : «Dans la volonté de puissance, la métaphysique de Nietzsche fait apparaître l'*avant*-dernière étape du processus par lequel l'étantité de l'étant exerce et déploie sa volonté comme volonté de volonté<sup>46</sup>». Comment interpréter la volonté de puissance comme volonté de volonté ? On peut le comprendre à la manière dont Heidegger interprète la dernière phrase de *La Généalogie de la morale* : «L'homme préfère encore avoir la volonté du *néant* que de ne *point* vouloir du tout...<sup>47</sup>.» La Volonté de puissance préfère vouloir le néant que de s'abolir comme volonté, elle n'aspire alors plus à rien d'autre qu'elle-même comme volonté.

Or, Nietzsche ne fait-il que penser métaphysiquement notre époque, celle de la technique, comme le suggère Heidegger lorsqu'il affirme : «L'«aujourd'hui» [...] se détermine à partir du

temps propre à l'histoire de la métaphysique : c'est la détermination métaphysique de l'humanité historique à l'époque de la métaphysique de Nietzsche<sup>48</sup>». Heidegger semble en effet le penser lorsqu'il définit le Surhomme comme : « l'espèce de l'homme inversé au sens nihiliste<sup>49</sup> ». Or, cette inversion, selon Heidegger, loin d'endiguer la conception classique de l'homme, n'est, dans les faits, rien d'autre que l'accentuation de la volonté humaine visant la domination de la Terre<sup>50</sup>. Domination déjà en germe dans l'histoire de la métaphysique, mais qui trouve sa pleine expression dans le règne de la machine et de la technique. Nietzsche le premier aurait reconnu « le caractère métaphysique de la machine<sup>51</sup> » dans son aphorisme « La machine en tant qu'enseignante<sup>52</sup> ».

De fait, dans son *Dépassement de la métaphysique*, Heidegger souligne à quel point la métaphysique de la Volonté de puissance de Nietzsche exprime l'essentiel des valeurs de la technique moderne. L'être de l'étant comme Volonté de puissance trouve son affirmation primordiale dans l'art créatif. C'est à travers l'art créateur de nouvelles valeurs que s'exprime à la fois la Volonté de puissance et le projet nietzschéen d'une transvaluation de toutes les valeurs. Or, ce projet est l'expression — peut-être la plus juste — à la fois de la modernité (dans son opposition aux anciennes valeurs traditionnelles) et du productivisme technique (la Volonté de puissance est essentiellement productive, s'exprimant d'abord et avant tout dans la création artistique)<sup>53</sup>. Heidegger dira à cet effet :

Dans le concept de volonté de puissance, les deux « valeurs » constitutives (la vérité et l'art) ne sont que des appellations détournées, d'une part pour la « technique », au sens essentiel de ce travail efficace qui, par plans et calculs, constitue des fonds et, d'autre part, pour la production des « créateurs » qui, dépassant la vie de leur époque, fournissent à la vie un nouveau stimulant et assurent le mouvement culturel<sup>54</sup>.

À l'opposé, pour Heidegger, la possibilité de se mettre à l'écoute de l'être doit passer par un certain moment contemplatif qui n'a rien à voir avec l'incessante création du Surhomme nietzschéen constamment prisonnier de la production et donc, toujours à côté du

sens, à l'abandon loin de l'être. La frénétique création du Surhomme, véritable expression de sa Volonté de puissance se retournant constamment en elle-même dans un retour éternel, s'inscrit parfaitement, à l'époque de la technique, dans l'image du moteur : « Qu'est-ce que l'essence du moderne "moteur", sinon *une* nouvelle forme du Retour éternel de l'Identique ?<sup>55</sup>. »

En rester à cette image d'un Nietzsche penseur de la technique, c'est toutefois oublier que le dialogue entre Heidegger et Nietzsche n'est pas un dialogue de sourds, mais un dialogue en développement<sup>56</sup>. Ainsi, si nous nous limitons à la figure du Surhomme nietzschéen, on assiste, dans *Qu'appelle-t-on penser?*, à une transformation radicale de sa compréhension par Heidegger. En effet, s'il en traçait préalablement, comme nous l'avons vu dans *Dépassement de la métaphysique*, un portrait définitivement inscrit dans la logique de la technique et de la modernité, ce ne semble plus être le cas dans le cours de 1951-1952 intitulé *Qu'appelle-t-on penser?*. Au contraire, dans ce dernier ouvrage, Heidegger rapproche le Surhomme nietzschéen du berger de l'être<sup>57</sup>, figure heideggérienne par excellence de celui qui est à l'écoute de l'être<sup>58</sup>. En effet, on ne peut ignorer le rapprochement lorsque Heidegger affirme : « Pour le Surhomme la perte de la mesure, le progrès incessant, en tant que catégorie purement quantitative, est précisément une question qui tombe. Le sur-homme est plus pauvre, plus simple, plus tendre, plus dur, plus calme, plus généreux, plus lent dans ses décisions et plus économe dans sa langue<sup>59</sup> ». On ne saurait être plus loin de la logique de domination à l'œuvre dans la métaphysique achevée comme technique.

### *Nietzsche, précurseur de la pensée de l'Être ?*

« Maintenant seulement — après la suppression de la distinction qui porte toute la métaphysique occidentale — commence Zarathoustra<sup>60</sup> ». Si Nietzsche est effectivement le dernier métaphysicien qui, tout en renversant le platonisme en maintient l'essence, il n'en reste pas moins que ce renversement ne laisse pas intacte l'essence de la métaphysique. On peut même penser qu'à travers ce renversement de la métaphysique, Nietzsche ouvre certaines brèches dont il reviendra à Heidegger de commencer

à penser la portée ontologique<sup>61</sup>. Peut-on voir en germe, dans la tentative nietzschéenne de dépasser la métaphysique, certains éléments — par exemple le Surhomme tel que Heidegger l'interprète dans *Qu'appelle-t-on penser?* — préfigurant cette pensée dont Heidegger parle à la fin de sa *Lettre sur l'humanisme*<sup>62</sup>? Il est certain que pour Heidegger le renversement opéré par Nietzsche n'est pas assez radical en ce qu'il maintient la hiérarchie entre sensible et suprasensible, en la renversant. Cependant, son impasse constitue en quelque sorte l'indication d'un autre chemin, voie d'accès à la pensée non-philosophique dont Heidegger parle à la fin de sa *Lettre*<sup>63</sup>. D'ailleurs, comme le rappelle Heidegger: «pour autant qu'il [Nietzsche] se tient lui-même dans sa propre pensée, il est déjà au-delà de ce qu'il pense, il est toujours plus loin qu'il ne le sait et ne le soupçonne lui-même<sup>64</sup>». Nietzsche constituerait ainsi un «pont<sup>65</sup>» entre la métaphysique et la pensée à l'écoute de l'être telle que prônée par Heidegger. En effet, ce pont que Nietzsche associe à l'homme en tant qu'il doit être dépassé pour mener à de nouveaux horizons, Heidegger l'associe à l'époque actuelle, «transition entre la fin de la métaphysique et “un nouveau commencement” de la pensée<sup>66</sup>», comme le rappelle Haar. Cette époque, selon Heidegger, c'est à Nietzsche qu'il est revenu de la penser. Autrement dit, l'époque qui est à la fois l'accomplissement de la métaphysique et l'ouverture à la possibilité d'une pensée à l'écoute de l'être est l'époque de Nietzsche<sup>67</sup>. C'est à partir de Nietzsche que l'on peut commencer à penser cette époque et ses possibilités de voilement ou de dévoilement de l'être.

Dès lors, Nietzsche, penseur de l'époque charnière entre métaphysique et pensée de l'être, n'est-il pas un élément indispensable dans une réflexion plus approfondie du rapport entre la pensée de Heidegger et la métaphysique? Comment concevoir le rapport entre pensée et métaphysique sans s'attarder à Nietzsche qui constitue pour Heidegger un point focal où ces deux approches de l'être se rencontrent?

### *Conclusion*

Par sa lecture de l'œuvre nietzschéenne, Heidegger dévoile ainsi la systématisme d'une pensée trop souvent associée à une simple

poésie (négligeant ainsi la dimension métaphysique de cette pensée). D'ailleurs, loin de faire de Nietzsche un métaphysicien parmi d'autres, Heidegger lui attribue le rôle historique d'un renversement du platonisme aboutissant à l'accomplissement de la métaphysique dans la figure historique de la technique. Ce faisant, Nietzsche constitue, pour Heidegger, la dernière figure du philosophe qui ouvre cependant — par certaines intuitions dont la tâche est revenue à Heidegger de les penser dans leurs portées — à une pensée à l'écoute de l'être qui s'inscrit, toujours selon Heidegger, en faux à la métaphysique<sup>68</sup>. Or, en quoi cette pensée heideggérienne de l'être n'est-elle pas métaphysique? Autrement dit, comment comprendre en quoi la pensée à l'écoute de l'être pratiquée par le poète et le penseur n'est pas, en un certain sens, une nouvelle tentative métaphysique de saisir le sens de l'être? Certes, cette pensée ne tombe pas dans ce que Heidegger appellerait l'erreur de la métaphysique classique consistant à vouloir dominer l'être en le déterminant par des principes rationnels. La pensée de l'être s'extrait-elle pour autant de l'entreprise métaphysique entendue comme découverte du sens de l'être? Alors que Heidegger récuse la prétention nietzschéenne à une sortie hors de la métaphysique, en quoi sa propre entreprise peut-elle prétendre échapper à ce même soupçon? La pensée à l'écoute de l'être est-elle non-métaphysique voire non-philosophique? C'est là une affirmation qui, si elle doit être mise à l'épreuve, ne peut l'être qu'à partir d'un approfondissement de la lecture heideggérienne de l'œuvre nietzschéenne, comprise à la fois comme accomplissement de la métaphysique et comme pré-ouverture à la pensée de l'être. Nietzsche comme figure faisant le pont entre métaphysique et pensée de l'être constitue en effet un point de départ privilégié à toute tentative visant à relier et à relire la pensée de l'être en des termes métaphysiques.

- 
1. Michel Haar, *Nietzsche et la métaphysique*, Paris, Gallimard, 2007, p. 7.
  2. Michel Haar, *La fracture de l'Histoire*, Paris, Millon, 1994, p. 202. Dans le même ordre d'idée, Olivier Huot-Beaulieu souligne dans son mémoire: «Heidegger parvient à coordonner de manière systématique et cohérente l'ensemble des grands thèmes de la pensée métaphysique

- de Nietzsche, tâche que personne n'avait accomplie avant lui.» (Olivier Huot-Beaulieu, *Nietzsche et le tournant dans la pensée de Martin Heidegger : Examen d'une thèse de Hannah Arendt*, Montréal, Université de Montréal, 2007, pp.37-38).
3. Dans la mesure où Heidegger affirme : «Nietzsche, le penseur de la pensée de la Volonté de puissance, est le *dernier métaphysicien* de l'Occident.» (Martin Heidegger, *Nietzsche I*, Paris, Gallimard, 1990, p.374). Affirmation d'autant plus importante que pour Heidegger la métaphysique est le propre de l'Occident (cf. Martin Heidegger, «Dépassement de la métaphysique» dans *Essais et conférence*, Paris, Gallimard, 2010, p.88), ce qui revient à faire de Nietzsche «*dernier métaphysicien*» tout court.
  4. Cf. Michel Haar, *La fracture de l'Histoire*, *op. cit.*, pp.189-190.
  5. *Ibid.*, p.199. À titre d'exemple, dans son texte «L'impensé ambivalent du surhomme», Michel Haar rappelle qu'il existe, chez Heidegger, environ huit lectures différentes du Surhomme nietzschéen entre 1935 et 1953 (Cf. *ibid.*, pp.160-162).
  6. Martin Heidegger, *Lettre sur l'humanisme*, Paris, Aubier-Montaigne, 1957, pp.49-51.
  7. Cf. Martin Heidegger, *Nietzsche II*, Paris, Gallimard, 1990, pp.205-266. C'est d'ailleurs essentiellement en s'inspirant de ce texte que nous exposerons la lecture heideggérienne de la métaphysique de Nietzsche. L'étude comparative consistant à analyser, dans les multiples leçons qu'Heidegger fit sur la métaphysique de Nietzsche, les éléments récurrents ou, au contraire, différents, dépasse malheureusement le cadre de cette étude.
  8. *Ibid.*, p.207.
  9. *Ibid.*, p.208.
  10. Martin Heidegger, *Lettre sur l'humanisme*, *op. cit.*, p.51.
  11. Heidegger le résume ainsi : «“La Volonté de puissance” désigne l'être de l'étant en tant que tel, l'*essentia* de l'étant. “Nihilisme” est le terme pour l'histoire de la vérité ainsi déterminé. L'“Éternel Retour du Même” exprime la manière dont l'étant dans sa totalité est l'*existentia* de l'étant. Le “Surhomme” caractérise cette humanité qui est requise par cette totalité. “Justice” est l'essence de la vérité de l'étant en tant que Volonté de puissance.» (Martin Heidegger, *Nietzsche II*, *op. cit.*, p.209).
  12. Cf. Martin Heidegger, *Nietzsche I*, *op. cit.*, pp.28-29. Néanmoins, il faut souligner que selon les développements de la compréhension de Nietzsche par Heidegger, l'emphase sera tantôt mise davantage sur la notion de Volonté de puissance, tantôt sur la pensée du Retour.

13. Cf. Martin Heidegger. *Nietzsche II*, op. cit., p. 213.
14. *Ibid.*, p. 215.
15. Cf. *ibid.*, pp. 218-219.
16. *Ibid.*, p. 221.
17. Cf. *ibid.*, p. 223.
18. *Ibid.*, pp. 228-229.
19. «*Que toutes choses reviennent, c'est l'extrême approximation d'un monde du devenir à celui de l'Être*» (*Ibid.*, p. 231).
20. *Ibid.*, p. 257.
21. *Ibid.*, p. 260.
22. *Ibid.*, p. 234.
23. Évidemment, cette écoute peut prendre plusieurs formes plus ou moins contemplatives ou agressives. Par exemple lorsque l'on *questionne* l'être de l'étant, il ne faut pas oublier le sens que «question» pouvait prendre, par exemple sous l'inquisition : c'est une écoute qui provoque une réponse et qui en un certain sens n'entend que les aveux qu'elle veut bien entendre.
24. «*La subjectivité achevée de la Volonté de puissance est à l'origine métaphysique de la nécessité d'essence du "surhomme".*» (*Ibid.*, p. 242.)
25. *Ibid.*, p. 243.
26. «A fundamental metaphysical position is characterized by the fact that it answers what Heidegger calls the guiding question of philosophy (What are beings?) in a specific and characteristic way. However, generally, the grounding question of philosophy (What is Being?) does not get asked and is even occluded in the fundamental metaphysical positions preceding Heidegger's own philosophy.» (Paul Catanu, *Heidegger's Nietzsche — Being and Becoming*, Montreal, 8th House Publishing, 2010, pp. 26-27). Cf. Martin Heidegger, *Nietzsche II*, op. cit., p. 208; Olivier Huot-Beaulieu, *Nietzsche et le tournant dans la pensée de Martin Heidegger: Examen d'une thèse de Hannah Arendt*, op. cit., pp. 38-39.
27. Frédéric De Towarnicki, «Le Nietzsche de Heidegger, entretien avec Jean Baufret», *Magazine littéraire*, n° 141, 1978, p. 26.
28. Comme le rappelle Baufret, philosophe et métaphysicien sont, pour ainsi dire, synonymes dans la pensée de Heidegger : «car philosophie et métaphysique, c'est tout comme, sauf que philosophie est un terme plus ancien.» (*Ibid.*)
29. Cf. Martin Heidegger, *Lettre sur l'humanisme*, op. cit., p. 93.

30. Jean Granier, *Le Problème de la Vérité dans la philosophie de Nietzsche*, Paris, Éditions du Seuil, 1966, p. 613.
31. Martin Heidegger, *Nietzsche I, op. cit.*, p. 171.
32. Cf. Platon, *La République*, 598b.
33. Martin Heidegger, *Nietzsche I, op. cit.*, p. 181.
34. Néanmoins, comme le rappelle Catanu, cette voie, certes privilégiée pour atteindre le suprasensible à partir de la beauté sensible, devient seconde dès lors que l'homme accède au suprasensible où se situe la vérité (cf. Paul Catanu, *Heidegger's Nietzsche — Being and Becoming, op. cit.*, p. 20).
35. Martin Heidegger, *Nietzsche I, op. cit.*, p. 181.
36. *Ibid.*, p. 193.
37. « C'est cette transvaluation qui constitue l'essence accomplie du nihilisme » (Martin Heidegger, *Nietzsche II, op. cit.*, p. 221).
38. Nietzsche pose en effet une équivalence entre le nihilisme et le platonisme comme le rappelle Heidegger : « le nihilisme, c'est-à-dire le Platonisme, pose le suprasensible en tant que le véritablement étant, en tant que ce qui participe véritablement à Être, à partir duquel tout autre étant est avalé, calomnié et déclaré nul, en tant que le proprement non-étant. » (Martin Heidegger, *Nietzsche I, op. cit.*, p. 131).
39. Cf. *ibid.*, pp. 72-73.
40. *Ibid.*, p. 148.
41. *Ibid.*, p. 182. Heidegger dira ailleurs : « Le renversement du platonisme, renversement suivant lequel les choses sensibles deviennent pour Nietzsche le monde vrai et les choses suprasensibles le monde illusoire, reste entièrement à l'intérieur de la métaphysique. » (Martin Heidegger, « Dépassement de la métaphysique » dans *Essais et conférence, op. cit.*, p. 91).
42. Cf. Martin Heidegger. *Qu'appelle-t-on penser*, Paris, Presses universitaires de France, 1967, p. 125. Cette interprétation heideggérienne de Nietzsche comme métaphysicien de la volonté qui ferait, à l'instar de Schelling, de la volonté l'être originaire de l'étant n'est toutefois pas exempt de problèmes d'un point de vue philologique, comme le souligne Paul Catanu (cf. Paul Catanu, *Heidegger's Nietzsche — Being and Becoming, op. cit.*, pp. 35-36).
43. Martin Heidegger. *Nietzsche I, op. cit.*, p. 40.
44. Martin Heidegger. « Dépassement de la métaphysique » dans *Essais et conférence, op. cit.*, p. 86.
45. *Ibid.*, p. 92.

46. *Ibid.*, p. 93. La dernière étape étant évidemment ce que Heidegger nomme la technique où la domination de l'étant par le calcul et la planification devient telle que la possibilité même d'un questionnement fondamental de l'être risque de disparaître de l'horizon de la réflexion humaine. La technique comme manifestation de l'être de l'étant risque de condamner l'homme, uniquement attentif à l'étant et non à l'être, à un «abandon loin de l'être» (Cf. *ibid.*, p. 105). C'est pourquoi Heidegger nous invite à faire un pas de côté afin de nous mettre à l'écoute de l'être plutôt que de chercher à le dominer sous sa forme technique : «La pensée ne surmonte pas la métaphysique en la dépassant, c'est-à-dire en montant plus haut encore pour l'accomplir on ne sait où, mais en redescendant jusqu'à la proximité du plus proche.» (Martin Heidegger, *Lettre sur l'humanisme*, *op. cit.*, p. 133).
47. Friedrich Nietzsche, *Généalogie de la morale*, Paris, Gallimard, 1969, p. 246.
48. Martin Heidegger, *Nietzsche II*, *op. cit.*, pp. 201-202. Comme le rappelle Catanu, pour Heidegger : «Nietzsche's metaphysics is both the last metaphysics of the West and the consummation of metaphysics into the essence of modern technology.» (Paul Catanu, *Heidegger's Nietzsche — Being and Becoming*, *op. cit.*, p. 15).
49. Martin Heidegger, *Nietzsche II*, *op. cit.*, p. 247.
50. Cf. *idem*.
51. *Idem*.
52. Cf. Friedrich Nietzsche, *Humain, trop humain II*, Paris, Gallimard, 2007, p. 276.
53. Martin Heidegger, «Dépassement de la métaphysique» dans *Essais et conférence*, *op. cit.*, p. 94.
54. *Ibid.*
55. Martin Heidegger, «Note sur le retour éternel de l'identique» in *Essais et conférence*, *op. cit.*, p. 147.
56. Comme le rappelle Paul Catanu : «there are notable developments across the [1936-1946's] lectures. Heidegger's position with respect to Nietzsche is not fully uniform and univocal in those lectures, but rather evolutive and progressive. Furthermore, Heidegger's position also evolves across the texts *Who is Nietzsche's Zarathustra*, *The Word of Nietzsche*, *God is Dead* and *What is Called Thinking* to name just a few of the later writing in which Heidegger still attempts to come to terms with Nietzsche's philosophy.» (Paul Catanu, *Heidegger's Nietzsche — Being and Becoming*, *op. cit.*, p. 364).

57. Dans sa *Lettre sur l'humanisme*, Heidegger propose la figure de l'homme comme berger de l'Être afin de l'opposer à l'homme qui, prisonnier de la technique, se conçoit comme le maître de l'étant : « L'homme n'est pas le maître de l'étant, il est le berger de l'Être. Dans ce "moins", l'homme ne perd rien, mais il gagne bien plutôt, dans la mesure où il parvient à la vérité de l'Être. Il gagne l'essentielle pauvreté du berger dont la dignité repose en ceci : être appelé par l'Être lui-même à la sauvegarde de sa vérité » (Martin Heidegger, *Lettre sur l'humanisme*, *op. cit.*, p. 105).
58. Sur ce rapprochement, cf. Michel Haar, *La fracture de l'Histoire*, *op. cit.*, p. 161.
59. Martin Heidegger, *Qu'appelle-t-on penser*, *op. cit.*, p. 111.
60. Martin Heidegger, *Nietzsche I*, *op. cit.*, p. 489.
61. « La pensée jusqu'ici connue est métaphysique et il est à présumer que la pensée de Nietzsche réalise son achèvement. Par là apparaît dans la pensée de Nietzsche quelque chose que cette pensée elle-même ne peut plus penser. Qu'une pensée reste en arrière de ce qu'elle pense caractérise ce qu'elle a de créateur. Et là même où une pensée conduit la métaphysique à son achèvement, elle indique, à la façon d'un cas privilégié, quelque chose qu'elle ne pense pas et qui est à la fois clair et confus. Mais où sont les yeux pour le voir ? » (Martin Heidegger, « Qui est le Zarathoustra de Nietzsche » in *Essais et conférence*, *op. cit.*, p. 142) (nous soulignons). Ce qu'il importe de voir ici, c'est comment ce que Nietzsche lui-même n'a pu penser est tout de même dans sa pensée ; autrement dit, comment l'achèvement de la métaphysique opérée par Nietzsche — au dire de Heidegger — constitue également le point de départ d'une nouvelle perspective — peut-être non-métaphysique si l'on en croit Heidegger — sur l'Être.
62. Cf. Martin Heidegger, *Lettre sur l'humanisme*, *op. cit.*, p. 167.
63. Cf. Martin Heidegger, *Nietzsche I*, *op. cit.*, pp. 189-190. C'est ainsi que pour Haar : « Heidegger admet que le renversement opéré par Nietzsche produit une métamorphose du schéma platonicien. Le renversement des valeurs devrait, dit-il, "frayer la voie à une interprétation nouvelle du sensible et du non-sensible". » (Michel Haar, *La fracture de l'Histoire*, *op. cit.*, p. 201).
64. Martin Heidegger, *Nietzsche I*, *op. cit.*, p. 250.
65. Sur cette expression cf. Michel Haar, *La fracture de l'Histoire*, *op. cit.*, p. 199.
66. *Ibid.*
67. Il serait légitime de se demander si cette époque n'est pas aussi celle de Heidegger lui-même.

68. Comme le rappelle Baufret : «Nietzsche est bien le dernier philosophe, soit qu'il ouvre un renouveau qui sera le stade définitif de la philosophie ; soit que sa philosophie prélude à un renouveau qui sera plus radical que la philosophie elle-même. Ce renouveau sera, selon Heidegger, le passage de la question de l'être à la question du *sens* de l'être, que même Nietzsche ne soupçonne pas encore, bien que sa philosophie, comme ultime ressource de la philosophie, en soit à sa façon la pré-apparition, et comme "l'aube grise" évoquée dans le *Crépuscule des idoles*.» (Frédéric De Towarnicki, «Le Nietzsche de Heidegger, entretien avec Jean Baufret», *loc. cit.*, p. 26).





# La complémentarité des formules de l'impératif catégorique

SIMON PELLETIER, *Université Laval*

À la fin de sa *Critique de la raison pure*, Kant dit admettre l'existence de lois pratiques pures, c'est-à-dire de lois morales produites par la raison pure commandant d'une manière absolue « ce qu'il faut faire et ne pas faire<sup>1</sup> », déterminant complètement *a priori* l'usage que tout être raisonnable doit faire de sa liberté. Le projet d'ériger une métaphysique des mœurs, dont l'objet serait l'étude de ces lois objectives de la liberté, requiert qu'on en pose d'abord les fondements en cherchant et en établissant le principe suprême et *a priori* de la morale loin de tout ce qui pourrait être emprunté à l'expérience. C'est ce que fera Kant dans les deux premières parties des *Fondements de la métaphysique des mœurs*<sup>2</sup>. La méthode qu'il emploie pour le faire doit attirer notre attention sur un fait bien particulier. Kant procède de manière *analytique*, c'est-à-dire en partant de la connaissance commune de la moralité par une série de concepts compris et imbriqués les uns dans les autres jusqu'à la découverte de l'impératif catégorique et l'exposition de ses multiples formules. La méthode utilisée par Kant implique que toutes les formules de l'impératif catégorique peuvent être tirées de la simple analyse de ce que contient et implique son concept; elle signifie, en d'autres mots, que toutes ses formules renvoient les unes aux autres, que, loin de s'ajouter les unes aux autres pour former un simple agrégat de prescriptions diverses, elles sont au contraire si intimement connectées qu'elles forment un tout organique, qu'elles ne font qu'un<sup>3</sup>. La méthode employée par Kant pour mener l'examen du principe suprême et *a priori* de la morale doit donc être prise en compte lorsque nous tentons d'en développer une interprétation satisfaisante et une compréhension approfondie.

Cet article se donne pour but de rendre compte du déploiement dynamique du concept d'impératif catégorique qui s'effectue au fil

de l'analyse présentée dans les première et deuxième sections des *Fondements de la métaphysique des mœurs*, et ce, pour mettre en évidence la complémentarité de ses formules et leur cohérence d'ensemble. Nous voulons, avec cette approche, prendre le contre-pied des démarches qui s'apparentent à celle qu'Alexis Philonenko a entreprise dans son *Œuvre de Kant*<sup>4</sup>, qui se contentent généralement d'empiler les différentes formules de l'impératif catégorique les unes sur les autres, en mentionnant leur unité sans la montrer. Cette étude se fera en deux temps : d'abord elle livrera l'essentiel de la démarche qui mène Kant à exprimer pour une première fois à l'aide d'une formule le contenu de l'impératif catégorique, car elle s'avère nécessaire, comme nous le verrons, pour comprendre comment il parviendra à l'exprimer par ses autres formules. Ensuite, elle rendra visible la manière dont chacune des autres formules peut être dérivée des implications de cette étude préalable du concept de l'impératif catégorique, tout en exposant les multiples connexions par lesquelles elles se lient les unes aux autres.

*Première partie : Vers une première formulation de l'impératif catégorique*

Jetons un bref regard sur les résultats de l'étude menée dans la première partie des *Fondements de la métaphysique des mœurs*, qui se proposait d'opérer un passage de la connaissance rationnelle commune de la moralité à la connaissance philosophique. Kant était parti de la notion de bonne volonté, conçue dans la conscience populaire comme le seul bien inconditionnellement moral, pour progresser vers le concept du devoir, qui contient celui de bonne volonté<sup>5</sup>. La bonne volonté est celle qui reconnaît pour elle-même des obligations. Le devoir est, en quelque sorte, le signe de notre finitude ; ses exigences se confrontent à la résistance de notre nature, ce pour quoi elles se présentent comme des contraintes. Agir par devoir signifie agir de telle sorte que ne soient pas déterminants pour notre action les multiples mobiles que nous propose notre sensibilité. Ni les buts, ni les fins ou encore les effets que la volonté se propose de réaliser par une action ne communiquent donc à celle-ci de valeur morale. La valeur morale d'une action, autrement

dit, n'est aucunement fonction des mobiles matériels et *a posteriori* de la volonté. Le principe qui détermine une bonne volonté, ce sont les exigences de la morale elles-mêmes, de sorte qu'en plus d'une conformité extérieure au devoir, il existe un « accord interne<sup>6</sup> » entre le mobile de l'action commise et le devoir.

La bonne volonté, selon la distinction devenue classique de Kant, agit non seulement *conformément* au devoir, mais encore *par* devoir. Pour comprendre ce point dans toute sa profondeur, il s'avère nécessaire de développer la notion de *respect*, intimement connectée à celle de *loi*, car, relève l'analyse, le devoir est « la nécessité d'accomplir une action par respect pour la loi<sup>7</sup> ».

Qu'est-ce donc qui détermine une bonne volonté à poser une action morale, à agir par devoir? Objectivement, c'est la loi morale; mais subjectivement, relève l'examen, c'est le *respect*. Le respect est le seul mobile approprié à une volonté qui agit par devoir. Le respect est certes un sentiment, mais il n'emprunte étrangement rien à la sensibilité ni n'est produit par elle. En effet, nous ne respectons en aucune manière ce que nous désirons et ce que nous visons à produire par nos actions, et nous ne pouvons pas non plus avoir du respect pour nos inclinations elles-mêmes. « Il n'y a que ce qui est lié à ma volonté uniquement comme principe et jamais comme effet, ce qui ne sert pas mon inclination, mais qui la domine, ce qui du moins empêche entièrement qu'on en tienne compte dans la décision, par suite la simple loi pour elle-même, qui puisse être un objet de respect et par conséquent être un commandement<sup>8</sup> ». C'est un concept de la raison, une idée pure, qui produit spontanément le respect. Le respect ne porte jamais sur une chose, mais seulement sur la représentation d'une valeur qui nous humilie tout en nous révélant simultanément notre dignité d'êtres libres<sup>9</sup>: la loi morale et « *sa majesté solennelle*<sup>10</sup> ». Si donc le respect a pour *objet* la loi, il est aussi à proprement parler *l'effet* de la loi sur le sujet: « Ce que je reconnais immédiatement comme loi pour moi, je le reconnais avec un sentiment de respect qui exprime simplement la conscience que j'ai de la *subordination* de ma volonté à une loi sans entremise d'autres influences sur ma sensibilité, la détermination immédiate de la volonté par la loi et la conscience que j'en ai, c'est ce qui

s'appelle le *respect*<sup>11</sup>». La reconnaissance de la valeur absolue de la loi morale est pour l'être raisonnable la source d'un abaissement, puisqu'elle fait faire pour un moment son amour-propre. C'est la raison pour laquelle Kant mentionne que le respect se rapproche (analogiquement) de la *crainte*<sup>12</sup>. Mais le respect ne nous fait pas uniquement faire l'expérience d'un abaissement; il est aussi la source d'une élévation<sup>13</sup>. Car, cette loi morale que nous respectons, c'est en quelque sorte *librement* que nous nous la donnons. Selon cet autre point de vue, la loi morale est pour ainsi dire la conséquence de notre volonté<sup>14</sup>, et le respect peut être vu comme se rapprochant aussi par analogie de *l'inclination*. Mais nous reviendrons plus tard sur cet élément pivot de la philosophie morale kantienne.

Récapitulons puis terminons cette courte incursion dans la première partie des *Fondements*. L'analyse de Kant parvient à montrer que la notion de bonne volonté est contenue dans le concept du devoir, et que ce dernier décomposé mène à celui du respect, puis enfin à l'idée de loi. La bonne volonté, en effet, agit par devoir, ce qui ne signifie rien de plus qu'elle agit non seulement conformément à la loi morale, mais aussi *pour* la loi morale<sup>15</sup>, car, disions-nous, le devoir est «la nécessité d'accomplir une action par respect pour la loi<sup>16</sup>». Se représenter la loi en elle-même, pourrait-on dire, et faire de cette représentation de principe déterminant de l'action, voilà ce qui caractérise la bonne volonté, voilà ce qui fait qu'une action a un caractère moral.

Et quelle est donc cette loi? L'analyse a dégagé jusqu'ici suffisamment de paramètres pour en déterminer le contenu, car, d'un certain point de vue, on pourrait dire qu'elle ne doit pas en avoir. En effet, comme la bonne volonté n'est déterminée en aucune façon par l'influence des inclinations, par les fins *matérielles* que celles-ci lui proposent, il ne lui reste plus qu'à se déterminer d'après un principe simplement *formel*: la nécessité de conformer ses maximes<sup>17</sup> à la l'idée même de loi<sup>18</sup>. «Puisque j'ai dépossédé la volonté de toutes les impulsions qui pourraient être suscitées en elle par l'idée des résultats dus à l'observation de quelque loi, il ne reste plus que la conformité universelle des actions à la loi en général, qui doit seule lui servir de principe<sup>19</sup>». Au terme de la première partie,

Kant fait toucher son lecteur à ce qui deviendra plus tard la première formule de l'impératif catégorique : « je dois toujours me conduire de telle sorte *que je puisse aussi vouloir que ma maxime devienne une loi universelle*<sup>20</sup> ».

La seconde partie des *Fondements*, quant à elle, vise à opérer le passage de la philosophie morale populaire à la métaphysique des mœurs. Par un chemin quelque peu différent, elle mènera le lecteur une nouvelle fois à la première formule de l'impératif catégorique. Pour ce faire, Kant fait remarquer que la faculté d'agir d'après la *représentation* d'une loi ne peut appartenir qu'à un être raisonnable. La raison, en effet, est requise « pour dériver des actions des lois<sup>21</sup> », et la volonté, pour cette raison, « n'est rien d'autre qu'une raison pratique<sup>22</sup> ». Pour un être à la fois raisonnable et sensible tel que l'homme, dont la volonté, parce qu'elle subit aussi l'influence des inclinations, n'est pas nécessairement docile à la raison, n'est pas pleinement et infailliblement déterminé par elle, la nécessité pratique d'une action se présente comme une *contrainte*, comme un *tu dois*, comme un *commandement*. Et la formule « du commandement s'appelle un IMPÉRATIF<sup>23</sup> ».

Tous les impératifs ne sont pas moraux. Certains impératifs, en effet, « représentent la nécessité pratique d'une action possible, considérée comme moyen d'arriver à quelque autre chose que l'on veut (ou du moins qu'il est possible qu'on veuille)<sup>24</sup> ». L'action à accomplir, à travers eux, est conçue comme bonne uniquement en tant que *moyen* en vue d'obtenir une *autre chose* qui est ou peut être objet de désir<sup>25</sup>. Ce sont les impératifs *hypothétiques*, que l'on dit tels parce qu'ils ne commandent que conditionnellement, ils ne prescrivent quelque chose que *si* la volonté se propose telle ou telle fin. Kant distingue parmi ceux-ci les impératifs de *l'habileté*, ou impératifs techniques, qui mettent, par des *règles*, la volonté en relation avec une fin *possible*<sup>26</sup>. Pour faire tel gâteau, par exemple, il faut mélanger tels ingrédients dans tel ou tel ordre. Les impératifs de la *prudence*, ou impératifs pragmatiques, sont également hypothétiques ; ils *conseillent* à la volonté les actions qui pourraient mener au bonheur, fin que se proposent nécessairement tous les hommes en tant qu'ils sont des êtres sensibles, mais qu'ils se

représentent tous différemment selon les aléas de leur imagination. Dans ces deux cas, la volonté se laisse alors déterminer de manière « pathologique<sup>27</sup> », c'est-à-dire par une représentation d'objet ainsi que du plaisir ou de la peine qui lui sont reliés. Ce genre de principe ne se fonde donc que sur une condition *subjective*, la capacité à ressentir du plaisir ou de la peine, condition « qui ne peut jamais être connue qu'empiriquement et ne peut être supposée à un degré égal chez tous les êtres raisonnables<sup>28</sup> ». Comme tels, ceux-ci peuvent bien servir de maximes « propres au sujet qui la possède<sup>29</sup> », mais ils ne peuvent certainement pas servir de *lois* pratiques, car ils manquent de nécessité objective et d'universalité.

Une loi pratique ne peut relever que d'un principe déterminant la volonté indépendamment de toute condition particulière ou subjective, d'un principe, par suite, qui n'est pas conditionné par une fin désirée. Et là où une action n'est pas prescrite par la raison comme moyen en vue d'autre chose, elle l'est pour elle-même, comme bonne *en elle-même*<sup>30</sup>. Un tel impératif peut être dit *catégorique* : il commande immédiatement une conduite<sup>31</sup>. Parce que l'impératif catégorique représente l'action comme bonne en elle-même, il contraint *inconditionnellement* et se présente à proprement parler comme un *commandement*. Conséquemment, il constitue une *loi* de la volonté, car « il n'y a que la loi qui entraîne avec soi le concept d'une nécessité inconditionnée, véritablement objective, par suite d'une nécessité universellement valable, et les commandements sont des lois auxquelles il faut obéir<sup>32</sup> ». L'impératif catégorique est l'impératif de la moralité, dont le simple concept, observe Kant, en fournit la formule<sup>33</sup>. En effet, si l'on ne peut pas d'avance savoir le contenu des impératifs hypothétiques, car ils sont conditionnés par les objets que la volonté se propose et qui dépendent de l'expérience, ce n'est pas le cas de l'impératif catégorique, qui est, lui, inconditionné. L'impératif catégorique ne concerne pas la matière ou le résultat de l'action, mais plutôt le principe dont celle-ci résulte : la maxime (ou l'intention)<sup>34</sup>. En dehors de la loi, l'impératif ne contient que « la nécessité, pour la maxime, de se conformer à cette loi<sup>35</sup> ». Or si on dépouille cette loi de toute matière, il ne reste que la simple forme universelle de la loi, avec la nécessité, pour la maxime, de s'y confirmer<sup>36</sup>. De là

la première formule de l'impératif catégorique : « *Agis uniquement d'après la maxime qui fait que tu peux vouloir en même temps qu'elle devienne une loi universelle*<sup>37</sup> ».

*Deuxième partie : Dérivation des autres formules*

Il n'y a qu'un impératif catégorique, et il est exprimé par la formule que nous venons d'énoncer. Pourtant, l'impératif catégorique peut être représenté d'autres manières. Celle qui suit attire notre attention sur cette *forme* bien particulière qu'il commande aux maximes de prendre : « *Agis comme si la maxime de ton action devait être érigée par ta volonté en loi universelle de la nature*<sup>38</sup> ». Cette nouvelle formule s'avère d'une importance cruciale : elle fournit, selon Kant, un « canon qui permet l'appréciation morale de notre action en général<sup>39</sup> ». Dans sa *Critique de la raison pratique*, Kant affirme même que c'est par cette formule, « en fait, que chacun juge si les actions sont moralement bonnes ou mauvaises (*gut ober böse*)<sup>40</sup> ».

Il pourrait sembler étrange que Kant, pour formuler une loi de liberté, ait recours à la notion de nature. Pourtant, il n'en est rien. Il parle ici de la nature considérée sous son aspect formel, « *das Formale der Natur* » (*Prolégomènes*, § 17<sup>41</sup>). La nature ainsi entendue est « la connexion universelle de l'existence des choses selon des lois universelles<sup>42</sup> ». Elle renvoie donc à la liaison systématique du donné de l'intuition selon des lois universelles, œuvre de notre entendement. L'universalité des lois de la nature et celle de la loi morale les rend, en quelque sorte, homogènes. Comme la première formule de l'impératif catégorique enjoint à la volonté d'adopter des maximes pouvant prendre la forme de lois universelles, et que « l'universalité de la loi constitue la forme de la nature<sup>43</sup> », Kant peut établir sans difficulté cette formulation de l'impératif catégorique.

Celle-ci semble comporter l'avantage d'être plus près de l'intuition commune, plus facilement évocatrice et donc plus facilement applicable que la première. C'est en effet grâce à elle que Kant trace les grandes distinctions entre les devoirs parfaits et imparfaits envers soi et envers les autres, et non simplement avec la formule originelle de l'impératif catégorique. Dans la *Critique de la raison pratique*, Kant nous indique que cet avantage tient au fait que les lois

de la nature posées par l'entendement jouent pour le jugement le rôle de modèles, de *types* lui permettant d'effectuer la subsumption des actions particulières sous une loi universelle<sup>44</sup>. Gilles Deleuze résume ce point de manière très synthétique: «C'est par analogie avec la forme des lois théoriques de la nature sensible que nous cherchons si une maxime peut être pensée comme loi pratique<sup>45</sup>». La loi de la nature prise comme type d'une loi universelle est au jugement ce que le schème de l'imagination est à l'entendement. L'imagination joue chez Kant le rôle d'une faculté intermédiaire par laquelle les concepts purs de l'entendement parviennent à informer le divers des intuitions sensibles pour construire l'expérience. Elle schématise; en elle, les catégories sont en acte<sup>46</sup>. Ici, c'est l'entendement qui joue le rôle de faculté intermédiaire. En fournissant au jugement le modèle d'une loi universelle, il lui permet d'appliquer la loi morale universelle et pure à des événements particuliers du monde sensible. C'est l'entendement qui permet de comprendre une maxime à partir de la loi morale, de juger de la moralité d'une maxime<sup>47</sup>.

Certes, on peut concevoir la volonté comme «une faculté de se déterminer soi-même à agir *conformément à la représentation de certaines lois*<sup>48</sup>», mais on peut aussi envisager en elle la faculté de se déterminer en fonction de fins. Pour cette raison, si jusqu'ici Kant s'était appliqué à illustrer la loi morale en vertu de la forme qu'elle commande aux maximes de la volonté de prendre, il le fera maintenant en exposant ce qu'elle prescrit à l'égard de la *matière* (ou de la fin) des maximes<sup>49</sup>. Il propose, pour le faire, une nouvelle façon de formuler l'impératif catégorique: «*Agis de telle sorte que tu traites l'humanité aussi bien dans ta personne que dans la personne de tout autre toujours en même temps comme une fin, et jamais simplement comme un moyen*<sup>50</sup>».

Cette nouvelle façon d'envisager l'impératif catégorique a surtout une valeur limitative<sup>51</sup>: elle pose des bornes claires en vue de l'action. Elle attribue à l'existence de l'être raisonnable la valeur d'une fin en soi et en fait la «condition suprême restrictive de toutes les fins subjectives<sup>52</sup>». Cela revient à dire que celui-ci ne peut simplement être considéré comme un *moyen*; il est une *personne*, non une chose.

Pour en arriver à cette formule, l'examen doit déterminer le genre de fin s'avérant compatible avec la notion d'impératif catégorique, car toutes ne le sont pas. Celles que se propose l'être raisonnable comme *effet* de son action (alors que l'action est considérée elle-même comme *moyen* de produire cet effet) sont relatives, c'est-à-dire qu'elles tiennent leur valeur «de la nature particulière de la faculté de désirer du sujet<sup>53</sup>». Ces fins «à réaliser<sup>54</sup>» par une action ne donnent lieu qu'à des impératifs hypothétiques, et, nous l'avons vu dans la première partie de cet article, elles ne peuvent donc fournir de principe pratique *universel*, de *loi* pratique. Une fin en mesure de le faire devrait être donnée par la seule raison, et donc «valoir pour tout être raisonnable<sup>55</sup>». Il faut donc chercher une fin objective, dont la valeur ne se fonde pas dans les dispositions, les inclinations ou les besoins de tel ou tel agent. Comme ces fins ne peuvent pas être comprises comme des fins à réaliser, «il faut que la fin soit conçue ici (...) comme une fin existant par soi<sup>56</sup>», une fin dont *l'existence* elle-même possède une *valeur absolue*; bref, il faut identifier une *fin en soi*<sup>57</sup>, car seule une telle fin s'avère compatible avec la notion d'impératif catégorique.

Or cette fin en soi ne peut être autre chose que l'être raisonnable lui-même, cet être capable de se proposer des fins et qui est le sujet de toutes les fins possibles. Tout être raisonnable se représente nécessairement sa propre existence comme une fin en soi, non par *amour de soi*, mais en vertu d'un principe rationnel qui vaut pour tous<sup>58</sup>, et que Kant promet d'aborder plus tard. Si donc nous comprenons jusqu'à maintenant la raison pour laquelle le respect du statut de fin en soi dont la nature raisonnable dote son existence peut bel et bien servir de principe objectif de la volonté (*tout* être raisonnable se considère comme une fin en soi), nous ne voyons pas immédiatement quel est ce «principe rationnel» auquel Kant fait allusion et qui permettrait de justifier cette prétention. Néanmoins, remarquons que, dans les premiers moments de son analyse, Kant a dégagé quelques éléments pouvant nous permettre de le comprendre. En effet, l'homme et en général tout être raisonnable peut se représenter ses fins en accord avec la loi morale, peut, autrement dit, se faire *bonne volonté*, seul bien dont la valeur est

inconditionnelle et absolue. Or une volonté absolument bonne « ne peut sans contradiction être mise au-dessous d'aucun autre objet<sup>59</sup> », c'est-à-dire être considérée dans l'action uniquement comme un moyen en vue de l'obtention ou de la réalisation d'autre chose. Il apparaît dès lors que cette *capacité* de l'être raisonnable à agir moralement, ou à universaliser ses maximes, fonde la valeur absolue de son existence et le distingue comme fin en soi<sup>60</sup>.

Nous pouvons remarquer ici l'intime proximité entre cette formule et les précédentes. Ne nous en étonnons pas : la troisième formule ne fait qu'un avec les deux premières, note Kant. Aussi permet-elle tout aussi bien de rendre compte des exemples qu'il avait traités à l'aide de la deuxième :

car dire que dans tout usage des moyens en vue d'une fin je dois imposer à ma maxime cette condition limitative, qu'elle vaille universellement comme une loi pour tout sujet, revient précisément à ceci : que pour principe fondamental de toutes les maximes des actions, il faut poser que le sujet des fins, c'est-à-dire l'être raisonnable même, ne doit jamais être traité simplement comme un moyen, mais comme une condition limitative suprême dans l'usage de tous les moyens, c'est-à-dire toujours en même temps comme une fin<sup>61</sup>.

Il est inconcevable que l'être raisonnable puisse *vouloir* ériger en loi pratique universelle une maxime qui pourrait le réduire à agir comme un simple moyen en vue de la réalisation d'une fin qui lui serait étrangère, puisque ce serait simultanément nier une des caractéristiques les plus fondamentales dont il est doté comme être de volonté : celle d'avoir la capacité de se déterminer pour des fins, de se donner des fins, et d'être le sujet de toutes les fins possibles. Autrement dit, un ordre des choses où les êtres raisonnables ne se respecteraient pas entre eux comme des fins en soi ne peut pas être voulu par eux comme une nature possible, si bien que cette formule qui leur enjoint de traiter en tant que fin en soi l'humanité dans chaque personne découle des premières façons de représenter l'impératif catégorique comme une conséquence directe et nécessaire<sup>62</sup>.

Or envisager l'unité des diverses formules qu'a jusqu'ici dégagée l'analyse permettrait selon Kant de former une nouvelle idée : celle « *de la volonté de tout être raisonnable conçue comme volonté instituant une législation universelle*<sup>63</sup> ». Cette idée, écrit-il, découle « *incontestablement*<sup>64</sup> » de ce rapprochement.

Comme l'a remarqué Paton, l'idée était déjà implicite dans les premières formules de l'impératif catégorique, qui commandaient d'agir d'après la maxime que l'on pouvait *vouloir* qu'elle devienne une loi universelle<sup>65</sup>, qu'elle soit érigée par notre volonté en loi universelle de la nature<sup>66</sup>. Mais maintenant que l'analyse a révélé que l'aptitude même de l'être raisonnable à se représenter ses maximes en accord avec la loi morale, ou, ce qui revient au même, à universaliser ses maximes, le distingue comme fin en soi, elle nous conduit aussi à le considérer comme la *source* et le *législateur* de la loi morale. Cependant, pour nous, une fois de plus, la raison du surgissement de ce nouveau principe de la volonté n'apparaît pas avec évidence, et mérite un certain effort de reconstitution pour être comprise.

D'une part, on pourrait relever comme l'a fait Delbos qu'on ne peut plus simplement se représenter l'être raisonnable comme *soumis* à une loi morale qui s'imposerait à lui du dehors ; il serait alors, pour l'accomplissement de celle-ci, tout à fait à son *service*, c'est-à-dire un simple *moyen*<sup>67</sup>. La volonté doit donc être regardée comme *autonome*, comme instituant elle-même la loi à laquelle elle se plie.

D'autre part, il s'agit ici d'explorer une perspective nouvelle qui permet d'expliquer le *motif* de l'obéissance à la loi morale. Kant, pourrait-on dire, a déjà esquissé la réponse à ce problème : une bonne volonté agit par seul *respect* pour la loi morale, c'est-à-dire de manière désintéressée, sans calcul d'intérêt, de sorte que ne sont pas déterminants pour son action les multiples mobiles que lui propose sa sensibilité. Mais il s'agit ici d'expliquer comment un tel respect désintéressé est possible. Qu'est-ce qui explique, autrement dit, *l'intérêt* simplement *moral*<sup>68</sup> que nous prenons pour la loi ? La réponse de Kant, en substance, est la suivante : si l'être raisonnable obéit à la loi morale, c'est qu'elle est pour ainsi dire le produit de sa volonté, elle est *sienne*, et il n'y est soumis « que

pour cette raison<sup>69</sup>». *Seul* ce motif s'avère compatible avec la notion d'impératif catégorique.

En effet, pour s'expliquer la raison pour laquelle un être raisonnable dont la volonté n'est pas parfaitement conforme à la loi morale obéit néanmoins à celle-ci, il faut supposer qu'il éprouve un intérêt envers elle, mais un intérêt qui ne peut en aucun cas s'identifier à l'intention d'agir *par* intérêt. La première partie des *Fondements* l'avancait déjà, et la deuxième — particulièrement le point de départ de celle-ci — permet tout aussi bien de comprendre pourquoi cette idée est une conséquence nécessaire de l'analyse déployée par Kant jusque-là. Si l'on ne conçoit pas que l'obéissance d'un être raisonnable est en vérité fondée sur cet intérêt «pur et indépendant des sens<sup>70</sup>» résidant dans le fait qu'il s'avère l'auteur de la loi, «celle-ci impliquerait nécessairement en elle un intérêt sous forme d'attrait ou de contrainte, parce qu'elle ne dériverait pas comme loi de *sa* volonté, et que sa volonté serait forcée conformément à la loi *par quelque chose d'autre* à agir d'une certaine manière<sup>71</sup>». Autrement dit, si l'on n'admet pas le principe de l'autonomie de la volonté, il faut alors se représenter la volonté de l'être raisonnable qui obéit à la loi morale comme contrainte conditionnellement : il fait l'action prescrite parce que celle-ci est le *moyen* d'obtenir telle *autre chose*. C'est alors en fait l'objet de la volonté qui lui donne sa loi, et cette détermination *hétéronome* de la volonté donne forcément lieu à un impératif hypothétique.

S'il y a une telle chose que l'impératif catégorique, celui-ci exprime et révèle forcément *l'autonomie* de la volonté, c'est-à-dire «cette propriété qu'à la volonté d'être à elle-même sa loi<sup>72</sup>», de sorte que l'on pourrait avancer, comme l'a fait Delbos<sup>73</sup>, que Kant soutient pour l'ordre moral ce que Rousseau avait soutenu pour l'ordre social légitime, cet ordre où règne la liberté, c'est-à-dire «l'obéissance à la loi qu'on s'est prescrite<sup>74</sup>». Bien que ce rapprochement soit incontestable, étude minutieuse de la notion d'autonomie nous fera plutôt voir la distance qui sépare Kant de Rousseau sur ce point, et nous permettra de cerner la spécificité de la manière dont Kant la conçoit.

En voyant dans les lois de l'État légitime l'expression de la volonté générale des citoyens, Rousseau caresse le projet de réconcilier nos devoirs et nos penchants en faisant dériver la justice de l'amour que chacun se porte naturellement. Il écrit à ce sujet dans son *Contrat Social* : « Pourquoi la volonté générale est-elle toujours droite, et pourquoi tous veulent-ils le bonheur de chacun d'eux, si ce n'est parce qu'il n'y a personne qui ne s'approprie ce mot *chacun*, et qui ne songe à lui-même en votant pour tous ? Ce qui prouve que l'égalité de droit et la notion de justice qu'elle produit dérivent de la préférence que chacun se donne et par conséquent de la nature de l'homme<sup>75</sup> ». Pour Kant, au contraire, le *devoir être* relève de la métaphysique<sup>76</sup>, non de l'étude de la nature, et le principe suprême de la morale, qui permet de fournir des lois pratiques universelles et inconditionnées, ne relève en aucune façon de la *constitution particulière de la nature humaine* :

ce qui est dérivé de la disposition naturelle propre de l'humanité, ce qui est dérivé de certains sentiments et de certains penchants, et même, si c'est possible, d'une direction particulière qui serait propre à la raison humaine et ne devrait pas nécessairement valoir pour la volonté de tout être raisonnable, tout cela peut bien nous fournir une maxime à notre usage, mais non une loi, un principe subjectif selon lequel nous pouvons agir par penchant et inclination, non un principe objectif d'après lequel nous *aurions l'ordre* d'agir, alors même que tous nos penchants, nos inclinations et les dispositions de notre nature y seraient contraires<sup>77</sup>.

La manière dont Kant comprend la liberté s'inscrit de manière cohérente dans le sillage tracé par son analyse jusque-là : elle s'avère pour lui la capacité de la volonté à se décider indépendamment des objets de l'expérience vers lesquelles tendent nos inclinations et nos penchants, indépendamment, autrement dit, de causes étrangères et *déterminantes*<sup>78</sup> ne générant chez nous que des impératifs hypothétiques. Cela revient à dire que la liberté constitue une certaine indépendance face au réseau des lois causales de la nature, qui soumettent les objets empiriques à des réactions nécessaires

et prévisibles lorsque d'autres objets extérieurs les influencent. La volonté peut au contraire déterminer entièrement sa maxime d'action auprès d'une loi purement rationnelle qui lui est propre et qui ne fait pas, elle, partie de la « nature<sup>79</sup> ». Cette loi, c'est l'impératif catégorique lui-même : « cette proposition : la volonté dans toutes les actions est à elle-même sa loi, n'est qu'une autre formule de ce principe : il ne faut agir que d'après une maxime qui puisse aussi se prendre elle-même pour objet à titre de loi universelle<sup>80</sup> ». La moralité se révèle en fait cette capacité de la volonté d'agir pour elle-même à titre de cause ; volonté autonome et volonté soumise à la loi morale sont une seule et même chose.

Voilà donc ce que Kant désigne lorsqu'il parle d'autonomie. Si donc Rousseau, en philosophe politique, souhaite donner voix aux mouvements premiers de la nature lors du processus de législation, Kant, en philosophe moral, dissocie nature et morale, penchants et législation de la volonté. L'examen de cette affinité entre la philosophie rousseauiste et kantienne nous mène alors au constat d'une différence de fond remarquable entre leurs pensées. Mais, fait étonnant, ce contraste n'empêche pas Kant de parvenir à la même conclusion que Rousseau : liberté et moralité ne font qu'un, puisque c'est la volonté qui légifère en matière de morale. On voit mieux, après toutes ces considérations, pourquoi Kant affirme que l'autonomie de la volonté est le « principe suprême de la moralité<sup>81</sup> ».

Un intérêt simplement moral envers la loi est donc possible parce que le respect doit se comprendre non seulement comme la reconnaissance d'une valeur qui nous humilie, comme la *conscience* de la subordination de notre volonté à une loi et de sa « détermination immédiate<sup>82</sup> » par elle, mais encore comme inséparable de l'idée de la « sublimité de notre nature<sup>83</sup> », c'est-à-dire de celle de notre *personnalité*, de notre qualité d'être libres et autonomes<sup>84</sup>, si bien que le respect pour la loi est simultanément un respect pour notre plus « haute détermination<sup>85</sup> ».

Cette découverte est importante, puisqu'elle permet de terminer l'analyse de l'impératif catégorique. Puisque les êtres raisonnables doivent tous agir selon des maximes universalisables, puisque leur capacité à le faire les distingue comme des fins en soi, et que, à

son tour, cette qualité nous conduit à envisager l'autonomie de leur volonté, nous sommes tout naturellement conduits à donner au concept d'impératif catégorique une détermination complète, à penser qu'un règne des fins est possible par liberté. Qu'est-ce donc qu'un «règne des fins»? Il s'agit d'un idéal et d'une idée de la raison, l'idée d'un tout de toutes les fins, aussi bien des fins en soi (les êtres raisonnables) que des fins relatives que ceux-ci se proposent (celles, du moins, qui sont compatibles avec la loi morale). Plus précisément, cette idée donne à penser la liaison et l'union systématique de tous les êtres raisonnables sous des lois communes et universelles dont ils sont les colégislateurs et par lesquelles ils se traitent tous réciproquement comme des fins en soi.

La dernière étape de l'analyse de Kant le mène donc à penser la moralité comme le rapport de toute action à la législation qui rend possible un tel règne des fins. De là la dernière formulation de l'impératif catégorique, qui commande à tout être raisonnable d'agir «comme s'il était toujours par ses maximes un membre législateur dans le règne universel des fins<sup>86</sup>».

On voit facilement en quoi cette formulation découle naturellement de l'analyse : elle effectue la synthèse de toutes les autres façons de représenter l'impératif catégorique et donne à *voir* leur unité et leur cohérence d'ensemble. Aussi permet-elle de mieux comprendre ce qui fonde la valeur absolue de l'être raisonnable et le distingue comme une fin en soi. En effet, l'idée d'un règne des fins permet de hiérarchiser les fins. Certaines, comme l'habileté ou l'application dans le travail<sup>87</sup> n'ont qu'un prix marchand (*Marktpreis*) et ne tirent leur valeur que des besoins ou des inclinations que peut avoir tel ou tel être raisonnable. D'autres, comme la vivacité d'imagination ou l'humour ont un prix de sentiment (*Affektionspreis*), et tirent leur valeur du goût. La valeur de ces deux types de fins est simplement conditionnelle et est fonction de dispositions subjectives. Chacun des objets auxquelles elles s'appliquent peut être remplacé par quelque chose d'une valeur équivalente. Mais il existe aussi un autre type de valeur, une valeur intrinsèque et absolue, la dignité, qui se rapporte à quelque chose de tout à fait irremplaçable. Une seule chose possède

une dignité, c'est la « condition qui seule peut faire que quelque chose est une fin en soi<sup>88</sup> » : la moralité.

Nulle chose, écrit Kant, « n'a de valeur en dehors de celle que la loi lui assigne. Or la législation même qui détermine toute valeur doit avoir précisément pour cela une dignité, c'est-à-dire une valeur inconditionnée, incomparable, que traduit le mot *respect*, le seul qui fournisse l'expression convenable de l'estime qu'un être raisonnable en doit faire. » Or lorsque l'être raisonnable agit moralement, lorsqu'il universalise sa maxime d'action, il rend en quelque sorte sa volonté identique à la loi, et sa volonté acquiert la valeur absolue de celle-ci<sup>89</sup>. L'intention que l'on devine à la source d'une action morale est alors l'objet du même respect que la loi elle-même ; elle constitue, en effet, l'exemplification de la loi, son image et son accomplissement<sup>90</sup>. « La moralité, ainsi que l'humanité en tant qu'elle est capable de moralité, c'est donc là ce qui seul a de la dignité<sup>91</sup> ». Par la moralité, par ce respect que l'être raisonnable porte à la loi dont sa raison est la source, l'être raisonnable prend valeur de fin en soi et devient membre législateur dans le règne des fins.

Ce qui devient visible en bout de parcours, c'est que tout l'édifice repose sur la notion d'autonomie, notion sans laquelle un impératif catégorique est tout à fait impensable et sans laquelle l'action morale même semble impossible. Démontrer la possibilité de l'impératif catégorique en approfondissant le concept d'autonomie, c'est ce à quoi Kant s'appliquera dans la troisième partie des *fondements*, où il prendra pour cela une méthode différente : une méthode *synthétique*.

### Conclusion

La double étude que nous nous sommes proposée dans cet article — celle de la démarche menant Kant à la première formule de l'impératif catégorique et celle de la manière dont chacune des autres formules peut être dérivée des implications de cette étude préalable — met en évidence la complémentarité des formules de l'impératif catégorique et leur cohérence d'ensemble. Le déploiement du concept d'impératif catégorique a montré successivement trois façons de le représenter. L'une d'elles considère la *forme* qu'il commande aux maximes de prendre. Une autre met en évidence

la *fin* que la volonté se doit de respecter. En dégageant au passage le principe suprême de la moralité, l'autonomie de la volonté, l'analyse permet également de fournir à l'impératif catégorique une *détermination complète*, qui rassemble en un tout systématique tous les éléments dégagés par l'analyse. Dans cette dernière formulation de l'impératif catégorique, l'unité des diverses formules par lesquelles il est possible de l'exprimer se donne à voir.

Au fil de cette étude, nous avons tenté de montrer que cette unité des diverses façons d'exprimer l'impératif catégorique était difficilement envisageable sans effectuer ce retour aux premiers moments de l'analyse, aux deux chemins empruntés par Kant afin d'établir la formule originelle de l'impératif catégorique. Loin de constituer une difficulté à la doctrine des *Fondements de la métaphysique des mœurs*, il faut au contraire y voir le signe de la solidité et de la précision de l'examen qu'y mène Kant, car cela signifie qu'il ne s'est bel et bien appliqué dans chaque cas qu'à expliciter et rendre évidentes des conséquences enveloppées dès le départ dans le concept même d'impératif catégorique, et qu'il n'en a fait, jusqu'au bout, qu'une simple, mais rigoureuse *analyse*.

- 
1. Kant, *Critique de la raison pure*, trad. André Tremesaygues et Bernard Pacaud, Paris, PUF, 2012, p. 544.
  2. Kant, *Fondements de la métaphysique des mœurs*, Trad. Victor Delbos, Paris, Librairie Générale Française, 1993.
  3. *Ibid.*, p. 115.
  4. Alexis Philonenko, *L'œuvre de Kant*, vol. II, Paris, Librairie philosophique J. VRIN, 1981, pp. 115-116.
  5. Bien que, nous dit Kant, ce soit avec «certaines restrictions» et «certaines entraves subjectives». Voir *Fondements de la métaphysique des mœurs*, p. 62.
  6. Victor Delbos, *Notes*, 41, dans Emmanuel Kant, *Fondements de la métaphysique des mœurs*, p. 220.
  7. Kant, *Fondements de la métaphysique des mœurs*, p. 66.
  8. *Ibid.*, p. 67.
  9. Voir Kant, *Critique de la raison pratique*, p. 83.

10. *Ibid.*, p. 81.
11. Kant, *Fondements de la métaphysique des mœurs*, p. 68.
12. *Id.*
13. Voir encore Kant, *Critique de la raison pratique*, p. 83.
14. Kant, *Fondements de la métaphysique des mœurs*, p. 68.
15. *Ibid.*, p. 53.
16. *Ibid.*, p. 66.
17. Le terme « maxime » renvoie au « principe subjectif du vouloir ». Voir Kant, *Fondements de la métaphysique des mœurs*, p. 67.
18. *Ibid.*, p. 69 ; voir aussi Emmanuel Kant, *Critique de la raison pratique*, trad. François Picavet, Paris, PUF, 2012, p. 26.
19. Kant, *Fondements de la métaphysique des mœurs*, p. 69
20. *Ibid.*, p. 69.
21. *Ibid.*, p. 83.
22. *Id.* ; voir aussi p. 67.
23. *Ibid.*, p. 83.
24. *Ibid.*, p. 85.
25. *Id.* ; voir aussi Kant, *Critique de la raison pratique*, p. 18.
26. Kant, *Fondements de la métaphysique des mœurs*, p. 86 ; Kant, *Critique de la raison pratique*, p. 19.
27. Voir à ce sujet Gilles Deleuze, *La philosophie critique de Kant*, Paris, PUF, 2011, p. 12.
28. Kant, *Critique de la raison pratique*, p. 20.
29. *Id.*
30. Voir là dessus H.J.Paton, *The Categorical Imperative*, London, Hutchinson's University Library, 1946, p. 115.
31. Kant, *Fondements de la métaphysique des mœurs*, p. 88.
32. *Id.* ; voir aussi p. 93.
33. *Ibid.*, p. 94.
34. *Ibid.*, p. 88.
35. *Ibid.*, p. 94.
36. *Id.* ; Kant, *Critique de la raison pratique*, pp. 26-30.
37. Kant, *Fondements de la métaphysique des mœurs*, p. 94.
38. *Ibid.*, p. 95.
39. *Ibid.*, p. 98.
40. Kant, *Critique de la raison pratique*, p. 72.

41. Kant, *Prolégomènes à toute métaphysique future*, trad. Louis Guillermit, Paris, Vrin, 2012, p. 78.
42. Kant, *Fondements de la métaphysique des mœurs*, p. 116; *Critique de la raison pratique*, p. 95; *Prolégomènes*, p. 75.
43. H. J. Paton, *The Categorical Imperative*, p. 157; Gilles Deleuze, *La philosophie pratique de Kant*, p. 49
44. Kant, *Critique de la raison pratique*, p. 71.
45. Gilles Deleuze, *La philosophie pratique de Kant*, p. 49.
46. Pour de plus longs développements sur le sujet, voir par exemple Kant, *Critique de la raison pure*, pp. 151-156.
47. Kant, *Critique de la raison pratique*, p. 72.
48. Kant, *Fondements de la métaphysique des mœurs*, p. 103.
49. *Ibid.*, p. 115.
50. *Ibid.*, p. 105.
51. *Id.*
52. *Ibid.*, p. 108.
53. *Ibid.*, p. 103.
54. *Ibid.*, p. 117.
55. *Ibid.*, p. 103.
56. *Ibid.*, p. 117.
57. H. J. Paton, *The Categorical Imperative*, p. 168.
58. Kant, *Fondements de la métaphysique des mœurs*, p. 105.
59. *Ibid.*, p. 117.
60. Voir aussi Kant, *Critique de la raison pratique*, p. 92.
61. Kant, *Fondements de la métaphysique des mœurs*, p. 117.
62. C'est également l'avis de Paton. Voir H. J. Paton, *The Categorical Imperative*, p. 178.
63. *Ibid.*, p. 108.
64. *Ibid.*, p. 117.
65. *Ibid.*, p. 94.
66. *Ibid.*, p. 95.
67. Victor Delbos, «La morale de Kant», dans Kant, *Fondements de la métaphysique des mœurs*, p. 179.
68. Le respect est une forme d'intérêt moral. Voir *Fondements*, p. 68, fin de la note de bas de page de Kant. Voir aussi Kant, *Critique de la raison pratique*, p. 83.
69. Kant, *Fondements de la métaphysique des mœurs*, p. 108.

70. Kant, *Critique de la raison pratique*, p. 83.
71. *Ibid.*, p. 110.
72. *Ibid.*, pp. 121, 128.
73. Victor Delbos, *La philosophie pratique de Kant*, p. 306.
74. Jean-Jacques Rousseau, *Contrat Social* dans Œuvres complètes, Gallimard, «Bibliothèque de la Pléiade», t. III, 1964, p. 361.
75. Rousseau, *Contrat Social*, pp. 373-374.
76. Kant, *Fondements de la métaphysique des mœurs*, p. 102.
77. *Ibid.*, p. 100.
78. *Ibid.*, p. 127.
79. Ce point est traité dans toute la troisième partie des *Fondements*.
80. Kant, *Fondements de la métaphysique des mœurs*, p. 128.
81. *Ibid.*, p. 121.
82. *Ibid.*, p. 68.
83. Kant, *Critique de la raison pratique*, p. 92.
84. *Ibid.*, p. 91.
85. *Ibid.*, p. 93.
86. Kant, *Fondements de la métaphysique des mœurs*, p. 118.
87. Les exemples utilisés sont tirés directement des *Fondements*. Voir p. 113.
88. *Id.*
89. Victor Delbos explique ce point d'une manière similaire. Voir Victor Delbos, *La philosophie pratique de Kant*, p. 308
90. Voir Kant, *Fondements de la métaphysique des mœurs*, p. 68 et 114; Voir aussi Kant, *Critique de la raison pratique*, p. 81
91. Kant, *Fondements de la métaphysique des mœurs*, p. 113.





# Autonomie et institution du social selon Debray, Gauchet et Castoriadis

THIBAUT TRANCHANT, *Université de Sherbrooke*

## *Introduction*

Une des interprétations qui a pu être faite des expériences totalitaires du XX<sup>e</sup> siècle consiste à les comprendre en fonction d'un hypothétique «retour du religieux» dans un monde sécularisé<sup>1</sup>. S'appuyant sur une définition extensive du religieux<sup>2</sup>, une telle interprétation voit dans certains aspects des totalitarismes — culte du chef, référence à un texte «sacré», idéologie à dimension eschatologique, etc. — des manifestations explicites d'une fondation religieuse de la société totalitaire. Les enjeux de ces interprétations, qu'on ne discutera pas comme telles ici, excèdent largement ces seules sociétés, car en cherchant à prouver l'hypothèse selon laquelle les sociétés totalitaires se fondent religieusement et, au-delà, qu'aucune société ne peut se fonder en dehors de médiations religieuses, ces interprétations du totalitarisme entrent en contradiction avec une des exigences normatives de la modernité philosophique, à savoir l'exigence d'une autofondation de la pensée et de la pratique humaine. En soutenant qu'aucune société n'existe sans une religion lui permettant de se donner corps et de perdurer dans le temps, l'autonomie, valeur centrale de la philosophie pratique moderne, est dite impossible: l'homme serait condamné à l'hétéronomie, condamné à chercher dans un ailleurs l'origine de son action et de sa pensée. Ce problème que suscite le totalitarisme est selon nous un des points de départ et un des axes autour duquel s'articule le «renouveau de la philosophie politique» en France. En effet, pour de nombreux intellectuels français à partir des années 1970, le détour par cet autre qu'est la société totalitaire déboucha sur une interrogation sur la place du religieux en tant qu'hypothétique structure transhistorique du social, mais aussi sur la possibilité de s'en émanciper, renouant par là avec un des motifs profonds de la philosophie politique

moderne. Cornélius Castoriadis fait paraître en 1975 son *Institution imaginaire de la société*<sup>3</sup>. En 1981, Régis Debray sort sa *Critique de la raison politique*<sup>4</sup>. Marcel Gauchet nous livre ensuite, en 1985, son *Désenchantement du monde*<sup>5</sup>. Tous ces textes, différents quant à leurs propos et à leurs enjeux internes, visent cependant à répondre à une question commune : à quelle condition une société est-elle possible ? Existe-t-il, en deçà et au-delà de l'infinie variabilité de l'histoire, un principe auquel toute société doit se soumettre afin d'exister comme telle et qui nous condamnerait à orienter notre action et notre pensée en fonction de normes sur lesquelles nous n'aurions aucun contrôle ? Si un tel instituant existait, quelle en serait la nature et pouvons-nous nous en émanciper ?

Dans les quelques pages qui vont suivre, nous présenterons d'abord deux réponses contraires au même problème : celles de Régis Debray et de Marcel Gauchet. Le premier soutient la thèse pessimiste selon laquelle le projet moderne d'autonomie politique est irréalisable en raison d'une loi naturelle voulant que toute société se fonde en référence à une transcendance, quelle que soit la nature de cette dernière. Le second soutient la thèse que l'instituant premier du social est le politique, et que la religion est un phénomène historique duquel les sociétés occidentales contemporaines se sont émancipées. Nous chercherons finalement à dépasser l'antinomie à laquelle aboutit l'opposition de ces deux auteurs. À cette fin, nous nous appuierons surtout sur le concept d'*autonomie relative* proposé par Castoriadis dans *L'institution imaginaire de la société*.

### *Régis Debray : l'hétéronomie structurelle du champ politique*

Dans le premier chapitre de la seconde section de sa *Critique de la raison politique*, Régis Debray propose ce qu'il appelle un « axiome d'incomplétude » visant à rendre compte de la genèse des groupes politiques et de leur condition de possibilité. Il emprunte librement le concept d'incomplétude à Kurt Gödel qui montre qu'une théorie en arithmétique est nécessairement insuffisante, puisqu'on trouve en elle des énoncés qui ne sont pas démontrables dans le cadre seul de la théorie<sup>6</sup>. L'axiome d'incomplétude de Debray est compris comme condition de possibilité *a priori* de l'organisation politique,

et s'applique à *toutes* les sociétés, passées, présentes, et futures. Ainsi écrit-il :

[La condition *a priori* de toute histoire politique] a la forme d'une loi logique, généralisation du théorème de Godel [*sic*] : Il n'y a pas de système organisé sans clôture, et *aucun système ne peut se clore à l'aide des seuls éléments intérieurs au système*. La fermeture d'un champ ne peut donc procéder contradictoirement que par ouverture à un élément extérieur au champ. Aucun ensemble de relations n'est relatif à lui-même, ou alors ce n'est plus un ensemble (ce n'est plus rien). Le reste s'en déduit : savoir « le conte plein de bruit et de fureur, raconté par un idiot, et ne signifiant rien<sup>7</sup> ».

Comme le remarque Debray, l'axiome d'incomplétude est strictement formel, et n'explique en rien, par exemple, quelle est la nature de l'élément servant à clôturer le groupe<sup>8</sup>. L'incomplétude debrayenne indique uniquement quelle est la structure formelle d'après laquelle tous les groupes politiques se fondent. C'est un « axiome », car cette proposition est évidente en soi, sans pour autant pouvoir être démontrée. Debray ajoute que l'incomplétude est descriptive et explicative. D'un côté, elle permet de décrire comment tous les groupes sont formés, d'un autre côté, elle permet de comprendre leurs comportements, notamment ceux liés à la légitimation du pouvoir et à la territorialisation.

Afin de faire comprendre à ses lecteurs ce qu'est l'incomplétude, Debray passe par une analogie géométrique. Une société peut se comprendre comme un cône. « Un groupe n'est pas un cercle, surface plane limitée par une ligne dont tous les points sont à égale distance d'un centre, car le point fixe du groupe n'est pas intérieur à son plan<sup>9</sup>. » Au même titre que la définition du cône inclut son sommet, la définition d'une société ne peut faire l'économie du « point fixe » qui la rend possible et qui se trouve en dehors de son plan d'immanence. Ce « point fixe » est une place vide, fonctionnelle. Il souligne que ce « point fixe » est sacralisé par les membres de la société ainsi réunie<sup>10</sup>. Debray est prudent, et fait remarquer que l'axiome d'incomplétude n'est pas une reformulation de la thèse durkheimienne sur la nature

sociale de la religion. Pour Debray en effet, l'axiome d'incomplétude se distingue de la thèse durkheimienne par sa dimension explicative alors que Durkheim n'aurait fait qu'enregistrer une régularité historique. En disant que l'effervescence collective est à l'origine de la spéculation religieuse, Durkheim n'aurait fait que hisser une observation, aussi juste soit-elle, au rang de théorie explicative alors que ce qu'il convient d'expliquer est la raison de cette effervescence. L'axiome d'incomplétude explique selon Debray les raisons et les effets de l'effervescence durkheimienne, non l'inverse<sup>11</sup>.

La constitution du *nous* sans un *autre* est donc clairement impossible dans la perspective de Debray, et il en tire les conséquences directes que cela implique pour une théorie de la légitimité du pouvoir. En visant Weber, il parle d'une «syntaxe» politique élémentaire qui fait abstraction des différents types d'autorité<sup>12</sup>. Selon Debray, si la domination peut prendre des formes variées, celle-ci ne peut qu'évoluer dans le cadre plus large qu'est la forme imposée par l'incomplétude. Que la domination soit rationnelle-légale, charismatique ou traditionnelle, elle se déploie invariablement selon le point fixe sacré à partir duquel la société se fonde. Cela fait dire à Debray que la domination est nécessairement représentative. Quelle que soit la nature du point de référence de la société, la domination n'est légitime qu'à condition qu'elle représente ce point sacré<sup>13</sup>. La représentation ne serait donc pas selon Debray une invention moderne : ce qui change avec la modernité est le représenté (la souveraineté populaire), mais non le mécanisme même. La légitimation de la domination dans un groupe est en outre selon lui nécessairement arbitraire. Il n'y a pas chez Debray de norme idéale du juste, mais plutôt la succession de différentes «satisfactions» du «trou fondateur<sup>14</sup>». Autre conséquence : la démocratie directe ou toute autre forme d'autogestion sont tout simplement renvoyées comme autant d'impossibilités logiques<sup>15</sup>.

Debray fait remarquer ailleurs que l'axiome d'incomplétude se vérifie dans les comportements du groupe social à l'égard de son territoire. Selon lui, le fait que la territorialisation de la société passe nécessairement par le point d'incomplétude a été relevé par Benveniste dans son *Vocabulaire des institutions indo-européennes*,

notamment au niveau de la racine commune qu'il a découverte entre le nom *rex*, roi, et le verbe *regere*, tracer une ligne. C'est toujours par référence à une transcendance que celui qui détenait l'autorité pouvait tracer la ligne du territoire, qui était alors investi d'une valeur sacrée<sup>16</sup>.

De l'axiome d'incomplétude découle finalement l'impossibilité de toute véritable autonomie politique; le groupe social trouvera toujours en dehors de lui-même le fondement de sa vie en commun, et la légitimité du pouvoir se fondera toujours sur une norme sacrée. Debray l'écrit d'ailleurs sans détour: «L'axiome d'incomplétude sociale a pour premier corollaire l'hétéronomie du champ politique<sup>17</sup>.»

La théorie debrayenne de l'incomplétude est séduisante, et elle rend compte des intuitions que nous pouvons avoir sur les tendances naturellement intolérantes des groupes politiques, mais aussi des pages les plus noires de l'histoire contemporaine de l'occident (et d'ailleurs). On peut toutefois se demander ce que vaut une telle théorie dans le contexte de nos sociétés contemporaines. En effet, deux traits complémentaires de nos sociétés complexes sont, dans le cadre moderne de la forme État-Nation des groupements politiques, l'interdépendance fonctionnelle des sous-systèmes sociaux (politique, juridique, économique, esthétique, etc.) et le déploiement interne des logiques de ces sous-systèmes en fonction de valeurs qui leur sont propres. Ainsi, en se concentrant sur le mécanisme qui permet à une société de se rassembler et de perdurer dans le temps (l'incomplétude), Debray passe à côté d'un des traits les plus importants de la modernisation des sociétés occidentales, à savoir le fait qu'elles se maintiennent comme unités nationales malgré un éclatement de leurs activités en sphères distinctes, éclatement pratique auquel correspond un éclatement normatif. Ce que nous reprochons en fait à Debray ici, est que son axiome d'incomplétude, du moins tel que formulé dans sa *Critique de la raison politique*, ne nous dit rien sur la tension proprement moderne entre l'imaginaire de l'unité nationale et la diversité effective des valeurs qui orientent différemment les activités au sein même de l'espace national. Il serait certes possible d'imaginer une solution théorique selon laquelle l'unité des sous-systèmes sociaux est

rendue possible grâce à une participation partielle de leurs valeurs à celles de l'imaginaire national (une économie «pour le peuple», etc.). Solution envisageable, mais solution incomplète, puisqu'il s'agit moins de penser la «participation» que le maintien de l'unité dans la différence, c'est-à-dire le fait que les États-Nations se maintiennent dans le temps malgré des oppositions pratiques et axiologiques incommensurables (par exemple, la contradiction entre l'orientation de l'activité politique en fonction du principe d'égalité, ne serait-ce que «formelle», et celle entre l'orientation de l'activité économique en fonction du principe de profit, impliquant une distinction inégalitaire entre dirigeants et dirigés, sanctionnée par une échelle, tout aussi inégalitaire, des salaires). Il faut en outre ajouter au pluralisme axiologique auquel aboutit l'autonomisation fonctionnelle des sous-systèmes sociaux, le pluralisme axiologique subjectif au niveau de la société civile — la séparation de la société civile et de l'État a pour corrélat la multiplication des préférences personnelles, c'est-à-dire la possibilité pour chacun de mener sa vie en fonction des valeurs qui lui semblent préférables. Si l'on peut émettre des réserves quant à l'influence des déterminants sociaux sur les conduites de vie, il n'en reste pas moins que la modernisation de la société coïncide avec la multiplication des morales individuelles et sous-culturelles. De ce point de vue, la théorie de Debray, plutôt que de comprendre cette autre caractéristique importante de la modernité, tend vers un diagnostic pessimiste selon lequel la multiplication des valeurs sur lesquelles s'appuient les conduites de vie est un facteur de désagrégation sociale et d'anomie. Ce n'est pas pour rien que Debray s'est inlassablement prononcé en faveur de la République qui, à ses yeux, est la seule forme de gouvernement apte à concilier modernisation de la société et fondement religieux du social.

Il faut également souligner qu'une telle théorie, qui fait de l'hétéronomie une conséquence nécessaire de la vie en commun, aboutit à un positionnement politique quiétiste et fataliste<sup>18</sup>. Comme nous l'avons évoqué, l'axiome d'incomplétude de Debray est un formalisme, c'est-à-dire qu'il est indifférent au contenu de la norme qui est censée structurer l'ensemble de la société à travers le mécanisme de sa représentation. La différence entre le droit divin

et la souveraineté populaire est indifférente aux yeux de la théorie debrayenne, ce qui importe n'étant que la puissance d'organisation du «point fixe». Le formalisme de la théorie debrayenne nous interdit finalement de critiquer le contenu des normes qui fondent l'être-ensemble, et refuse à la philosophie la légitimité du rôle qu'elle s'est attribué dès sa naissance, à savoir l'interrogation sans limites sur l'institué et l'appropriation de cette interrogation par tous au profit de tous<sup>19</sup>. Il se pourrait que d'un point de vue sociologique, les sociétés aient pour condition de fonctionnement la soumission à un «point fixe», mais, d'un point de vue philosophique, une telle théorie est inacceptable en ceci qu'elle conduit à l'acceptation sans réserve de l'institué sous prétexte qu'il ne peut en être autrement en raison d'une loi de nature des groupements sociaux. Contre les théories qui font de l'hétéronomie la conséquence nécessaire de l'existence à plusieurs, il nous semble au contraire qu'il faille maintenir l'idée moderne selon laquelle les sociétés sont des êtres *historiques* et *artificiels*, et que les hommes sont virtuellement ou effectivement capables de produire consciemment les conditions de leur vie en commun. Au contraire, la théorie de Debray nivelle toutes les différences historiques en se référant à une hypothétique fondation transhistorique de la société sur la religion et propose une conception naturaliste de l'agrégation sociale.

*Marcel Gauchet : l'histoire comme dévoilement de l'autonomie*

La philosophie de l'histoire de Marcel Gauchet, dont la dimension «téléologique» ou non peut faire question<sup>20</sup>, s'articule autour de la tension entre «condition politique» et «condition historique», cherchant ainsi à dégager dans l'histoire humaine la part du variable et celle de l'invariable. Si d'un côté toutes les sociétés doivent faire face à la division et au conflit qui les habitent et les structurent (la condition politique), elles trouvent d'un autre côté dans l'histoire des façons particulières d'y apporter une réponse. Gauchet trouve dans l'élucidation des rapports entre politique et religion un terrain privilégié d'analyse afin de comprendre les rapports entre conditions politique et historique. Comprenant la religion comme l'institué historique par excellence, Gauchet propose une philosophie de

l'histoire dans laquelle la condition politique des sociétés se révèle peu à peu à elle-même comme l'instituant transhistorique du social, alors que la religion échoue peu à peu à jouer sa fonction intégratrice et pacificatrice. En reconstituant une histoire politique de la religion où le « désenchantement » est compris comme moment de vérité, Marcel Gauchet assume dès lors une vision des rapports entre politique et religion dans laquelle la sortie de la religion coïncide avec la prise de conscience de ce qu'elle avait toujours camouflé : la condition politique, la conflictualité structurante de toute société, l'invariable social primordial à laquelle la condition historique ne fait que donner des formes différentes.

Le point de départ de l'investigation de Marcel Gauchet est la définition de la religion dans les sociétés pré-étatiques, religion « première », religion « pure », idéal type à la lumière duquel toutes les religions historiques ne sont que des sortes de compromis. La religion, écrit-il, est la « dépossession radicale des hommes quant à ce qui détermine leur existence<sup>21</sup> ». Fidèle aux concepts fondamentaux de la littérature ethnographique, Gauchet pense la religion primitive *via* l'opposition du profane et du sacré. Ainsi, la religion primitive est tout entière vouée à la satisfaction et à la conservation du temps passé (temps sacré) par l'intermédiaire des rites dans le temps présent (temps profane). Pour l'homme religieux, l'expérience du monde réel se ramène entièrement à l'expérience d'une *dette* envers les fondateurs de l'ordre du cosmos. Ce conservatisme radical qu'est la religion assume une fonction politique essentielle. En effet, l'altérité sacrale du fondement social permet à la société de trouver un point d'équilibre par lequel le conflit à l'intérieur de la communauté est résorbé. D'un côté, la domination des chefs sur leurs dominés est minimisée en faisant de ceux-ci de simples relais du sacré dans la communauté ; l'altérité radicale du fondement représentée dans la religion primitive pose tous les êtres comme égaux, malgré les effets de prestige et de gloire<sup>22</sup>. D'un autre côté, elle fonde les règles apparemment immuables sur lesquelles les rapports sociaux s'appuient, permettant de désamorcer systématiquement les conflits par la solution que les ancêtres imaginaires proposent.

Tout le propos de Marcel Gauchet dans *Le désenchantement du monde* consiste à recomposer le parcours symbolique qui a mené les sociétés humaines du stade pré-étatique jusqu'à la démocratie moderne telle que nous la connaissons désormais où, selon lui, la religion ne serait plus structurante. Au sujet de ce parcours, nous ne dirons rien ici ; plaçons-nous directement au niveau de la ligne d'arrivée, à savoir la spécificité historique de la démocratie moderne.

Dans la dernière partie du *Désenchantement du monde*, Gauchet commence par expliquer que la démocratie moderne s'est construite contre la souveraineté absolue. Selon Gauchet, la souveraineté absolue est à son tour une réponse à la crise théologique engendrée par la critique protestante selon laquelle nulle médiation entre Dieu et le croyant n'est possible du fait de l'éloignement de Dieu. Les hommes étant abandonnés par Dieu, et le roi étant un homme, la place de celui-ci cesse d'être pensée comme un maillon dans la hiérarchie des êtres allant du plus grand au plus petit. La coupure du Roi d'avec Dieu fait qu'il devient celui autour duquel la société se forme. Avec la souveraineté absolue, l'ordre social n'est plus créé du dehors par le divin, mais est tenu tout entier par la volonté d'un seul homme : le corps du roi devient le pôle symbolique autour duquel la société s'organise, une altérité immanente lui permettant de se doter d'une représentation d'elle-même comme un tout, mais aussi des ressources lui permettant de s'auto-transformer — ce retour de la société sur elle-même grâce à la figure du roi est ce que Gauchet appelle « la naissance de la société comme sujet à travers le mécanisme de la représentation<sup>23</sup> ». Toutefois, une telle institution du social, par son exigence de trouver dans le mécanisme représentationnel le moyen de se saisir comme un tout fusionnel, ne rompt aucunement avec le réflexe religieux de médiation des conflictualités. C'est en réponse à cette nouvelle morphologie du social que la démocratie se construit, tout en développant les prémisses qu'elle contient<sup>24</sup>. On voit, soit dit en passant, se dessiner une opposition radicale entre Gauchet et Debray sur la question de la représentation. Nous avons vu que pour Debray, la représentation est celle du « point fixe » sacré, son incarnation dans le souverain, indépendamment de son contenu. Debray propose une conception clairement anhistorique et formaliste

de la représentation : ce qui est représenté n'a aucune importance, ce qui importe est de connaître le mécanisme de la représentation, par où l'on détient la preuve de l'incomplétude de corps social. Gauchet, au contraire, voit dans la représentation une invention *moderne*, et est inséparable de son contenu, à savoir la souveraineté populaire dont la logique se développe de Hobbes à Rousseau.

La démocratie conserve le projet proprement moderne de se penser comme sujet, mais rompt avec l'idéal de la « conjonction avec soi ». Plutôt que de chercher la liaison par la médiation du souverain, la démocratie va chercher à se fonder dans la division, tout en demeurant une unité. Selon Gauchet un tel tour de force passe par un nouveau « déplacement » du pôle symbolique par le détour duquel la société s'institue : le tout Autre passe dans le mondain. C'est parce que les modernes ont incorporé l'altérité religieuse au sein de leurs affaires les plus profanes qu'a été rendue possible la démocratie. Il y donc une certaine continuité entre les sociétés religieuses et les sociétés démocratiques puisque les deux se fondent par la référence à une altérité ; mais il y a aussi rupture, puisque sa nouvelle localisation modifie profondément sa nature<sup>25</sup>. La conséquence d'une telle transformation est l'apparition du trait le plus distinctif de la démocratie, c'est-à-dire l'institutionnalisation du conflit, subvertissant ainsi toute une histoire des rapports entre le politique et la religion où celle-ci visait à recouvrir les conflits et en minimiser les effets. Désormais tournées vers le futur et débarrassées des altérités sacrales, les sociétés démocratiques sont autonomes, c'est-à-dire qu'elles se sont débarrassées des médiations religieuses dans le cadre de la gestion des conflictualités.

Gauchet, contrairement à Debray, soutient donc que l'autonomie politique est non seulement réalisable, mais aussi d'ores et déjà réalisée<sup>26</sup>. Ce diagnostic rentre en contradiction avec d'autres interprétations de la modernité, en particulier celles qui font du capitalisme une forme nouvelle d'hétéronomie religieuse. On ne pourra dans le cadre de ce texte discuter en détail la thèse selon laquelle le capitalisme serait une *religion séculière décentralisée*<sup>27</sup>, mais puisqu'elle met directement en jeu les thèses de Gauchet sur les rapports entre capitalisme, démocratie et désenchantement, il

nous paraît opportun de la mettre brièvement en rapport avec celles de Gauchet.

Dans le cadre de la philosophie de l'histoire de Marcel Gauchet, les développements conjoints du capitalisme et de la démocratie ne posent aucun problème à la thèse selon laquelle la modernité se déploie comme désenchantement et institution autonome de la société, puisqu'il conçoit les rapports entre capitalisme et démocratie comme complémentaires<sup>28</sup>. Gauchet reconnaît bien entendu les enjeux sociaux considérables qu'implique le capitalisme, mais selon lui, celui-ci doit être analysé comme une des dimensions de l'émancipation des hommes de l'institution religieuse de la société. Il faut, pour bien comprendre cela, souligner que la notion de démocratie chez Marcel Gauchet ne s'épuise pas dans celle d'un ensemble de règles formelles qui réglerait la répartition et l'accès au pouvoir, mais se confond plutôt avec l'idée d'un certain type d'institution symbolique, d'un certain type d'accès au réel structurant les actions individuelles. La démocratie est un *régime* au sens fort du terme<sup>29</sup>, c'est-à-dire, en plus d'une organisation du pouvoir, une façon historiquement située de vivre dont la spécificité est de se rapporter au monde de façon active et dynamique, ouverte sur le futur et consciente d'elle-même — raison pour laquelle la notion de démocratie chez Gauchet, ainsi que pour Lefort avant lui, est presque identique à celle de modernité. Le capitalisme est dès lors à comprendre selon Gauchet comme une facette de ce type inédit d'institution symbolique du social dans l'histoire : le capitalisme est dans le domaine de la production, de la consommation et des échanges, l'instanciation en même temps qu'un moteur de cette nouvelle façon de se rapporter au monde. Ainsi, pour Gauchet, confondre religion et capitalisme, faire du capitalisme et de ses composantes la nouvelle religion des temps modernes, c'est confondre tous les registres, ce n'est comprendre ni ce qu'est la modernité démocratique, ni le capitalisme, ni la religion : conforme à la modernité démocratique, le capitalisme introduit dans le domaine économique la dimension de l'historicité, de la prise en charge consciente de son devenir en vue du futur et à la lumière du passé, tandis que la religion, comme nous l'avons vu ci-dessus, est tout le contraire selon lui, à savoir l'interdiction de l'innovation,

la stagnation dans le présent en fonction de normes prescrites par l'au-delà. La démocratie et le capitalisme sont pour Gauchet deux dimensions complémentaires de la mutation moderne du rapport symbolique au monde et doivent être compris ensemble plutôt que l'un contre l'autre. Gauchet est bien entendu sensible au fait que la modernité se crée ses nouvelles « altérités », dont certains éléments du capitalisme tels que la technique, la croissance ou le progrès peuvent servir d'exemples, mais rappelle toujours qu'elles trouvent leurs conditions de possibilité dans un univers symbolique où ont été reconnues l'autofondation et l'historicité du monde humain<sup>30</sup>.

Selon Marcel Gauchet, le désenchantement du monde a donc pour vertu d'ouvrir l'homme à son propre pouvoir créateur, composante centrale de son autonomie. Un tel constat est partagé par les critiques « théologiques<sup>31</sup> » du capitalisme, mais, à la différence de Gauchet, celles-ci mettent en lumière le fait que *la création humaine dans le capitalisme est orientée contre elle-même*. Si le désenchantement et l'autonomie supposent l'ouverture de l'homme à la création, ils auraient aussi pour conséquence l'orientation de la création humaine de telle sorte qu'une interrogation radicale sur celle-ci ne serait plus accessible. C'est selon nous cette orientation de la création dans le capitalisme, orientation unilatérale vers la croissance et le développement dont un des corrélats est le sacre de la technique, qui a pu faire dire à de nombreux de ses critiques qu'il est une sorte de religion séculière. Ce qui structure les critiques de type « théologique » du capitalisme (au-delà d'une tendance à faire du religieux une structure transhistorique du social) n'est pas tant l'idée selon laquelle le capitalisme introduirait de nouvelles personnalités divines et un nouvel au-delà, que celle selon laquelle le désenchantement du monde est compensé par une orientation absurde de la créativité humaine qui, en dotant d'un sens économique toute activité humaine et inversant dès lors fins et moyens de l'existence, *ressemble* à quelque chose comme de la religion. S'il est ainsi parfois question dans de telles critiques de « nouveaux dieux » ou de « nouvelles idoles », l'argent, la technique, le progrès ou la croissance, c'est toujours pour faire remarquer en contrepartie qu'il ne s'agit que de simulacres de divinités dans un contexte désenchanté, simulacres d'autant plus

absurdes que les sacrifices qui leur sont offerts, parfois humains, ne seraient compensés par aucune immortalité, aucune grâce, aucune rédemption.

Ce problème que pose le capitalisme, à savoir l'orientation absurde de la création humaine selon une direction qui n'est ni consciemment voulue, ni négociable, est précisément un des problèmes centraux qui anime la pensée de Cornélius Castoriadis, vers laquelle nous nous tournons finalement.

*Cornélius Castoriadis : la démocratie comme projet et comme autonomie relative*

Selon nous, la pensée castoriadienne possède les ressources afin de se dégager de la thèse d'un fondement transhistorique de la société sur le religieux, mais aussi d'éviter de faire de la démocratie parlementaire la fin heureuse d'une histoire plurimillénaire des rapports entre politique et religieux. Elle permet aussi de rendre compte de la thèse selon laquelle le capitalisme serait une forme nouvelle d'hétéronomie religieuse et propose un concept substantif de démocratie comme autonomie relative qui fait de la démocratie un *projet* sans cesse à renouveler plutôt qu'un régime dont la définition s'épuise dans celle du constitutionnalisme juridique. Dans ce qui suit, nous allons (1) nous concentrer sur la façon dont Castoriadis entend l'institution religieuse de la société comme modalité hétéronome de l'institution imaginaire de la société; pour ensuite (2) montrer en quoi la démocratie telle que l'entend Castoriadis est une institution autonome de la société.

La théorie de l'institution imaginaire de la société exposée dans *L'institution imaginaire de la société* et esquissée dans «Marxisme et théorie révolutionnaire<sup>32</sup>» est une forme très marquée d'holisme sociologique, principalement dirigée contre l'épistémologie monocausaliste et empiriste du marxisme orthodoxe, mais aussi contre toutes les théories sociologiques qui font des institutions sociales une réponse à des besoins invariants (fonctionnalisme et structuralisme), et dont l'enjeu est le renouvellement du projet d'autonomie dans le contexte de l'effondrement du mouvement ouvrier. Ce que reproche principalement Castoriadis au marxisme,

fonctionnalisme et structuralisme est qu'ils sont incapables de comprendre la multiplicité historique des sociétés humaines en soutenant notamment que toute institution sociale vise à répondre à un besoin, et que les besoins sont les mêmes pour toute l'humanité<sup>33</sup>. Nous ne nous engagerons pas dans le débat consistant à savoir s'il est possible d'attribuer une telle thèse aux marxisme, structuralisme et fonctionnalisme ; un des artifices rhétoriques de Castoriadis consiste bien souvent à se doter d'un épouvantail afin de faire valoir ses propres conceptions. Contre ces théories, Castoriadis soutient la thèse selon laquelle l'être est indéterminé<sup>34</sup>, et que l'activité humaine (mais aussi le règne du vivant dans son ensemble) ne peut se comprendre en dehors de la notion de *création*, laquelle aurait été occultée par le rationalisme philosophique (la « pensée héritée ») duquel procèdent les trois familles théoriques incriminées. C'est ainsi qu'afin de rendre compte de la multiplicité des sociétés à travers l'histoire, Castoriadis est amené à postuler un *imaginaire radical*, source de la multiplicité historique des formes sociales et de la création humaine. La théorie de l'institution imaginaire de la société de Castoriadis est un holisme sociologique prononcé en ceci qu'elle postule comme source des institutions sociales une puissance ontologique irréductible à l'action des individus composant la société<sup>35</sup>. L'imaginaire radical est la faculté créatrice du social-historique, lui-même compris comme « strate » irréductible de l'être. La difficulté que le lecteur rencontre face à cette notion d'imaginaire radical consiste à ne pas substantialiser la notion castoriadienne d'imaginaire, c'est-à-dire à ne pas chercher à rattacher l'imaginaire à un sujet quelconque, malgré la thèse selon laquelle le social-historique est un domaine de l'être à part entière.

Le mécanisme réel de l'institution de la société selon Castoriadis est la production de « significations imaginaires sociales » par l'imaginaire radical de la société<sup>36</sup>. Ces significations imaginaires sociales sont comprises comme sortes d'idées directrices qui informent la totalité des activités pour une société donnée. Castoriadis explique que l'unité de la multiplicité des institutions<sup>37</sup> d'une société est donnée dans leur participation aux significations imaginaires, qu'elles en sont des *incarnations*. Par exemple, l'utilisation d'un

outil, déterminée en partie par la nécessité de transformer la nature en vue de la satisfaction d'un besoin, est significativement orientée en fonction des significations imaginaires propres à la société dans laquelle il est utilisé: on n'utilise pas un marteau chez les Yanomamis pour les mêmes raisons et de la même façon que chez un cordonnier de la Suisse calviniste de la fin du XVI<sup>e</sup> siècle, et ce en raison de la différence des significations imaginaires qui structurent inconsciemment et distinguent ces deux sociétés.

Selon Castoriadis, la religion est une modalité de l'institution imaginaire de la société où l'autocréation de la société par elle-même (la production immotivée de significations imaginaires sociales par l'imaginaire radical) est occultée et attribuée à des êtres fictifs. Castoriadis estime que la religion est le dispositif social de l'hétéronomie en ceci qu'elle camoufle aux individus qui composent la société le fait que ce sont eux qui produisent les conditions de leur existence<sup>38</sup>. Plus précisément, la religion est comprise par Castoriadis comme l'institué qui vient verrouiller et cadenasser le mouvement créateur de l'imaginaire radical; la religion, création humaine qui ne se reconnaît pas comme telle, entrave le déploiement de la création. La définition de la religion par Castoriadis hérite clairement de celle donnée par Marx comme aliénation à la suite de la critique par Feuerbach de Hegel. Feuerbach définissait l'aliénation comme le processus par lequel un contenu de conscience est extériorisé puis objectivé de telle sorte qu'il devient une puissance étrangère pour celui qui en est l'origine<sup>39</sup>. Ce qui distingue toutefois Castoriadis de Marx ou de Feuerbach est que ce qui est aliéné dans la religion n'est pas la conscience (Feuerbach) ou le travail social (Marx), mais la création.

La théorie castoriadienne de l'institution imaginaire de la société aboutit donc à un découplage de la notion d'imaginaire et de religion en faisant de la religion une modalité de l'institution imaginaire de la société. S'il y a un invariant transhistorique du social, ce n'est, selon Castoriadis, ni le religieux (Debray), ni le politique (Gauchet), mais l'imaginaire, puissance sans sujet et source de la création dans le domaine du social-historique en particulier et du vivant en général. C'est avec cet arrière-fond théorique que Castoriadis identifie

le capitalisme et la religion. Le capitalisme est défini comme «quasi-religion», car il en possède les traits structuraux, l'occultation de la création, mais se distingue des religions traditionnelles sur son plan formel, à savoir la distinction entre visible et invisible, naturel et surnaturel<sup>40</sup>. Dans le capitalisme, ce sont les sciences et les techniques qui sont, plutôt que les dieux et les héros, les représentants et vecteurs de l'hétéronomie, car c'est à travers le prisme rationaliste des sciences et des techniques que l'intégralité du monde vécu est perçue au détriment de ses potentialités créatrices. Ceci fait dire à Castoriadis que l'imaginaire du capitalisme n'a pas de «chair propre<sup>41</sup>», recoupant par là la critique francfortienne de l'expansion de la rationalité instrumentale sous le capitalisme<sup>42</sup>.

Ce découplage de l'imaginaire et de la religion fournit la base sur laquelle se déploie sa théorie de la démocratie comme autonomie relative. L'analyse de la religion, modèle paradigmatique de l'institution hétéronome de la société, permet à Castoriadis de définir l'hétéronomie comme, d'une part, l'impossibilité de se reconnaître comme l'origine de sa propre loi, d'autre part, comme autonomisation de l'institué et verrouillage du moment instituant. De ce point de vue, une société est autonome à condition qu'elle se reconnaisse comme l'origine de sa propre loi, conformément au mécanisme réel de l'institution imaginaire de la société, mais aussi à condition qu'elle mette au diapason le moment proprement instituant du social avec l'institué<sup>43</sup>. En paraphrasant Trotsky, on peut dire que l'autonomie castoriadienne est une révolution permanente. L'autonomie comprise en ce sens est selon Castoriadis nécessairement démocratique, car ce n'est qu'à partir du moment où sont partagées collectivement les décisions sur les fondements de la vie en commun et qu'est mis à jour et approprié par tous le moment créateur de l'institution qu'une société est autonome<sup>44</sup>. Elle doit être dite «relative», puisqu'elle ne se comprend que comme un rapport particulier à l'institution où la possibilité de sa subversion, de sa remise en question par tous au profit de tous, est ouverte sans limites. L'hétéronomie advient dans le rapport à l'institution, mais n'est pas ce rapport.

L'autonomie est donc un projet, et elle doit toujours être remise en jeu afin qu'elle puisse être conforme à sa nature. Pour qu'un tel

projet puisse se réaliser, il faut selon Castoriadis commencer par se débarrasser d'une conception technicienne de l'action politique qui fait de celle-ci un simple moyen en vue d'une fin qui serait intégralement déterminée par le savoir<sup>45</sup>. La politique authentique est une *praxis*, un faire non-technique qui vise l'autonomie de tous, sans pour autant se doter d'une représentation exhaustive de ce qu'elle devrait être. La *praxis* est le faire non-technique de la liberté. En définissant le projet démocratique comme *praxis*, Castoriadis assume ainsi une conception révolutionnaire de la politique (l'institué étant rabattu sur le moment instituant), mais cherche à se distinguer de la conception léniniste de la révolution selon laquelle une avant-garde éclairée possédant une théorie prétendument exhaustive peut produire des bouleversements sociaux grâce à son application. Pour Castoriadis, la *praxis* n'est pas un *plan*, et c'est, en partie, la confusion entre ces deux modalités de l'action qui est la cause de l'échec du mouvement ouvrier et de sa pétrification dans la bureaucratie stalinienne<sup>46</sup>.

On voit en outre toute la différence avec Gauchet, qui pense que l'autonomie est déjà réalisée et qu'elle est le fruit du développement politico-religieux de l'occident.

### *Conclusion*

Nous avons ouvert ce texte en évoquant le problème classique du totalitarisme comme religion séculière. Nous avons montré que la thématique de la religion séculière peut se comprendre comme une toile de fond sur laquelle se sont dessinées de nouvelles réflexions sur la question du politico-religieux, c'est-à-dire sur la nature du principe instituant du social dans son rapport à la question moderne de l'autonomie collective et individuelle. Nous avons brièvement présenté ce que nous estimons constituer deux théories contraires pour le même problème, celles de Debray et de Gauchet, pour ensuite en indiquer quelques limites. En ce qui concerne la théorie de Régis Debray, nous avons suggéré qu'elle ne parvient pas à rendre compte de certains aspects centraux de la modernisation des sociétés occidentales. En soutenant que toute société se fonde religieusement, la théorie debrayenne est incapable d'expliquer comment les sociétés modernes s'articulent comme des tous

complexes. L'idée selon laquelle toute société se maintient dans le temps et l'espace par référence à un «point sacré» est selon nous extrêmement limitée pour expliquer l'interdépendance fonctionnelle des sphères d'activité dans les sociétés modernes, mais aussi pour comprendre la multiplication des morales individuelles au sein de la société civile. Elle nous semble aussi dangereusement conservatrice, ainsi que nous l'avons fait remarquer. S'agissant de la théorie du désenchantement de Marcel Gauchet, nous avons cherché à mettre en lumière ce qu'a de problématique la thèse selon laquelle la démocratie et le capitalisme seraient les deux facettes d'un même processus de découverte par l'homme de son propre pouvoir créateur. Pour ce faire, nous l'avons rapidement mis en perspective avec les critiques «théologiques» du capitalisme, lesquelles l'interprètent comme une réorientation hétéronome de la créativité humaine en contradiction avec les exigences normatives de la modernité. Nous avons clos notre texte en soumettant l'idée que le concept d'autonomie relative proposé par Castoriadis nous semble plus apte à relever la question du politico-religieux dans les sociétés contemporaines, en particulier dans le contexte d'une diffusion sans précédent de la logique du capitalisme<sup>47</sup>.

- 
1. La dimension religieuse du totalitarisme a été commentée par de nombreux auteurs, et on ne saurait citer toutes les études qui l'ont analysée selon une telle perspective. Voici, néanmoins, quelques études notables à ce sujet : Raymond Aron, *L'Opium des intellectuels*, Paris, Presses pocket, 1986 ; Jacques Ellul, *Les nouveaux possédés*, Paris, Mille et une nuits, 2003 ; Jules Monnerot, *Sociologie du communisme*, Paris, Gallimard, 1949 ; Jean-Pierre Sironneau, *Sécularisation et religions politiques, with a summary in English*, La Haye, Paris, New-York, Mouton, 1982 ; Eric Voegelin, *Les religions politiques*, Paris, Éd. Du Cerf, 1994.
  2. La sociologie contemporaine de la religion distingue deux types principaux de définition de la religion : fonctionnelle et substantielle. Le premier type se concentre sur la fonction sociale qu'assume la religion, indépendamment de son contenu et incluant ainsi des phénomènes qui ne paraissent pas intuitivement «religieux». Le second type se donne en amont certaines caractéristiques précises du fait religieux (par exemple

- la croyance en un arrière-monde). Cf. Camille Tarot, *Le symbolique et le sacré : théories de la religion*, Paris, La Découverte, 2008.
3. Cornélius Castoriadis, *L'institution imaginaire de la société*, Paris, Éditions du Seuil, 1975.
  4. Régis Debray, *Critique de la raison politique*, Paris, Gallimard, 1981.
  5. Marcel Gauchet, *Le désenchantement du monde*, Paris, Gallimard, 1985.
  6. Il importe moins ici de savoir quelle est la signification exacte de l'incomplétude selon Gödel que de rendre compte de la bizarrerie que représente le transfert par Debray d'un concept d'ordre mathématique au domaine du politique. Debray a reconnu personnellement en 1996, après avoir été nommément visé dans «l'affaire Sokal», que cet emprunt aux mathématiques est «métaphorique ou isomorphique» et que l'on ne saurait «assimiler un système politico-social à un système logico-déductif». Cet emprunt est donc à comprendre comme métaphore heuristique, et non comme l'ébauche d'une tentative de réduire le champ politique au champ mathématique (même si le texte de 1985, que nous commentons ici, ne fait rien pour expliciter un tel positionnement «métaphorique»). Cf. Régis Debray «L'incomplétude, logique du religieux?» dans *Bulletin de la société française de philosophie, séance du 27 janvier 1996*, vol. 90, n° 1 (Janvier-Mars 1996), p. 7.
  7. Régis Debray, *Critique de la raison politique*, p. 256. Souligné dans le texte.
  8. *Ibid.*, p. 256.
  9. *Ibid.*, p. 258.
  10. *Ibid.*, p. 259.
  11. *Ibid.*, p. 259.
  12. *Ibid.*, p. 265.
  13. *Ibid.*, p. 266.
  14. *Ibid.*, pp. 264-265.
  15. Le rejet de l'autogestion par Debray doit être mis en perspective. Comme le remarque Christofferson, le conseillisme autogestionnaire français des années 1970-80, dont Castoriadis est un représentant de premier ordre, a été une position politique d'ultra-gauche visant à réaffirmer stratégiquement l'émancipation des travailleurs dans le contexte de l'annexion des partis communistes d'Europe par l'URSS. Pour un intellectuel de gauche à cette époque, refuser l'autogestion signifie soit l'abandon du projet révolutionnaire, soit la collaboration avec l'URSS via les partis communistes traditionnels (ex: Sartre ou Althusser). En intégrant le gouvernement Mitterrand après son emprisonnement,

Debray a clairement fait le choix de l'abandon du projet révolutionnaire. Cf. Michael Scott Christofferson, *Les intellectuels contre la gauche : l'idéologie antitotalitaire en France, 1968-1981*, Marseille, Agone, 2009.

16. Régis Debray, *Critique de la raison politique*, p. 258.
17. *Ibid.*, p. 263.
18. Cette critique a déjà été formulée par Pierre Manent : Pierre Manent, « Régis Debray révolutionne la science », dans *Commentaire*, vol. 4, n° 16 (1981), pp. 592-597.
19. La définition de la philosophie comme projet *critique et démocratique* est une idée proprement castoriadienne, que nous nous approprions dans le cadre de ce développement. Si l'on admet que la philosophie, dès sa naissance, va de pair avec l'exigence normative selon laquelle l'activité humaine doit trouver en elle-même ses propres ressources (l'autonomie), ce que l'enseignement socratique ne semble pas démentir, alors on peut déduire d'une telle exigence l'idée comme quoi la pratique d'une telle activité réflexive n'a de sens qu'à condition qu'elle soit universalisée et universalisable : non seulement la pratique de la philosophie est impossible là où n'est pas disponible l'idée d'une orientation critique de la pensée, mais il est aussi absurde de ne vouloir l'autonomie que pour soi-même (doublement absurde en vérité, puisque, d'une part, il m'est impossible d'être libre dans un monde où personne ne le serait et, d'autre part, parce qu'interdire à l'autre l'ouverture à la pensée critique est en contradiction flagrante avec l'idée même de cette ouverture). Que la philosophie ne se soit pas pensée à son origine comme « démocratique » ne signifie pas pour autant que la démocratie ne se trouve pas en aval de l'exercice du discours philosophique. Cf. Cornélius Castoriadis, « La polis grecque et la création de la démocratie », dans *Les carrefours du labyrinthe 2. Les domaines de l'homme*, Paris, Seuil, 1999, pp. 325-382.
20. Cf. Daniel Tanguay et Gilles Labelle, *Vers une démocratie désenchantée ? Marcel Gauchet et la crise contemporaine de la démocratie libérale*, Québec, Fides, 2013.
21. Marcel Gauchet, *Le désenchantement du monde*, p. 13.
22. *Ibid.*, p. 16.
23. *Ibid.*, p. 249.
24. *Ibid.*, p. 252.
25. *Ibid.*, p. 278.
26. Il faudrait toutefois rendre justice à Marcel Gauchet, et ne pas en faire un apôtre de la démocratie parlementaire tel que nous avons tendance à le faire. Marcel Gauchet, en particulier dans *La démocratie contre*

*elle-même*, est sensible aux contradictions internes auxquelles aboutit le processus de sortie de la religion. Il montre en effet que le déploiement de la logique de l'autonomie débouche de nos jours à une hypertrophie de son moment juridique au détriment de son moment politique, conduisant à une véritable dissolution du corps social. Ainsi, pour Gauchet, la tâche authentique de la philosophie politique est de remettre au jour le fondement politique du social, plutôt que d'élaborer une doctrine normative de la justice. Cf. Marcel Gauchet, «Les tâches de la philosophie politique», dans *La condition politique*, Paris, Gallimard, 2005, pp. 505-557,

27. Pour une critique théologique du capitalisme: Hugo Assmann et Franz J. Hinkelammert, *L'Idolâtrie du marché: critique théologique de l'économie de marché*, Paris, Les Editions du Cerf, 1993; Robert H. Nelson, *Reaching for Heaven on Earth: The Theological Meaning of Economics*, Savage, Rowman & Littlefield Publishers, 1991; Robert H. Nelson, *Economics as Religion: From Samuelson to Chicago and beyond*, University Park, Pennsylvania State University Press, 2001; Jung Mo Sung, *The Subject, Capitalism, and Religion: Horizons of Hope in Complex Societies*, New York, Palgrave Macmillan, 2011.

Pour une critique sociologique du capitalisme comme religion: Paul Dumouchel et Jean-Pierre Dupuy, *L'enfer des choses: René Girard et la logique de l'économie*, Paris, Seuil, 1979; Jean Pierre Dupuy, *L'avenir de l'économie: sortir de l'économystification*, Paris, Flammarion, 2012. La littérature marxiste est également pleine de comparaisons entre le capitalisme et la religion, à commencer par Marx au niveau de sa théorie du fétichisme de la marchandise. À part Marx, on peut citer par exemple: Lafargue, Paul, *La religion du capital*, Castelnau-le-Lez, Climats, 1995; Jappe, Anselm, *Les aventures de la marchandise: pour une nouvelle critique de la valeur*, Paris, Denoël, 2003; Benjamin, Walter, «Le capitalisme comme religion», dans *Fragments philosophiques, politiques, critiques, littéraires*, pp. 111113, Paris, Presses universitaires de France, 2000.

28. Nous nous rapportons pour ce qui suit à: Marcel Gauchet, *L'avènement de la démocratie. I, La révolution moderne*, Paris, Gallimard, 2007.
29. Définition que Gauchet hérite selon nous de Lefort. On trouvera une définition de la notion de régime selon Lefort dans le texte suivant: Claude Lefort, «Avant-propos», dans *Essais sur le politique: XIXe-XXe siècles*, Paris, Seuil, 1986), p. 9.

30. Cf. les textes de Gauchet sur le religieux après la religion. Marcel Gauchet, *Un monde désenchanté? : débat avec Marcel Gauchet sur « le Désenchantement du monde »*, Paris, Cerf, 1988; Luc Ferry et Marcel Gauchet, *Le religieux après la religion*, Paris, Grasset, 2004; Marcel Gauchet, « Le religieux après la religion » dans *Le désenchantement du monde*, pp. 292-303.
31. Le lecteur remarquera que nous faisons ici un usage générique du concept de « critique théologique ». Nous admettons comme critique « théologique » du capitalisme celles qui mettent en avant le problème proprement axiologique que suscite selon elles le capitalisme, à savoir une « inversion des moyens et des fins », problème mis en perspective avec la thématique plus générale de la sécularisation et de la mort de de Dieu. Une critique théologique met en outre en avant le problème du sens « existentiel » du capitalisme plutôt que des problèmes d'ordre sociaux, de répartition des richesses, etc.
32. L'essai intitulé « Marxisme et théorie révolutionnaire » est désormais la première partie de *L'institution imaginaire de la société*. Il s'agissait néanmoins d'un essai indépendant, rédigé entre 1964 et 1965, c'est-à-dire 10 ans avant la première publication de *L'institution*. Il s'agit d'une synthèse de sa critique du marxisme débouchant sur les prémisses de sa théorie de l'institution imaginaire de la société. Warren Breckman, « From the Symbolic Turn to the Social Imaginary: Castoriadis's Project of Autonomy », dans *Adventures of the symbolic: Post-Marxism and Radical Democracy*, New-York, Columbia University Press, 2013, pp. 96-138.
33. Cornélius Castoriadis, *L'institution imaginaire de la société*, pp. 173-174.
34. Cornélius Castoriadis, « L'imaginaire: la création dans le domaine social-historique », dans *Les carrefours du labyrinthe. 2. Domaines de l'homme*, Paris, Seuil, 1990. pp. 272-274.
35. Cornélius Castoriadis, *L'institution imaginaire de la société*, pp. 160-161.
36. « Les significations centrales ne sont pas des significations « de » quelque chose — ni même, sinon en un sens second, des significations « attachées » ou « référées à » quelque chose. Elles sont ce qui fait être, pour une société donnée, la co-appartenance d'objets, d'actes, d'individus en apparence les plus hétéroclites. Elles n'ont pas de « référent »; elles instituent un mode d'être des choses et des individus comme référé à elles. Comme telles, elles ne sont pas nécessairement explicites pour la société qui les institue. » *Ibid.*, p. 525.
37. La définition du concept d'institution par Castoriadis est très englobante, puisqu'il regroupe aussi bien les manières de faire que de penser, mais

- aussi les individus. L'extension castoriadienne du concept d'institution recoupe à peu près tout ce qui peut se trouver dans une société.
38. Cornélius Castoriadis, «Institution de la société et religion», dans *Les carrefours du labyrinthe. 2, Domaines de l'homme*, pp.462-463.
  - 39 Ludwig Feuerbach, *L'essence du christianisme*, Paris, F.Maspero, 1982, p. 149.
  - 40 Cornélius Castoriadis, *op. cit.* pp.463-464.
  - 41 Cornélius Castoriadis, *Ibid.*, pp.235-240.
  - 42 Pour une synthèse de la critique francfortienne de la rationalité instrumentale : Jürgen Habermas, *Théorie de l'agir communicationnel : rationalité de l'agir et rationalisation de la société*, Paris, Fayard, 1987, pp. 371-402.
  - 43 Cornélius Castoriadis «Pouvoir, politique, autonomie», dans *Les carrefours du labyrinthe. 3, Le monde morcelé*, Paris, Seuil, 1990, p. 156.
  - 44 *Ibid.*, pp. 166-167.
  - 45 Cornélius Castoriadis, *L'institution imaginaire de la société*, p. 107.
  - 46 Cornélius Castoriadis, *La société bureaucratique. 1. Les rapports de production en Russie*, Paris, U.G.E, 1973, p. 19
  - 47 Je souhaite remercier Benoît Côté qui a suivi de près l'élaboration de ce texte, notamment à l'occasion de la présentation publique de ses différentes moutures.



# Utilité scientifique des classes d'espèces : orthodoxie et originalité de Leibniz dans les *Nouveaux Essais*

LÉA DEROME, *Université de Montréal*

À première vue, la position de Leibniz quant au statut des espèces peut sembler conservatrice : les espèces existent dans la nature et nous sont connues par une différence spécifique. Il n'y a rien là de bien novateur dira-t-on ; ces thèses pourraient tout aussi bien être celles de Porphyre. Le contexte philosophique et intellectuel dans lequel écrit Leibniz partage cependant bien peu avec celui qui a vu fleurir le néo-platonisme : la philosophie des écoles est virulemment critiquée, la méthode expérimentale du *Novum Organum* de Bacon gagne en popularité, le finalisme est progressivement remis au profit d'un mécanisme intégral... Bref, s'intensifiait alors ce qu'on qualifiera plus tard, à tort ou à raison, de « révolution scientifique ». Et celle-ci s'accomplissait au nom d'une conception de la science qui se voulait radicalement nouvelle, c'est-à-dire libérée de l'autorité d'Aristote et de la tradition scolastique qui avaient jusqu'ici déterminé les modèles du savoir et de la science. Difficile, donc, d'être à la page en défendant la pertinence des prédicables de genre et d'espèce. C'est pourtant ce que fait Leibniz, au troisième livre des *Nouveaux essais sur l'entendement humain*<sup>1</sup> (*NE*), en réponse à John Locke qui dans *An Essay concerning Human Understanding*<sup>2</sup> (*Essay*) avait fait des espèces des conceptions de l'esprit humain sans correspondants extra-mentaux dans les êtres.

Pour expliquer l'attachement de Leibniz à une position de facture plus orthodoxe, différents motifs peuvent être invoqués, notamment le fait qu'en répliquant aux arguments anti-essentialistes de Locke, Leibniz consolide la définition canonique de l'être humain comme animal rationnel. C'est une intention que les commentateurs n'ont pas manqué de souligner, jusqu'à en faire la principale motivation de

Leibniz dans sa critique de Locke. Nicholas Jolley est de cet avis : «For Leibniz, the really disturbing element in Locke's theory of nominal essences is what he takes to be its degrading consequences for the definition of man<sup>3</sup>». Jolley est suivi en cela par Susanna Goodin<sup>4</sup> et Brandon Look<sup>5</sup>.

Que Leibniz tiennne à ce que l'être humain se distingue du reste de la création en termes d'essence, cela ne fait aucun doute. Il faut pourtant se garder de réduire toute la démarche leibnizienne à cet unique motif. Car, de l'avis de Leibniz, les notions de genre et d'espèce font plus que simplement préserver à l'être humain son statut privilégié, mais constituent, pour la science, des catégories logiques et taxinomiques de première importance, et dont l'utilité se vérifie aux progrès qu'elles permettent dans la connaissance de la nature et à l'intelligibilité additionnelle qu'elles jettent sur les phénomènes empiriques. La position de Locke n'entraîne donc pas uniquement des conséquences dégradantes pour l'être humain, mais constitue également une limite excessive, aux yeux de Leibniz, tracée à la connaissance humaine.

Aussi souhaitons-nous, après avoir restitué les thèses principales de nos deux auteurs, insister, dans le présent article, sur la dimension non pas ontologique, mais épistémologique de la position de Leibniz. Nous tâcherons notamment de montrer que, depuis les prédicables de genre et d'espèce, Leibniz en vient à fonder un programme scientifique novateur à bien des égards, de sorte à relativiser cette autre affirmation de Jolley : «Yet, though Leibniz advances arguments that have modern parallels, the motivation of his critique is highly traditionnal<sup>6</sup>». Si les motivations sont traditionnelles, bien qu'en partie seulement, nous verrons que le résultat final a toutes les marques de l'originalité.

### *1. Conceptualisme et théorie de l'abstraction dans l'Essay<sup>7</sup>*

Rappelons d'abord que pour Locke les mots ne signifient immédiatement rien d'autre que les idées du locuteur. Ce qui veut dire que les termes généraux doivent être signes d'idées générales. Ces dernières, nous les fabriquons au niveau de la faculté réflexive, lorsque, grâce à notre capacité d'abstraction puis de combinaison,

nous dépouillons nos idées simples, d'abord reçues de l'expérience, de ce qui les particularise : les conditions de temps, de lieu, ou toute autre composante circonstancielle. Il est ensuite libre à nous de les combiner entre elles, de manière à ce que ces idées abstraites aboutissent à représenter une pluralité d'individus. Le résultat de ces opérations est une idée complexe, qui totalise plusieurs idées simples, et à partir de laquelle est formée l'essence nominale, indépendamment de l'essence réelle.

L'essence réelle est ce qui constitue le fondement insensible et la cause des propriétés observables retenues dans les essences nominales<sup>8</sup>. Toutefois, comme la constitution interne (corpusculaire) des êtres ne peut faire l'objet d'une idée simple (elle n'est jamais directement intelligible), les essences réelles nous sont rigoureusement inconnues et, par suite, le lien causal entre ces essences et les qualités observables l'est tout autant. C'est somme toute sans égard pour les essences réelles, de toute manière inconnaissables, que l'entendement forme les essences nominales, regroupant les individus selon les propriétés qu'il juge les plus déterminantes, puis annexant à ces classes générales des noms eux aussi généraux. Cette généralité, cependant, c'est celle des signes (mots et idées), et elle tient au fait que les humains ont besoin de s'élever au-dessus du particulier pour fabriquer un langage efficace et se comprendre entre eux<sup>9</sup>. Les classes générales d'espèce et de genre, idées complexes par excellence<sup>10</sup>, n'ont donc pas de réalité extra-mentale : ce ne sont que des représentations et des significations abstraites de plusieurs particuliers, rien de plus que des « inventions » et des « créatures » de l'entendement<sup>11</sup>. Aussi est-ce l'entendement qui introduit la généralité dans le monde, qui, lui, à en juger par la sensation, n'est pas déjà organisé en classes, mais présente toujours des êtres singuliers<sup>12</sup>.

Précisons encore que, tous singuliers qu'ils soient, les individus affichent néanmoins des ressemblances entre eux, surtout les êtres vivants, plantes et animaux<sup>13</sup>. Or, c'est parce qu'elle se fonde sur ces ressemblances que l'essence nominale jouit d'une objectivité relative<sup>14</sup>. Il y a certes ce travail de l'esprit (*workmanship of the understanding*), qui sélectionne les traits jugés déterminants, mais on n'invente pas de toutes pièces les essences nominales. Grâce à

elles, l'esprit unifie un amas d'idées simples co-existantes, et qui sont habituellement liées, à un niveau phénoménal, dans l'expérience. Ce processus cognitif n'est donc pas aléatoire, puisqu'il se fonde sur des ressemblances et des régularités effectivement observées<sup>15</sup>. L'argument de Locke est plutôt de dire qu'il est strictement humain et donc que les classes générales de genre et d'espèce qui en résultent sont entièrement de notre fabrication, la position adverse voulant qu'elles nous soient en quelque manière dictées par la nature.

## 2. *Équivocité du vocable d'espèce*

Parmi les nombreux reproches que Leibniz adresse à Locke, le principal concerne l'ambiguïté que comporte le terme d'*espèce*, et dont l'auteur de l'*Essay* ne se serait pas suffisamment avisé. Pour Leibniz, il convient en effet de soigneusement distinguer l'espèce prise au sens physique de l'espèce prise au sens mathématique. Dans le cas de l'espèce mathématique, la moindre différence conceptuelle suffit. Ainsi deux êtres qui ne sont pas tout à fait semblables ne seront pas de même espèce, ce qui laisse entrevoir la possibilité d'une infinité de genres et d'espèces mathématiques<sup>16</sup>. Et c'est effectivement la perspective adoptée par Locke, lequel estime que les marques extérieures doivent être articulées avec la nature intérieure de façon si étroite que lorsque deux êtres n'affichent pas exactement les mêmes propriétés, alors c'est qu'ils ne partagent pas non plus la même structure interne: «[It] is as impossible, that two Things, partaking exactly of the same real *Essence*, should have different Properties, as that two Figures partaking of the same real *Essence* of a circle should have different Properties<sup>17</sup>».

Or, pour Leibniz, on ne saurait concevoir les différences spécifiques dans le domaine empirique sur le modèle des distinctions logiques ou mathématiques sans accumuler les invraisemblances. Considérant que les êtres physiques ont une existence temporelle marquée par le mouvement, un même individu pourrait successivement cumuler une pluralité d'appartenances spécifiques, la moindre altération étant susceptible de le faire changer d'espèce. De plus, comme deux individus physiques ne seront jamais en tous points semblables, chacun, à lui seul, pourrait constituer une dernière

espèce logique. Dans l'hypothèse où l'on respecte la même rigueur qui prévaut en mathématique, le nombre d'espèces physiques serait dès lors équivalent au nombre total d'individus et l'appartenance spécifique, chez un même individu, connaîtrait un mouvement perpétuel. Le concept d'espèce deviendrait ainsi parfaitement caduc au plan épistémique, en perdant tantôt sa généralité, tantôt sa fixité, deux qualités qui, sinon, justifient sa contribution à la démarche scientifique. Leibniz redoute ces conséquences, puisqu'il souhaite attribuer aux catégories d'espèce et de genre un rôle épistémique dans le processus d'élaboration de connaissances empiriquement fondées. Aussi est-ce dans cette intention que Leibniz opère une distinction entre espèces mathématiques et physiques, s'octroyant, par la même occasion, une marge de manœuvre appréciable dans la détermination des types physiques, ce que ne permettait pas la rigueur mathématique. Et cette marge de manœuvre, Leibniz entendait bien la mettre à profit, non seulement contre les stratégies argumentatives mobilisées par Locke dans le troisième livre de l'*Essay*, mais aussi afin de renouveler la méthode scientifique en sciences naturelles.

### *3. Utilité scientifique des espèces physiques*

Dans sa critique anti-réaliste, Locke revient à plusieurs reprises sur l'incapacité évidente des taxinomies de l'époque, d'une part à instituer des critères précis pour discriminer les espèces, d'autre part à se justifier devant les phénomènes inclassables. Les arguments de Locke jouent donc sur les deux terrains du normal et du pathologique : non seulement trouve-t-on de nombreux exemples de créatures monstrueuses, mais nous sommes en outre incapables de nous entendre sur ce qui fait la norme. D'où le fait que Locke va, par la suite, abondamment recourir à des exemples de monstruosité ; monstres et hybrides constituant, à ses yeux, des arguments de poids contre le réalisme traditionnel<sup>18</sup> : «The frequent Productions of Monsters, in all the Species of Animals, and of Changelings, and other strange Issues of humane Birth, carry with them difficulties, not possible to consist with this *Hypothesis* [i.e. le réalisme traditionnel]<sup>19</sup>». Or il est intéressant de constater que la position de Leibniz, une fois augmentée du concept d'espèce physique, l'autorise à concéder qu'il

s'agit effectivement d'une tâche fort laborieuse, dans l'état actuel des savoirs, que de trouver un critère de discrimination pour une majorité d'espèces, voire toutes les espèces, hormis l'espèce humaine.

Nous savons en effet que Leibniz correspondait avec un nombre important de savants et qu'il était parfaitement au fait, comme Locke, et même peut-être davantage, des échecs répétés de la taxinomie animale et végétale<sup>20</sup>. En témoigne un passage des *NE*, où Leibniz commente un débat de botanique portant sur le meilleur critère de distinction spécifique :

Les botanistes modernes croient que les distinctions prises des fleurs approchent le plus de l'ordre naturel. Mais ils y trouvent pourtant encore bien de la difficulté, et il serait à propos de faire des comparaisons et arrangements non seulement suivant un seul fondement, comme serait celui que je viens de dire, qui est pris des fleurs, et qui est peut-être le plus tolérable et commode à ceux qui apprennent, mais encore suivant d'autres fondements pris des autres parties et circonstances des plantes, chaque fondement de comparaison méritant des classes à part, sans quoi on laissera échapper bien des genres subalternes, et bien des comparaisons, distinctions et observations utiles<sup>21</sup>.

On voit bien que Leibniz n'est nullement gêné par la recherche à tâtons et un peu confuse des botanistes. De fait, non seulement Leibniz admet-il ici qu'on puisse procéder à des classifications spécifiques selon divers fondements — alors que la tradition essentialiste avait plutôt tendance à n'admettre qu'un seul attribut spécifique fixe par espèce, suffit de penser à la systématique des substances de l'*Isagoge* de Porphyre —, mais il considère en outre que c'est par l'usage de cette diversité de critères, par ces essais et erreurs, que la science progresse dans son approximation de l'ordre naturel.

Et ce qui lui permet cette flexibilité, c'est précisément la distinction entre espèces physiques et mathématiques. Car si nos définitions peuvent être réelles et causales en mathématique — la définition du cercle donne la possibilité de son essence et quiconque la comprend devient alors en mesure de tracer un cercle, etc.<sup>22</sup> —, dans le domaine

physique, nos définitions spécifiques sont « provisionnelles », c'est-à-dire proportionnées à l'état actuel de nos connaissances, et par conséquent elles sont aussi nominales, c'est-à-dire qu'elles se contentent de circonscrire un nombre de réquisits suffisant à la reconnaissance du défini, mais sans en assurer la possibilité ni en épuiser l'essence. Sur ce point, Locke et Leibniz parlent à l'unisson : le savoir spécifique est nominal<sup>23</sup>.

La rupture entre nos deux auteurs survient donc ailleurs, nommément dans le fait que, pour Leibniz, contrairement à Locke, notre incapacité à décrire l'ordre naturel des espèces de manière adéquate — c'est-à-dire au moyen de définitions réelles et d'une connaissance des natures internes — ne suffit pas pour en réfuter l'existence, ni pour révoquer la possibilité que nous avons de le connaître et de le représenter partiellement. La raison en est simple : Leibniz estime que les définitions nominales sont instructives quant à l'essence des êtres, puisqu'elles véhiculent une connaissance positive et réelle au sujet du défini. Cette connaissance reste cependant incomplète, d'où le fait que, pour ce qui est des espèces, on ne puisse juger précisément des frontières qui les départagent, surtout quand les individus exemplifient des ressemblances physiologiques importantes<sup>24</sup>.

Pour Locke, à l'inverse, le fait qu'on puisse parler diversement d'une même espèce, c'est-à-dire admettre un nombre variable d'idées simples dans l'idée complexe, fait office d'argument supplémentaire : si les espèces naturelles étaient concevables, elles le seraient par tous et toujours de la même manière. Donc que la réalité des espèces ne préoccupe qu'un petit nombre, soit ceux formés au « langage des écoles<sup>25</sup> », et que, même au sein de cette élite, les bornes spécifiques ne fassent pas l'unanimité, ce sont là des constats éloquents qui militent en faveur du caractère entièrement construit des classes d'espèces : « [F]or were they [les espèces] nature's workmanship, they could not be so various and different in several men, as experience tells us they are<sup>26</sup> ». Pour sa part, Leibniz estime que l'imprécision et la variabilité sont des conséquences normales de la nature empirique des contenus qui servent à la formation de nos définitions spécifiques<sup>27</sup>. Pour Leibniz, en effet, la détermination des espèces

physiques s'opère par l'identification de relations de similitude : les constitutions individuelles des êtres réalisés expriment des similarités fondamentales en termes de structures, de formes et de fonctions. Nous savons que Locke fait lui aussi des constats comparables<sup>28</sup>. La différence — elle est de taille — c'est que pour Leibniz on peut légitimement supposer que ces similarités observables s'ancrent dans la structure interne<sup>29</sup>, et donc qu'elles expriment, au niveau macroscopique, une appartenance spécifique réellement partagée. Une fois induites et combinées, les ressemblances sensibles peuvent ainsi être agencées, jusqu'à offrir une approximation distincte du tableau général des espèces.

Aussi n'est-il pas étonnant que, dans cette perspective, les bornes spécifiques ne soient jamais définitivement fixées. Bien plus, elles ne sont pas non plus appelées à l'être, puisqu'elles demeurent relatives à l'état des connaissances disponibles et à l'exhaustivité des observations empiriques. Comme nous fait défaut la connaissance de la structure intérieure des êtres, on juge de ces bornes de manière « conjecturale », en prenant appui sur les connexions qui existent entre les propriétés observables qui sont régulièrement constatées dans l'expérience. Or, dans ce domaine, ajoute Leibniz, il y a de la « latitude », si bien que « disputer alors si une différence est spécifique ou non, c'est disputer du nom<sup>30</sup> ». Cette latitude, toutefois, ne signifie pas que toutes les définitions se valent entre elles en matière d'exactitude. Au contraire, comme notre connaissance des espèces est proportionnelle à l'état des savoirs empiriques et que ces savoirs ne sont pas distribués également parmi les individus, Leibniz est conduit à hiérarchiser les sujets connaissant, c'est-à-dire à reconnaître l'autorité épistémique d'experts<sup>31</sup>. Admettre pareille diversité et pareille hiérarchie dans la connaissance est par ailleurs solidaire avec le fait que, selon Leibniz, il est possible de mettre au point une pluralité de définitions, plus ou moins distinctes selon les cas, pour exprimer une même essence<sup>32</sup>.

La position leibnizienne présente donc une nette originalité, en cela qu'elle reconnaît à la fois l'imperfection et la validité de nos distinctions spécifiques et de nos tables classificatoires. *L'imperfection*, parce que Leibniz s'entend avec Locke pour dire

qu'il nous manque la connaissance des véritables attributs essentiels ; mais aussi la *validité*, puisqu'en remplacement, « nous nous servons des attributs qui nous paraissent les plus commodes à distinguer et à comparer les choses, et en un mot à en reconnaître les espèces et les sortes : et ces attributs ont toujours leurs fondements réels<sup>33</sup> ». Ainsi, les attributs qui servent au travail de distinction du naturaliste, en dépit de leur imperfection, n'en demeurent pas moins fondés en nature, ce que contestait Locke.

Leibniz est conforté dans cette pensée par le fait que, selon lui, toute distinction ou comparaison, pour peu qu'elle soit rigoureuse et intelligible, possède un fondement réel dans la constitution interne des êtres et ainsi participe en quelque sorte au dévoilement de leur nature spécifique : « [T]out ce que nous distinguons ou comparons avec vérité, la nature le distingue ou le fait convenir aussi, quoiqu'elle ait des distinctions et des comparaisons que nous ne savons point et qui peuvent être meilleures que les nôtres<sup>34</sup> ». Certes, les distinctions naturelles surpassent les nôtres en précision et en exhaustivité, toutefois l'inverse est impossible : l'entendement humain, sous réserve qu'il raisonne convenablement, ne saurait excéder l'ordre naturel en se figurant des espèces que la nature n'aurait pas elle-même déjà distinguées ou en élaborant des distinctions trop sophistiquées. C'est ce qui explique que Leibniz en vienne à nourrir un certain optimisme à l'égard du progrès scientifique.

Le fondement de nos taxinomies étant assuré — toute relation de similarité a son ancrage dans les espèces —, il est tout à notre avantage d'intensifier la recherche scientifique, que ce soit en multipliant les expérimentations, en risquant plusieurs comparaisons ou, comme on l'a vu ci-dessus, en admettant différents critères discriminatoires. Comme l'écrit Leibniz, il faudra « beaucoup de soin et d'expérience pour assigner les genres et les espèces d'une manière assez approchante de la nature<sup>35</sup> ». D'autant plus que, dans le domaine empirique, c'est par la considération de plusieurs ressemblances, jamais d'une seule, que l'entendement procède au groupement des individus en espèces, d'où l'importance d'accroître le recours à l'expérience, car « plus on amasse de circonstances, moins la définition est provisionnelle<sup>36</sup> ». Grâce à l'observation,

on connaît toujours davantage l'essence de l'or — pour reprendre l'exemple canonique —, puisque comme les phénomènes sont des réalités, notre expérience répétée de l'or atteste à chaque fois d'un ensemble de propriétés compossibles. Découvrir une nouvelle propriété de l'or, ce serait découvrir une nouvelle possibilité et donc une nouvelle sous-espèce. Ainsi, quand Locke affirme qu'on ne peut établir de connexion nécessaire entre les propriétés observables qui existent simultanément, Leibniz répond que cette simultanéité suffit à nous faire progresser dans la connaissance des types naturels, en cela qu'elle dévoile un ordre réel de possibilités.

Dans ces conditions, il n'est pas surprenant que ce soit la comparaison systématique des ressemblances morphologiques qui passe, aux yeux de Leibniz, comme la méthode scientifique la plus prometteuse pour établir des taxinomies relativement fiables d'espèces physiques. Pareil programme de recherche empirique — basé sur des processus d'induction pratiqués sur divers individus en diverses circonstances et balisé par les lois naturelles qui règlent, dans ce monde-ci, les combinaisons compossibles — mettrait en évidence les propriétés les plus stables et donc sans doute les plus essentielles aux individus. À cela pourrait aussi s'ajouter une méthode exploratoire des types, qui, elle, fonctionnerait *a priori*<sup>37</sup>. C'est du moins ce que Leibniz laisse entendre, lorsqu'il commente les succès de certains jardiniers, passés maîtres dans l'art de créer de nouvelles variétés d'agrumes. À cette occasion, il affirme qu'il « dépend souvent de nous de faire des combinaisons des qualités pour définir encore les êtres substantiels avant l'expérience<sup>38</sup> ». Cette liberté combinatoire *a priori* a toutefois ses limites, et nécessite des connaissances empiriques préalables, puisqu'encore faut-il, ajoute immédiatement Leibniz, entendre assez les qualités « pour juger de la possibilité de la combinaison<sup>39</sup> ».

De manière spéculative, nous pouvons être certains que nos bornes spécifiques sont conformes à celles de la nature tant et aussi longtemps que les idées que nous combinons *a priori* sont compatibles, lorsqu'on suit les règles logiques d'enchaînement des définitions et notre expérience passée des lois de la nature physique<sup>40</sup>. De manière empirique, c'est l'existence actuelle de la

combinaison qui est garante *a posteriori* de la possibilité. À l'intérieur de ces limites s'offre néanmoins à la raison un terrain d'exercice légitime, afin qu'elle puisse progresser dans sa connaissance des types spécifiques :

Si nous combinons des idées compatibles, les limites que nous assignons aux espèces sont toujours exactement conformes à la nature ; et si nous prenons garde à combiner les idées qui se trouvent actuellement ensemble, nos notions sont encore conformes à l'expérience ; et si nous les considérons pour provisionnelles seulement pour des corps effectifs, sauf à l'expérience faite ou à faire d'y découvrir davantage, et si nous recourons aux experts lorsqu'il s'agit de quelque chose de précis de ce qu'on entend publiquement par le nom, nous ne nous y tromperons pas<sup>41</sup>.

Pour Leibniz, on peut dès lors raisonnablement s'attendre à ce qu'en fonction de l'avancement des savoirs et de la science, nos critères discriminatoires se précisent, approchant toujours davantage des propriétés véritablement essentielles, et à ce que nos définitions se rectifient progressivement, pour ainsi traduire l'ordre naturel de manière toujours plus satisfaisante, sans qu'il n'y ait — cela va sans dire — parfaite expression. En ce qui concerne les espèces organiques, ces méthodes combinatoires et comparatives appliquées aux individus réalisés apparaissent cependant non suffisantes. Au projet d'une anatomie comparée des types, Leibniz considère nécessaire de joindre une étude scientifique de la génération, toujours dans l'espoir de discriminer les propriétés essentielles des propriétés accidentelles<sup>42</sup>, de manière à assigner les appartenances spécifiques le plus conformément à la nature<sup>43</sup>. Des traits structuraux qui subsistent d'une génération à l'autre seront plus volontiers tenus pour spécifiques que ceux qui, présents chez les géniteurs, ne sont pas transmis à la progéniture. La génération fait cependant plus que mettre en valeur les propriétés phénoménales stables, car, pour Leibniz, les processus génératifs et la descendance peuvent parfois suffire à déterminer l'espèce, tout spécialement quand les marques externes viennent à manquer. Évidemment, les membres d'une

même espèce affichent, de manière générale, des ressemblances morphologiques. Toutefois, leur origine est également ce qui assure leur intégration sous une espèce commune<sup>44</sup>.

Bien que les lois générales de la ségrégation sexuelle spécifique, encore une fois, ne nous sont pas parfaitement transparentes et intelligibles, Leibniz nous pourvoit tout de même d'un second critère classificatoire, lequel permet de pallier aux possibles insuffisances du premier. De sorte que, même lors de circonstances où il est difficile de reconnaître les espèces en se fiant exclusivement aux traits apparents (par exemple si ces derniers s'écartent de beaucoup des archétypes attendus, comme dans les cas de monstruosité), la génération fournit malgré tout une « forte présomption<sup>45</sup> », qui est suffisante à ce que l'individu, quoique déviant en apparence, ne soit pas d'office ostracisé. Son « billet d'entrée » lui vient en outre de sa naissance, jamais uniquement de ses propriétés manifestes, ce qui lui permet d'en déroger, et ce, même de façon radicale, sans que le travail du naturaliste soit durablement entravé.

Voyons à présent comment Leibniz met à profit ce cadre théorique renouvelé afin de sauver la définition classique d'animal rationnel.

#### *4. Justification de la définition d'animal rationnel*

La souplesse supplémentaire que Leibniz s'octroie dans sa détermination des espèces grâce à la combinaison de critères discriminatoires et grâce au concept d'espèce physique apparaît clairement au moment de justifier son attachement à la définition traditionnelle de l'homme comme animal rationnel contre les deux cas de figure, constamment invoqués par Locke, du monstre et de l'idiot :

There are Creatures in the World that have shapes like ours, but are hairy, and want Language and Reason. There are Naturals amongst us that have perfectly our shape, but want Reason, and some of them Language too. There are Creatures, as 'tis said, that with Language and Reason, and a shape in other Things agreeing with ours, have hairy Tails ; others where the Males have no Beards, and others where the Females have<sup>46</sup>.

L'argument de Locke est ici de faire remarquer qu'on accorde l'humanité sur la base de critères différents : en l'absence de la raison chez l'idiot, on invoque la figure corporelle et quand c'est la figure corporelle qui fait défaut chez le monstre, on recourt au critère de la rationalité. De l'avis de Locke, ce genre de pratique dévoile le caractère arbitraire de nos découpages, ce qui finit de le convaincre que nous sommes effectivement libres de définir l'espèce humaine comme bon nous semble<sup>47</sup>. Dans le cas du monstre ou de l'idiot, on ne se règle pas sur une essence naturelle qui se ferait voir et connaître ; on accorde l'appartenance à l'espèce humaine selon nos essences nominales, et comme celles-ci sont de notre fabrication, libre à nous d'y inclure, selon ce qui est pour nous le plus commode, tantôt la forme, tantôt la raison. Les théologiens et les juristes doivent donc admettre davantage de flexibilité définitionnelle en théorie, comme ils le font déjà en pratique, en renonçant, écrit Locke, à leur « définition consacrée<sup>48</sup> » d'animal rationnel.

Leibniz partage pourtant cela avec les théologiens et juristes qu'il tient, lui aussi, à cette définition canonique, qu'il estime être « réelle et nominale en même temps<sup>49</sup> ». Réelle, parce que la rationalité est bel et bien l'attribut spécifique de l'espèce humaine. De sorte que, quand elle est présente, elle éclipse toutes les autres déterminations. Si bien que, reprenant l'exemple de Locke, Leibniz affirme que « la barbe et la queue ne seront point considérées auprès d'elle<sup>50</sup> ». Nominale, parce qu'en l'absence de la raison, on se voit forcé de juger à partir des marques extérieures et d'après la génération, comme dans le cas des autres espèces physiques. C'est d'ailleurs du fait qu'on leur reconnaît, par provision, la rationalité que les enfants sont d'emblée considérés comme des humains et ont droit au baptême, bien qu'ils n'aient pas encore démontré le moindre signe d'intelligence. Dans l'attente de cette démonstration, on recourt à d'autres indices. S'ils suffisent habituellement, ils sont bien loin d'être infaillibles, ce qui, pour Leibniz, confirme que nos définitions, « quand elles dépendent de l'extérieur des corps » — et c'est le cas de toutes nos définitions spécifiques exception faite de celle d'animal rationnel — « sont imparfaites et provisionnelles<sup>51</sup> ». Ainsi attribue-t-on une appartenance spécifique aux individus physiques

un peu comme on attribue l'humanité aux enfants, soit en nous basant sur la figure et la naissance. Locke a donc raison de constater que l'appartenance à l'humanité ne repose pas seulement le critère de la rationalité et qu'on peut parfois rencontrer des hésitations. Or, si ces constats risquent de placer un réalisme de forme plus classique dans l'embarras, ils ne suffisent pas à ébranler Leibniz, lequel, on l'a noté précédemment, s'accommode fort bien d'une imprécision définitionnelle et de l'usage de multiples critères discriminatoires.

##### 5. Conclusion : *Orthodoxie et originalité de la posture leibnizienne*

Sur la question des espèces et des genres, les thèses lockéennes et leibniziennes présentent tout compte fait d'appréciables recoupements<sup>52</sup>. Locke et Leibniz s'entendent pour dire que les corps ont une structure interne qui est liée, de manière causale, avec les propriétés observables et ils soutiennent que nos représentations abstraites des espèces expriment des ressemblances réellement observées dans l'ordre naturel. Ils admettent aussi tous deux le caractère incertain et changeant des frontières spécifiques retenues dans nos taxinomies et en imputent conjointement la faute à l'imperfection et aux limites de l'entendement humain. Bien plus : pour réduire la gravité de ces limites cognitives, Locke et Leibniz prescrivent le même remède, soit le recours plus soutenu à l'expérience. Bien entendu, il reste que, de ces constats partagés et de cette même injonction — il faut accroître nos connaissances empiriques —, Locke et Leibniz n'aboutissent pas, loin de là, aux mêmes conclusions.

Pour Locke, la méprise (scolastique) concernant l'existence objective d'espèces naturelles est avant tout le signe qu'on gagnerait à modérer nos penchants essentialistes, induits par le langage, en faisant preuve de davantage de vigilance dans notre emploi des termes, surtout des termes généraux. Aussi les recommandations qu'il formule, dans les chapitres qui clôturent le livre III, visent-elles essentiellement à limiter ce qui finalement se rapporte, pour Locke, à des abus de langage (*abuses*). Leibniz énonce également certaines recommandations, seulement celles-ci concernent moins le langage que la démarche scientifique. Car même s'il s'entend avec Locke pour

dire que notre description des espèces est condamnée à demeurer nominale<sup>53</sup>, Leibniz estime que cette description est non seulement instructive quant à l'être réel des choses, mais qu'elle est en outre perfectible. D'une part, il n'y a pas lieu de penser que nos définitions, quoique nominales, soient dépourvues d'ancrage dans l'essence des êtres, puisqu'elles s'appuient sur des ressemblances fondées en nature. D'autre part, nos définitions spécifiques peuvent gagner en distinction et entrer en correspondance toujours plus étroite avec l'ordre naturel des espèces. On peut même maximiser les progrès descriptifs en appliquant certaines méthodes, décrites plus haut, telles que la comparaison systématique des similitudes et l'étude de la génération.

Chose certaine, dans ce débat entourant l'existence des classes génériques et spécifiques, il serait réducteur d'assimiler la position de Leibniz à une simple défense de la définition canonique de l'être humain comme animal rationnel. Leibniz y souscrit bel et bien par ailleurs, et nous avons pris soin, en terminant, de montrer comment sa conception plus souple des espèces physiques lui permettait d'effectivement préserver cette définition et de parer les attaques de Locke. Seulement, on peut minimalement conclure qu'il ne s'agit pas là d'une motivation exclusive. Autrement, les critiques qu'il adresse à Locke l'auraient conduit à de plus amples considérations au sujet de la suprématie de l'espèce humaine et des conséquences dommageables, entre autres pour la morale et les mœurs, advenant la contestation de cette suprématie. Or ce qui semble préoccuper Leibniz, au troisième livre des *NE*, c'est plutôt l'intérêt scientifique de présumer l'existence d'un ordre cohérent des espèces, c'est-à-dire de postuler cet ordre, faute de le connaître, à titre d'hypothèse, de manière à ce que le naturaliste soit mieux outillé dans son investigation du monde naturel.

Une fois instrumentalisées et mises au service de la science, les notions de genre et d'espèce permettent en effet de décrire et de définir nominalement les êtres physiques, de sorte à hisser la compréhension humaine des phénomènes à davantage de distinction et d'universalité. Les notions de genre et d'espèce sont donc loin de constituer, pour Leibniz, les reliques périmées de l'ontologie

aristotélicienne. Au contraire, elles apparaissent nécessaires au déploiement de sciences de la nature empiriquement fondées dans les phénomènes, parce que ces notions sont les seules à rendre compte de l'arrangement harmonieux des individus et à permettre une définition nominale des essences.

Il n'y a pas trace de pareilles ambitions chez Locke. Pour ce dernier, les ressemblances affichées par les individus demeurent irrémédiablement mystérieuses, sans véritable portée explicative, et elles échouent à fonder nos taxinomies de manière réelle. Au mieux peuvent-elles s'assimiler à du savoir expérimental<sup>54</sup>. Pour Leibniz, en revanche, l'existence de relations de similitude authentifie la pertinence de la notion d'espèce, sur laquelle, dans les *NE*, Leibniz fonde tout entier un programme scientifique audacieux et original, qui suppose un éventail de critères spécifiques obtenus de l'expérience, une spécialisation des savoirs et le développement d'une expertise technique en taxinomie.

Que Leibniz mette ainsi de l'avant la valeur scientifique des concepts génériques et spécifiques est d'ailleurs conforme au propos qui est celui des *NE*, un propos, rappelons-le, prioritairement épistémologique. Dans le présent article, nous avons tenté de montrer que la portion des *NE* destinée aux genres et aux espèces ne faisait pas exception à l'orientation générale de l'ouvrage. Aussi peut-on émettre des doutes vis-à-vis de l'accusation selon laquelle Leibniz aurait échoué à rencontrer l'auteur de l'*Essay* sur son propre terrain, en offrant, dans le débat sur les espèces, une réponse métaphysique à une question qui, chez Locke, en était une d'épistémologie<sup>55</sup>. En fait, bien que Leibniz dispose, de par sa métaphysique et son ontologie, de plusieurs éléments doctrinaux susceptibles de venir fonder le postulat des classes d'espèces, il ne les mobilise pas expressément, ou du moins, il ne fait pas reposer sur eux l'essentiel de ses arguments, comme il nous a été donné de le voir.

Il y a bien certains principes théoriques qui, en amont, pèsent lourd dans la conception que Leibniz se fait du monde naturel, et qui peuvent en partie expliquer le plaidoyer leibnizien en faveur de la réalité des classes spécifiques. Soulignons par exemple que, pour Leibniz, le monde a été divinement institué selon un ordre

fixe et immuable. Aussi en dépit du fait qu'il y ait place, dans ce monde, pour des transformations radicales (même au plan spirituel, suivant le réarrangement des monades immortelles en différentes organisations de domination), ces transformations s'opèrent toujours en vertu de structures préétablies et donc, vraisemblablement, dans l'ordre du vivant, en vertu de structures spécifiques, qui garantissent ordre, intelligibilité et harmonie dans la nature. Certains postulats théologiques<sup>56</sup>, qui lui inspirent par exemple son anthropocentrisme, ainsi que plusieurs autres éléments de sa pensée — sa philosophie du langage<sup>57</sup>, sa logique ou sa métaphysique modale des mondes possibles<sup>58</sup> — pourraient encore être invoqués et tous ces éléments ont sans conteste leur importance dans une justification globale de la posture leibnizienne à l'égard des espèces.

Pour notre part, nous avons préféré insister sur la dimension épistémologique, espérant par là montrer que le postulat des espèces autorise le philosophe de Hanovre à formuler des énoncés programmatiques à l'endroit d'une méthode et d'une taxinomie scientifiques, veillant ainsi au déploiement optimal des sciences de la nature, et de la nature vivante en particulier. Ces conclusions sont aussi propices à faire apparaître le caractère proprement original de la théorie leibnizienne des espèces, malgré les éléments plus conservateurs qu'elle comporte. Car bien que Leibniz fasse preuve d'un authentique attachement à l'égard de la tradition philosophique qui le précède — d'où sa relative orthodoxie —, la récupération qu'il fait de cette tradition est toujours novatrice, jamais servile. Et la question des genres et des espèces illustre admirablement cette règle générale.

- 
1. Gottfried Wilhelm Leibniz, *Nouveaux essais sur l'entendement humain*, Paris, GF, 1990 [1765], 440 p.
  2. John Locke, *An Essay concerning Human Understanding*, Oxford, Oxford University Press, 1975 [1704], 748 p. On cite habituellement l'*Essay* selon le modèle suivant : E livre, chapitre, paragraphe. Comme les *Nouveaux essais* calquent très précisément la structure de l'*Essay*, le même système de référence est d'usage.

3. Nicholas Jolley, *Leibniz and Locke : a study of the New essays on human understanding*, Oxford, Oxford University Press, 1984, p. 145.
4. Susanna Goodin, «Debate over Species», dans *New Essays on the Rationalists*, Oxford, Oxford University Press, 1999, p. 170.
5. Brandon Look, «Leibniz and Locke on Natural Kinds», dans *Branching Off: The Early Moderns in Quest for the Unity of Knowledge*, Bucarest, Zeta Books, 2009, p. 404.
6. Nicholas Jolley, *op. cit.*, p. 145.
7. Nous sommes contraints de schématiser. Pour des distinctions plus fines concernant l'abstraction, voir : E 2.11.9; 3.2.4; 3.3.6-7. Sur la typologie des idées, voir : E 2.2.2; 2.6.2; 3.30.1-5. Sur la rupture lockéenne par rapport à la tradition, voir : Micheal Ayers, «Locke Versus Aristotle on Natural Kinds», *The Journal of Philosophy*, vol. 78, n°5, (1981), pp. 247-272.
8. E 3.6.2.
9. E 3.3.2-3. Cf. E 3.6.28 -30.
10. E 3.3.12 -14.
11. E 3.3.11 : «Inventions and Creatures of the Understanding».
12. E 3.6.4.
13. E 3.3.13 : «I would not here be thought to forget, much less to deny, that Nature in the Production of Things makes several of them alike: there is nothing more obvious, especially in the Races of Animals, and all Things propagated by Seed».
14. Ce qui n'est pas le cas des modes mixtes par exemple. Cf. E 3.5.2-5.
15. E 3.3.13
16. NE 3.6.14.
17. E 3.3.17 Le paradigme mathématique est usuel. Cf. E 3.6.8 et 3.6.22.
18. Même constat chez Jolley (*op. cit.*, p. 150: «Locke thinks that the existence of monsters and borderline cases is evidence that nature does not divide things up into natural kinds») et Ayers («Locke Versus Aristotle on Natural Kinds», *loc. cit.*, p. 268).
19. E 3.3.17
20. Sur la correspondance scientifique de Leibniz, qui est d'un grand intérêt pour la question des espèces, voir : Justin Smith, *Divine machines*, Princeton, Princeton University Press, 2010, pp. 251-255 et pp. 305-312. Leibniz tente même une explication fixiste pour rendre compte de la découverte récente d'un fossile de mammouth (*Ibid.*, p. 351).
21. NE 3.6.14.
22. NE 3.3.18.

23. Seulement, quand l'un parle d'essence, l'autre parle de définition. Au-delà de l'écart terminologique, essence et définition nominales partagent cela qu'elles sont toutes deux une suite de propriétés obtenues de l'expérience et qui servent à la reconnaissance. Certains éléments interdisent toutefois une identification trop rapide. Notons par exemple le fait que, pour Leibniz, la définition nominale est instructive quant à l'essence et que l'impossibilité d'atteindre à des définitions réelles, dans le cas des individus physiques, est chez lui tributaire d'une thèse relative à l'irréparable confusion des contenus sensibles ainsi qu'à l'infinité des prédicats individuants, une infinité que seul l'entendement divin peut envelopper. Cf. NE 3.3.6.
24. NE 3.6.38.
25. E 3.6.24
26. E 3.6.26
27. NE 3.6.31.
28. Cf. note 13.
29. Ce point est déterminant pour la suite de l'argument leibnizien. Cf. Jan-Erik Jones, «Leibniz and Locke and the Debate over Species», dans *Leibniz selon les Nouveaux Essais sur l'entendement humain*, Paris-Montréal, Bellarmin-Vrin, pp. 146-147.
30. NE 3.6.36. Voir aussi : 3.6.39.
31. NE 3.3.15.
32. *Ibid.* : «[I] faut considérer qu'il n'y a qu'une essence de la chose, mais qu'il y a plusieurs définitions qui expriment une même essence».
33. NE 3.6.13.
34. NE 3.6.14.
35. *Ibid.*
36. NE 3.6.34.
37. François Duchesneau parle ici d'une «théorie combinatoire» qu'il qualifie d'abord «génétique» (*Leibniz et la méthode de la science*, Paris, PUF, 1993, p. 374), puis finalement de «généralisatrice» (*Leibniz, le vivant et l'organisme*, Paris, Vrin, 2010, p. 116).
38. NE 3.6.28.
39. *Ibid.*
40. Parce qu'elles nous renseignent sur les possibilités de combinaisons, les lois naturelles nous renseignent aussi sur les espèces. Cf. Jan-Erik Jones, «Leibniz and Locke and the Debate over Species», *loc. cit.*, p. 148 : «The laws of nature determine the compatibility of properties and thus help us constrain our species definitions by providing empirical

evidence that same properties are members of homeostatic clusters of properties.»

41. NE 3.6.31. Cf. NE 3.3.15.
42. Cf. Richard Arthur, «Animal Generation and Substance in Sennert and Leibniz», dans *The Problem of Animal Generation in Early Modern Philosophy*, New York, Cambridge University Press, 2006, p. 170.
43. Dans le cadre de l'exemple de botanique qui nous occupait plus haut, Leibniz ajoutait que si le processus de génération des plantes nous était mieux connu, «les variétés qu'on y remarquerait [fourniraient] un fondement à des divisions fortes naturelles», et la discussion qui suit immédiatement, au sujet de l'hypothèse du pollen pour expliquer la fécondation, montre bien que Leibniz juge les structures reproductives susceptibles de fournir des distinctions précieuses quant à l'ordre des espèces végétales (NE 3.6.14). Cf. Evelyn Vargas, «Nominal Species and Generation in the Nouveaux Essais», dans *Leibniz selon les Nouveaux Essais sur l'entendement humain*, *op. cit.*, p. 163 : «Leibnizian species are grounded in the intra-organic features that generation preserves». Cela dit, depuis les indications des *NE*, il n'est pas évident de déterminer en quoi consisterait au juste une étude systématique de la génération au-delà de l'analyse des embryons et semences, ni en quoi cette méthode est liée à la théorie combinatoire (cf. note 37). Dit plus simplement : Leibniz envisage-t-il *concrètement*, à la manière d'un Maupertuis, de dévoiler les mécanismes de la génération par le croisement et le métissage ?
44. NE 3.6.14 : «[D]ans les corps organiques [...] nous définissons l'espèce par la génération, de sorte que ce semblable, qui vient ou pourrait être venu d'une même origine ou semence, serait une même espèce. Dans l'homme, outre la génération humaine, on s'attache à la qualité d'animal raisonnable».
45. NE 3.6.23.
46. E 3.6.22
47. Cette conclusion est indéniable, notent les commentateurs. Cf. Roger S. Woolhouse, *Locke's Philosophy of Science and Knowledge*, Oxford, B. Blackwell, 1971, p. 83 : «According to Locke, there is no correct answer to the question 'What is man?' which say what man really is, for we are free to define as we will». Cf. Nicolas Jolley, *op. cit.*, p. 155 : «Locke generally claims that the question whether a particular individual is a man requires an element of arbitrary human decision».
48. E 3.4.26 : «sacred definition».
49. NE 3.6.22.

50. *Ibid.*

51. *Ibid.*

52. Susanna Goodin et Mark Kulstad y insistent également. Cf. «Leibniz, Locke et l'essence de l'or», dans *Leibniz et les puissances du langage*, Paris, Vrin, 2005, p. 217.

53. Nous savons déjà (cf. note 23) que l'équivalence n'est pas parfaite.

54. E 4.3.29

55. La thèse se trouve chez Susanna Goodin («Debate over Species», *loc. cit.*, pp. 173-174) et fait l'objet d'importantes critiques de la part de Jan-Erik Jones («Leibniz and Locke and the Debate over Species», *loc. cit.*, pp. 144-152) et de François Duchesneau (*Leibniz, le vivant et l'organisme, op. cit.*, pp. 103-119).

56. Cf. Richard Arthur, «Animal Generation and Substance in Sennert and Leibniz», *loc. cit.*, p. 163 ; Brandon Look, «Leibniz and Locke on Natural Kinds», *loc. cit.*, p. 407.

57. Cf. Susanna Goodin et Mark Kulstad, «Leibniz, Locke et l'essence de l'or», *loc. cit.*, p. 211-240.

58. L'article de Brandon Look («Leibniz and Locke on Natural Kinds», *loc. cit.*, pp. 380-407) exploite pratiquement toutes ces pistes.



# Connaissance et sagesse dans la philosophie de Louis Lavelle

NICOLAS COMTOIS, *Université Laval*

«*Nul ne peut connaître la vie de l'esprit si la douceur lui est étrangère.*<sup>1</sup>»

L'interrogation fondamentale qui nous anime ici fut d'abord celle de Pierre Hadot, et reste celle de plusieurs de nos contemporains, à savoir si la philosophie peut encore être une sagesse, c'est-à-dire si elle peut, tout en conservant les acquis de siècles de travail théorique (et sans trahir son exigence conceptuelle), constituer également une pratique<sup>2</sup>. Il nous semble que l'on peut trouver une réponse originale à cette question chez un philosophe français du siècle dernier qui reste aujourd'hui méconnu, mais dont l'œuvre est incontournable dès lors que l'on porte un intérêt au problème qui nous occupe ; il s'agit de Louis Lavelle. Nous nous demanderons donc spécifiquement comment, sans sortir du cadre universitaire (Lavelle a été professeur à la Sorbonne et au Collège de France) et tout en assumant l'état, en son temps, du questionnement philosophique, il lui a été possible de lier ensemble connaissance philosophique et «sagesse de la conduite<sup>3</sup>». Cela nous conduira à étudier sa pensée en trois temps, qui nous semblent marquer les trois étapes fondamentales de son développement : tout d'abord, la décision de partir de l'expérience de l'être, au plus intime du questionnement philosophique ; ensuite, l'élaboration d'un système entier de métaphysique, malgré le discrédit dans lequel était tombée une telle démarche ; enfin, l'approfondissement existentiel de ce savoir théorique, de manière à constituer un enseignement susceptible de jeter une lumière nouvelle sur la vie quotidienne.

## *1. L'expérience de l'être*

Il faut expliquer tout d'abord le choix de Lavelle de prendre une expérience (il s'agit ici d'une expérience intérieure, dont la portée est

métaphysique) pour point de départ de la connaissance philosophique. Explicitement, ce choix est tributaire de l'appartenance de Lavelle à la tradition spiritualiste française, qui peut être décrite comme une série de développements fondés sur l'expérience cartésienne du *cogito*<sup>4</sup>. De manière plus implicite, mais peut-être plus vive, il est lié à l'atmosphère apparue en philosophie dans la foulée du bergsonisme. L'expérience de l'être est considérée avant tout par Lavelle comme un fait immédiat de la conscience, sans l'exigence préalable d'une démarche conceptuelle ou méthodique ; l'expérience telle qu'il l'envisage s'inscrit dans le cadre d'un « retour au concret<sup>5</sup> » et Bergson en est le principal inspirateur : « c'est à M. Bergson que je dois peut-être cette certitude que la philosophie ne vaut qu'à condition qu'elle exprime un contact direct, sincère, immédiat et profond avec la vie concrète et personnelle de la conscience, indépendamment et au-delà de toutes les théories et de tous les systèmes<sup>6</sup> ». Lavelle ne centrera pas son travail, comme Bergson, sur la notion d'intuition, et il tirera avant tout de l'intériorité une ontologie plutôt qu'une métaphysique de la durée. Cependant, on pourra reconnaître dans sa manière de traiter de la vie de l'esprit une préoccupation constante, comme chez son prédécesseur, pour parler de la conscience telle qu'elle peut être appréhendée directement<sup>7</sup>.

### *1.1 Expérience primitive*

La philosophie de Lavelle part donc d'une expérience, et cette expérience, de son point de vue, est celle de l'être. D'ailleurs, si l'on veut avoir une connaissance de l'être, quelle qu'elle soit, fait remarquer Lavelle, il n'y a pas d'autre voie possible : « si la logique nous permet d'atteindre l'intelligibilité d'une chose, c'est-à-dire son idée, la chose ne peut être atteinte elle-même que d'une manière immédiate, et par suite, dans une expérience : or l'être est précisément le caractère qui fait que les choses sont des choses<sup>8</sup> ». Mais de quelle expérience s'agit-il ? Et comment, si elle ne concerne que moi, peut-elle me permettre d'atteindre l'être ?

Tout d'abord, il convient de souligner que, dans sa forme première, cette expérience n'est pas spécifique au philosophe : elle n'est pas le fruit d'une méditation, elle ne dépend pas d'une méthode, mais elle

constitue au contraire l'expérience la plus simple, la plus immédiate et, quoiqu'elle puisse être ignorée ou niée, la plus prégnante de la conscience. « Chacun avouera qu'elle est primitive, ou plutôt qu'elle est constante, qu'elle est la matière de toutes nos pensées et l'origine de toutes nos actions, que toutes les démarches de l'individu la supposent et la développent.<sup>9</sup> » Ainsi, quoiqu'elle soit en liaison intime avec l'activité du moi, elle précède même la découverte de la vie subjective ; c'est le moi qui découvre l'être, mais il le découvre avant même de prendre conscience de son inscription en lui : « on peut dire que je découvre la présence toute pure, qui est la présence de l'être au moi, avant de découvrir la présence subjective, qui est la présence du moi à l'être<sup>10</sup> ». Mais sur quoi repose au juste cette présence de l'être, et comment est-elle rendue possible ? Ce fait primitif, le seul où le moi pourra trouver légitimement son « point d'insertion », se trouve selon Lavelle dans l'expérience « de ma présence active à moi-même<sup>11</sup> ». C'est en effet de plain-pied avec l'esprit que l'être se trouve présent : il s'agit de l'être de son activité même. Ainsi, Lavelle peut relier la présence d'un être qui précède et dépasse l'activité subjective précisément à l'action du sujet la plus simple — par exemple le « pouvoir que j'ai de remuer le petit doigt » —, en tant que celle-ci me permet de commencer « à saisir le réel par le dedans<sup>12</sup> ». Je sens de façon manifeste que l'être m'est présent lorsque je déploie ma propre activité, et il m'est indubitablement présent quelle que soit l'action que je produise, car il n'y a que l'être qui puisse donner son caractère d'éternelle nouveauté à cette activité. Et à la manière de Fichte — pour qui ce fait permet, paradoxalement, d'effacer la notion d'être et de ne plus parler qu'en termes de Moi et d'Acte — Lavelle a pu parler ici d'intuition intellectuelle<sup>13</sup> : « cette expérience est donc une intuition intellectuelle, elle est la racine commune de la pensée logique, qui la suppose pour ne pas être un vain jeu dialectique, et de l'expérience sensible, qui l'approprie aux conditions particulières où nous sommes placés<sup>14</sup> ». C'est que l'action du moi, qu'elle concerne la pensée conceptuelle ou la réalité sensible, implique chaque fois implicitement la conscience de l'être que je suis (un être libre, en acte), seul accès qui me soit donné à l'être comme tel.

L'être est donc présent au moi parce qu'il est «présent tout entier avec chaque opération de la pensée<sup>15</sup>». Mais la signification — et *a fortiori* la justification théorique — de cette intimité de l'être ne peut être identifiée par celui qui n'en a qu'une intuition implicite : il lui faut encore faire l'effort de la réflexion pour entrer en philosophie.

### 1.2 *Expérience réflexive*

C'est en effet grâce à la réflexion que l'expérience primitive peut, pour la première fois, se doubler d'une connaissance. Celle-ci constitue une première analyse de l'être<sup>16</sup>, par laquelle le moi se constitue et dépasse la seule intuition de la présence de l'être pour établir de manière éclairée sa relation avec lui : «Dans une seconde démarche, la présence de l'être devient notre présence à l'être. Et sans doute cette seconde phase de l'expérience initiale était impliquée dans la précédente, mais elle n'en était pas encore distinguée.<sup>17</sup>» La réflexion est donc elle aussi une expérience (voire : elle est toujours la même expérience), mais son contenu est tel qu'elle permet d'initier cette dialectique de l'être et du moi qui est la forme caractéristique de l'investigation philosophique. Mais avant d'exposer les principaux éléments du discours que permet de déployer la connaissance réflexive, il convient d'examiner attentivement d'où procède la réflexion selon le point de vue lavellien et quelle est exactement la vérité qu'elle donne à connaître.

Le terme de «réflexion», à en croire l'exposé qu'en fait Lavelle, correspond essentiellement à l'expérience du *cogito* telle que Descartes en a ménagé l'accès et l'a décrite. Cependant si, comme nous le verrons, cela reste vrai en substance, il convient de nuancer cette affirmation et de bien considérer le point de vue qui est celui de Lavelle lorsqu'il reprend la première certitude cartésienne pour son compte. Tout d'abord, si la réflexion implique un retournement par rapport «à la direction naturelle de la vie» et si elle exige un travail «singulièrement difficile et aride<sup>18</sup>», il ne s'agit nullement pour Lavelle d'engager celui qui veut l'exercer dans l'entreprise du doute, ou même dans quelque exercice de méditation préalablement établi. De fait, Lavelle ne peut envisager la réflexion, comme le fait Descartes, à partir d'une ascèse radicale, puisqu'il n'est pas

question pour lui de distinguer l'esprit du corps. Au contraire, dans l'interprétation qu'il fait du *cogito*, il souligne l'importance de la sensation, seule capable selon lui de marquer le caractère fini du moi et de tracer la limite entre celui-ci et le domaine intellectuel considéré dans sa pleine extension : « si l'argument cartésien pose non pas la pensée en général, mais une pensée qui est la mienne et non pas celle des autres, il faut que ce soit aussi une pensée qui m'affecte, de telle sorte que le moi qui pense toutes choses, là où il est présent, ne peut que se sentir et non point se penser<sup>19</sup> ». Il ne s'agit donc plus de mettre en doute la réalité du corps pour envisager distinctement la pensée, mais, alors même que l'on prend conscience de la pensée comme telle, d'envisager sa propre singularité à travers la limitation que lui impose le corps. Il n'en reste pas moins que si, fondamentalement, la réflexion est « la conscience même se saisissant dans son exercice pur<sup>20</sup> », sa proximité avec le *cogito* est indéniable. La réflexion est, comme le *cogito*, un retour de l'esprit sur son activité même, à l'encontre de la spontanéité naturelle ; c'est cela même qui la rend difficile, même lorsque ce retour sur soi n'implique pas que le corps doive être révoqué en doute.

L'expérience réflexive lavellienne tire donc son origine du *cogito* de Descartes, tout en en offrant une interprétation originale, dont il faut remarquer qu'elle permet d'en renouveler l'expérience (plutôt que d'en faire le produit d'une méditation devant être faite « une fois en sa vie<sup>21</sup> »). Mais le problème pour Lavelle est moins de savoir comment parvenir à la réflexion que d'en établir d'une part la possibilité et d'autre part les implications sur le plan métaphysique. Le premier point a déjà été traité, il s'agit de l'expérience de la présence de l'être. Quant au second, il débute avec la thèse centrale de la pensée de Lavelle<sup>22</sup> : l'identité de l'être et de l'acte. Mais comment la réflexion peut-elle permettre de l'établir ? Il devrait déjà apparaître, tout d'abord, que si la réflexion est possible, c'est parce que la pensée, dans sa pureté, est une activité ; c'est précisément ce qui lui permet de *se réfléchir*. Or, fait remarquer Lavelle, si nous étions le point de départ de cette activité, nous n'aurions nullement besoin, pour en prendre conscience, de revenir ainsi sur nos propres pas, dans une démarche rendue difficile ; il s'ensuit que son origine est ailleurs : « si

elle [la réflexion] est un retour, c'est que l'activité même que nous exerçons, nous l'avons reçue au lieu de la produire<sup>23</sup>». L'activité de notre pensée est donc également une activité qui n'est pas nôtre, et le moment même de la saisie du moi est tout à la fois saisie de cet acte qui me dépasse de toutes parts et dont témoigne l'expérience dans son ensemble: «Elle [la réflexion] fait du moi l'origine de lui-même, mais en l'obligeant de remonter jusqu'à un acte qui est l'origine commune de tout le réel.<sup>24</sup>» Cet acte, c'est aussi l'être, ou aussi bien Dieu, comme le caractère circulaire de l'expérience que nous en faisons nous le confirme. Car nous touchons chaque fois dans la réflexion l'acte comme une origine, ou plus précisément comme une affirmation originaire: il est «éternellement le premier commencement de lui-même<sup>25</sup>». L'acte est donc ce dont tout dépend, alors qu'en lui-même il jouit d'une indépendance parfaite et pour ainsi dire nécessaire. Et il peut être saisi par la réflexion précisément parce qu'il est au fondement de tout être, et d'abord de celui du moi.

C'est ainsi que trouve sa justification une expérience qui, dans sa forme primitive, était encore confuse. Non seulement l'être nous est toujours présent, mais il peut nous être connu: nous le retrouvons chaque fois que nous opérons volontairement un retour sur l'exercice de notre propre pensée. C'est aussi de cette façon que s'ouvre pour Lavelle la possibilité toujours renouvelée d'un approfondissement spirituel.

## 2. La théorie de la participation

Mais avant de parvenir à cet approfondissement, une question va se poser: comment, à partir d'une expérience intérieure, expliquer le monde tel qu'il apparaît dans l'expérience? Comment expliquer le passage de l'être tel qu'il est appréhendé dans la réflexion à la réalité telle qu'elle est perçue ordinairement et telle qu'elle est pensée? Lavelle, dont l'orientation la plus fondamentale est platonicienne, répondra par une théorie de la participation. Si, chez Platon, celle-ci permettait d'expliquer les rapports du monde sensible à la réalité intelligible (tel objet était dit «beau» parce qu'il *participait de la beauté en soi*), chez lui elle permettra d'expliquer la relation caractéristique du moi fini à l'être absolu, précisément dans

son rapport au monde. C'est là l'objet particulier des œuvres que Lavelle a placées sous le titre *La Dialectique de l'éternel présent*, qui comprennent *De l'Être*, *De l'Acte*, *Du Temps et de l'Éternité* ainsi que *De l'âme humaine* et qui devaient — pertinemment — se terminer par un ouvrage intitulé *De la Sagesse*<sup>26</sup>.

### *2.1 Expérience du monde*

La notion de participation implique que l'homme, plutôt que d'être simple spectateur ou au contraire pur producteur du monde qu'il habite, «participe à l'opération par laquelle le monde se fait<sup>27</sup>». Le problème devient alors de savoir quelles sont les modalités de cette opération, afin de connaître le meilleur moyen pour le moi de s'inscrire dans le monde et d'y tracer une trajectoire qui lui soit propre. La solution de Lavelle passe par le concept d'intervalle, représentant cet écart entre le moi et l'être où est suscité le monde sensible et celui des idées. Elle se prolonge dans une véritable théorie de la connaissance, révélatrice de l'unité de sa pensée.

Si l'on demande à quoi correspond, concrètement, le concept d'intervalle, il faudra répondre : à l'expérience, tant intellectuelle que sensible, que le sujet a du monde, c'est-à-dire à la façon dont l'être lui est donné considérant la condition finie qui est la sienne. La théorie de la participation suppose en effet que cette expérience est susceptible d'être pensée. Nous nous attarderons ici, pour exemplifier cette notion, à la déduction<sup>28</sup> lavellienne de l'espace et du temps. En tant que l'espace et le temps constituent les formes caractéristiques de l'expérience subjective, notre auteur soutient en effet que c'est par eux que, de manière privilégiée, l'intervalle s'exprime, et que c'est par leur biais que sont liées ses diverses modalités. Car ce qui caractérise essentiellement l'intervalle, c'est l'absence, manifestation de cette insuffisance première dans laquelle se trouve le moi à l'égard de l'acte qui le suscite ; or le temps et l'espace seront précisément décrits par Lavelle comme les moyens par lesquels l'absence est véhiculée : «les deux notions d'espace et de temps apparaissent l'une et l'autre sous la forme d'un vide qui doit être rempli [...] c'est là le sens même que nous donnons à l'intervalle<sup>29</sup>». Cela signifie, comme nous le verrons dans le détail plus loin, que l'espace et le temps, s'ils

se manifestent à prime abord sous la forme du néant, représentent plus profondément un appel concret à l'engagement et à l'action. Mais de manière plus immédiate, cela veut dire que c'est à travers eux que l'être est rendu manifeste, c'est-à-dire que nous est donnée une expérience dont les divers termes puissent être liés de manière cohérente. C'est ce que Lavelle choisit d'illustrer à travers l'étude des rapports qui unissent entre eux, d'une part, l'un et le multiple et, d'autre part, le fini et l'infini. Il montre par exemple que le multiple ne surgit pour la conscience que grâce à la répétition du temps, par laquelle elle se donne un à un les objets — «le multiple est un acte qui s'engage dans le temps<sup>30</sup>» —, et que la multiplicité demande ensuite à être appréhendée dans un espace où ses manifestations puissent être juxtaposées, et qui permette par le fait même «de l'embrasser par l'unité d'un même regard<sup>31</sup>». Il souligne également le fait que les limites du sujet doivent se situer dans l'espace et dans le temps pour être déterminées : ne demandent-elles pas, si elles sont finies, à être tracées à l'intérieur d'une étendue et d'une durée sans limites ? et ne faut-il pas que l'existence soit ainsi circonscrite pour que nous puissions la surmonter par la pensée, et «occuper» intellectuellement «tous les lieux et tous les instants possibles<sup>32</sup>» ? Sans avoir à multiplier les exemples de cette sorte, on voit bien la richesse que Lavelle est capable de tirer du concept d'intervalle et de son expression à travers l'espace et le temps : c'est parce que le moi est séparé de l'être total, et parce qu'il doit trouver dans cet écart même les moyens d'y participer, que son expérience prend la forme qui est la sienne, marquée par le manque et par des rapports d'opposition qui à la fois l'isolent des autres êtres et potentiellement l'y relie.

Mais avant d'entrer plus en profondeur dans le «jeu» de la participation, il convient de se demander quelle valeur, dans ces conditions, Lavelle peut attribuer à la connaissance : si l'intervalle, où se constitue le monde où nous vivons, représente la distance entre le moi et l'être, et qu'il est caractérisé essentiellement par l'absence, dans quelle mesure est-il possible d'attribuer à la connaissance objective une validité, et dans quel rapport peut-elle nous faire entrer avec l'être ?

Pour répondre à ces interrogations, il faut d'abord évaluer le statut, chez Lavelle, de la représentation ; il faut ensuite exposer les notions d'analyse et de relation, qui caractérisent la connaissance selon lui. Pour comprendre tout d'abord ce qu'il en est de la représentation chez Lavelle, il faut situer celle-ci par rapports aux principes métaphysiques qui sont les siens. On découvre alors qu'elle ne peut se trouver ailleurs qu'au confluent de l'activité subjective, qui relève intimement de l'être, et de la manifestation objective de ce qui dépasse les limites de cette activité, mais qui doit témoigner de l'être tout autant (en tant qu'il s'agit précisément de l'être qu'il lui manque). Réfléchissant ensuite, avec Lavelle, à la justesse de l'accord entre l'activité du moi et le monde qui lui apparaît, il devient manifeste que la représentation ne peut dépendre que d'une vue déterminée sur l'être, et qu'elle constitue à ce titre ni plus ni moins qu'une réponse de l'être à la vie subjective : « la représentation se place d'emblée au cœur de l'objet, mais [...] elle éclaire seulement en lui les aspects qui sont en accord avec notre constitution propre, avec notre nature et avec nos besoins<sup>33</sup> ». Ainsi pour Lavelle, l'être, c'est *l'acte*, qui peut être appréhendé réflexivement par la conscience, mais tout aussi bien il s'agit du *tout* des phénomènes qui lui sont donnés et sont susceptibles de lui être donnés. Mais c'est ici qu'apparaît un paradoxe : le tout, s'il s'agit de l'être en soi, est nécessairement un, et néanmoins nous ne le connaissons qu'à travers une diversité de représentations. Pour surmonter cette difficulté, Lavelle soutiendra que la connaissance des phénomènes, tant intellectuelle que sensible, constitue fondamentalement une *analyse*<sup>34</sup> de l'être : l'être fini dévoile la richesse du tout en en tirant une à une des qualités. La connaissance prendra alors pour lui un caractère relatif. Si elle dépend des règles générales (logiques, par exemple) qui fondent le rapport de l'être fini à l'être total, elle dépend également de la relation particulière qu'entretient un certain sujet, dans ses limites et à partir de ses circonstances propres, avec le tout : il s'agit de *son* analyse. La pensée témoigne d'ailleurs de cette relation du fait de son unité (pensons à la relation implicite qui unit des concepts opposés, ou à celle qui relie de manière cohérente les éléments d'un système) : « La relation exprime précisément cette distinction entre les termes

qui est la condition et le moyen de leur union par la pensée, laquelle à son tour est pour ainsi dire l'attestation de leur union dans le tout.<sup>35</sup> » Le monde, c'est l'être tel qu'il doit apparaître pour nous, et c'est l'être tel que sommes arrivés à le connaître. Sa diversité témoigne du fait qu'il s'agit aussi d'un tout : un ensemble constitué d'une infinité de parties et où « sont situés tous les chemins<sup>36</sup> ».

On voit ainsi comment les conceptions de Lavelle, à partir de leur fondement dans l'expérience de l'être, permettent de donner une signification à l'expérience dans son ensemble : celle-ci ne paraît plus être le fait du hasard, mais apparaît au contraire comme la mise en forme du rapport qu'a avec l'être total tel être singulier. Cela serait encore insuffisant, cependant, s'il ne s'agissait pas là des conditions d'exercice qui conviennent à la liberté.

## *2.2 Inscription dans le monde*

Toute l'entreprise spirituelle proposée par Lavelle n'a pas d'autre but que de permettre de toujours mieux assumer le fait découvert par la réflexion, qui n'est autre que celui de la liberté. En effet, découvrir l'être comme acte, c'est le découvrir comme étant à l'origine de toute liberté, et la liberté comme étant l'intériorité de tout être, quelles que soient par ailleurs ses limites : « L'être absolu n'agit que par la communication de ce qu'il est, c'est-à-dire qu'il ne crée que des êtres et non point des choses ; mais le propre d'un être, c'est aussi de se déterminer, c'est-à-dire d'être jusqu'à un certain point capable de se suffire : car dans la mesure même où il est lui-même un être, il est une liberté.<sup>37</sup> » Il faut donc non seulement connaître le monde, mais en user comme du domaine où s'inscrit la liberté et où elle prend forme. Ici encore, l'espace et le temps joueront un rôle essentiel, dans un travail pour convertir la spontanéité naturelle en une spontanéité nouvelle, gouvernée par l'esprit.

Comme nous l'avons vu précédemment (section 2.1), l'espace et le temps sont l'expression même de l'intervalle et à ce titre, ils permettent de lier entre elles les diverses modalités de l'expérience. Mais considérant ensuite quel est le rôle de l'intervalle lui-même, en tant qu'il constitue la condition d'exercice de la liberté, il doit nous apparaître que l'espace et le temps jouent également un rôle plus

intime : ils sont au cœur de la vie même de la conscience, capables d'une part de la situer, et ensuite de lui procurer ce caractère dynamique qui est caractéristique de la liberté vécue. Car l'action d'un être fini, tout d'abord, ne saurait s'exercer sans une situation concrète : cette situation, ce sera celle qu'il occupe dans l'espace, c'est-à-dire celle de son corps<sup>38</sup>. On voit bien en effet que le corps donne son attache à la liberté ; la vie de la conscience, qui demande à se réaliser, ne peut guère avoir de signification sans lui : elle ne saurait alors trouver d'où elle part (c'est le corps qui la limite) ; elle ne saurait trouver où son activité s'inscrit (c'est le corps qui permet d'en avoir la perception, et qui permet cet engagement même). Par ailleurs, l'activité libre est pour la conscience un mouvement, et c'est progressivement qu'elle trouve son inscription dans l'être : dans le temps. En toute rigueur, aucune réalité ne correspond à ce déroulement. L'identité de l'être avec l'acte oblige en effet à penser celui-ci en termes d'éternité, et seul le présent offre alors un caractère positif : « nous n'en sommes jamais sorti, nous n'en sortirons jamais<sup>39</sup> ». Mais l'explication lavellienne dépassera le seul fait objectif du temps pour en faire ressortir la portée existentielle, c'est-à-dire le rôle pour la conscience. Lavelle dira ainsi que le futur, c'est le domaine du possible : c'est en lui que la pensée projette pour elle-même un avenir ; à ce titre, il s'agit d'une « présence pensée<sup>40</sup> ». Quant au passé, il dira qu'il est le lieu de la connaissance : c'est en lui que sont puisés les éléments de l'analyse. Il dira aussi qu'il éclaire de ce fait, de manière privilégiée, tant le présent que l'avenir : « C'est donc le passé qui donne sa signification au temps lui-même et nous ne comprendrons la signification de chaque chose que lorsqu'elle aura elle-même pénétré dans le passé<sup>41</sup> ». Espace et temps acquièrent une signification différente dès lors qu'on les considère du point de vue de l'intériorité : le corps apparaît comme un moyen dont la liberté dispose, puis le passé, le présent et l'avenir, comme les rouages nécessaires de sa réalisation. C'est qu'il faut à la liberté de l'être particulier une situation concrète où s'inscrire, par un geste qui soit susceptible d'être pensé et assimilé dans le temps.

Mais qu'est-ce en définitive que ce mouvement par lequel le moi s'inscrit dans un monde marqué par le manque et, en approfondissant

cette expérience même, s'approche continûment de la plénitude qui cherche à s'exprimer dans le présent? Il s'agit pour Lavelle d'une conversion, celle d'une spontanéité naturelle en une spontanéité spirituelle. C'est, tout d'abord, un fait élémentaire de la vie consciente que son enracinement dans la nature. Le moi est sollicité à chaque instant par cette réalité qui lui échappe encore et dont les effets sont sentis dans son propre corps. Mais y aurait-il une liberté possible s'il ne vivait pas d'abord cette forme d'asservissement? «Il faut [...] être d'abord extérieur à soi pour devenir par degrés intérieur à soi. De telle sorte que la nature est une condition de la liberté parce que celle-ci doit toujours être conçue comme une libération.<sup>42</sup>» Par la réflexion, qui est le fondement même du travail de la philosophie telle que l'entend Lavelle, le sujet pourra retrouver le principe de son action, et il sera d'autant plus libéré qu'il le prendra mieux en charge — c'est-à-dire, puisque c'est de l'être qu'il s'agit, qu'il lui accordera son consentement. Une nouvelle forme de spontanéité est donc susceptible de se faire jour, qui doit être dite spirituelle puisqu'elle procède d'une initiative purement intérieure. Mais à quelle occasion, concrètement, est-elle susceptible de s'exprimer? Ce sera au moment d'effectuer des choix. Le sujet, en effet, puisqu'il n'est pas créateur du monde (pas même de *son* monde), ne décide pas des conditions dans lesquelles s'exerce sa liberté; c'est par le biais du libre-arbitre qu'il en fait usage. C'est donc au moment de choisir qu'il devra retrouver le chemin de son intériorité pour choisir le meilleur. Cette forme d'engagement, qui semble momentanée, dépend en fait d'un équilibre qui est long et difficile à acquérir, et dont la carence peut avoir de lourdes conséquences: «Il y a entre ma nature et le choix que je fais une correspondance qu'il m'appartient de régler, faute de quoi je manque ma vocation<sup>43</sup>»; pour retrouver un chemin, il faut déjà en avoir balisé le tracé.

La philosophie de Lavelle est donc avant tout une philosophie de l'intériorité, ou plutôt de l'intériorisation: «l'être nous étant présent tout entier, il s'agit moins de donner plus d'étendue à la représentation que nous en avons que de donner toujours plus de vie et de profondeur à l'acte intérieur par lequel nous fondons en lui notre être participé<sup>44</sup>». Ceci étant dit, elle ne fait jamais l'impasse sur le

monde. Au contraire, elle cherche tout d'abord à fonder la validité de l'expérience commune, et avec elle celle de la connaissance objective, pour ensuite en proposer le dépassement à travers un effort qui se résout par des choix eu égard au cours du monde lui-même. Et on peut dire que si la trajectoire de Lavelle commence avec la réflexion, la légitimation du monde objectif (seul domaine où le moi puisse exercer ses puissances) est chez lui la première leçon de la sagesse : elle permettra de considérer toutes les nuances de la vie morale, et les mille manières dont le moi est aux prises avec sa nature.

### *3. L'Erreur de Narcisse*

La part théorique de la philosophie lavellienne ne se résume pas, bien entendu, aux éléments que nous venons de mettre en lumière. Nous nous sommes cependant efforcé d'en développer les aspects les plus fondamentaux afin de faire voir comment, à partir d'une certaine expérience, son auteur a voulu construire un discours philosophique susceptible de transformer le regard de son lecteur sur la vie quotidienne. Il convient maintenant de s'attarder à la part proprement pratique de cette pensée, ce qu'il nous sera possible de faire à partir d'une œuvre morale que l'on peut considérer comme exemplaire : *L'Erreur de Narcisse*. C'est en effet dans cet ouvrage, selon nous, que prend sa pleine signification le choix initial duquel procède toute la philosophie lavellienne, et qui peut être caractérisé comme un engagement en faveur de la réalisation spirituelle des hommes : « on cherchera où est aujourd'hui le danger le plus grave pour *la salut de la personne*, si c'est de la livrer à la séparation et à tous les délices de l'amour de soi et du jugement propre, ou de chercher à l'assujettir dans une réalité infinie dont elle ne se sépare point sans retomber au néant, qui l'appelle à la vie, à condition qu'elle écoute sa voix et qu'elle y réponde avec docilité<sup>45</sup> ». Cela demandait, dans une philosophie qui repose sur l'expérience la plus haute de l'*ego*, de surmonter le piège que tout autant il constitue, afin de le voir s'épanouir dans ses rapports avec autrui et avec Dieu.

Comme l'exprime avec justesse Jean-Louis Vieillard-Baron dans la préface qu'il a écrite pour la réédition de l'ouvrage aux éditions La Table ronde en 2003, dans *L'Erreur de Narcisse*, « Lavelle désigne

l'amour-propre — ou la vanité de vouloir donner une fausse image de soi — comme ce qui empêche la conscience de vivre<sup>46</sup>». C'est bien le sens de la référence au mythe transmis par Ovide (*Les Métamorphoses*; version à laquelle se réfère le plus volontiers Lavelle): selon la perspective ouverte par la *Dialectique de l'éternel présent*, le fait même de fixer une image de soi est une atteinte à la vie de l'esprit, dont l'essence est la liberté et donc une activité qui peut être envisagée réflexivement, mais dont le mouvement ne doit se porter en définitive qu'au-dehors. Comme le souligne Lavelle: «Il [Narcisse] commet ce péché contre l'esprit de vouloir se saisir comme il saisit les corps<sup>47</sup>». C'est la mort de l'esprit que de vouloir s'en faire une image arrêtée, car c'est en toute antinomie vouloir en faire une chose; le désir qui nous y porte est suscité par l'amour-propre, c'est-à-dire par le refus de s'engager dans l'être et la tentation de voir le moi se refermer sur lui-même et s'auto-suffire. Cela amène d'ailleurs Lavelle à dénoncer les excès dans lesquels peut faire tomber la réflexion (dans la conscience de soi). Car le retour que le moi effectue sur lui-même, s'il lui est nécessaire pour trouver son origine, peut aussi, s'il le prolonge, devenir un geste d'idolâtrie susceptible de causer sa perte: «Il ne faut jamais mettre trop de complaisance dans la conscience de soi. [...] Dans cet intérêt exclusif qu'il montre pour lui-même, le moi pense se relever, mais finit par défaillir. Car il tient toute son existence de l'objet qu'il connaît et l'être qu'il aime.<sup>48</sup>» Il y a ici une dénonciation de la tendance de la philosophie moderne à affirmer la primauté du vouloir<sup>49</sup>: plutôt que de s'enivrer des succès remportés par la volonté, le moi doit y voir autant d'occasions de discerner sa vocation, et de régler sur elle son usage. C'est d'ailleurs la notion de vocation qui semble à Lavelle rendre le mieux le sens de la liberté lorsqu'elle trouve véritablement à se réaliser dans le monde. Dans la continuité de la spontanéité spirituelle exposée plus haut (section 2.2), celle-ci repose sur «le sentiment d'un accord entre ce que nous avons à faire et les dons que nous avons reçus<sup>50</sup>». À terme, elle doit permettre que l'action se défasse de son caractère arbitraire pour prendre le visage de la nécessité: nous aurons choisi une voie si appropriée à notre nature et à la situation dans laquelle nous sommes placé que nous ne

pourrons plus dire « si c'est ce qui nous arrive qui prend la forme de ce que nous voulons ou ce que nous voulons qui prend la forme de ce qui nous arrive<sup>51</sup> ». Encore faut-il, pour y parvenir, considérer que l'action n'est jamais celle de tel sujet à lui seul, mais qu'elle le relie toujours à autrui.

Car ce n'est pas tout d'échapper à la tentation de Narcisse et d'éviter le piège de la conscience de soi ; pour découvrir sa vocation, il faut encore trouver le moyen d'avoir un rapport authentique avec ceux auxquels le destin nous lie. Et c'est encore à la réflexion qu'il faudra revenir : s'il ne faut pas la prolonger indéfiniment, il est toujours possible de l'approfondir. C'est la voie que propose Lavelle. La réflexion doit permettre d'entrer dans un rapport d'intimité croissant avec l'être. Or considérant que cette intimité nous place à la source de laquelle découlent toutes les libertés, c'est en elle que nous trouverons le fondement véritable de notre rapport aux autres : « le point où chacun se ferme sur lui-même est aussi le point où il s'ouvre véritablement à autrui<sup>52</sup> ». C'est donc la pure liberté qui donnera son sens à notre rapport avec le prochain. Non pas que l'égalité entre les hommes se résolve dans l'identité de leur essence spirituelle — les hommes sont tous libres, mais précisément ils sont appelés, parce qu'ils sont des êtres incarnés, à des libertés différentes —, mais parce qu'aucun homme ne peut se trouver libre sans encouragement et sans réciprocité. Si la liberté doit s'articuler dans une vocation, qui elle-même est fondée sur un usage harmonieux des potentialités particulières à une personne, alors ces potentialités doivent être éprouvées dans le rapport que peuvent avoir entre elles, lorsqu'elles en trouvent la générosité, des singularités inimitables : « le plus grand bien que nous faisons aux autres hommes n'est pas de leur communiquer notre richesse, mais de leur découvrir la leur<sup>53</sup> ». Il faut parvenir à considérer l'autre à partir de l'usage qu'il fait librement des virtualités qui sont les siennes. Lavelle parlera ici de pureté, ce qu'il faut rapporter à la notion de présence pure : celle de l'être. Il lui semble que plus le sujet se rapproche de la source qu'il trouve dans l'être, plus son rapport avec le monde se purifie ; il trouve en lui-même une simplicité, une force paisible qui lui étaient inconnus auparavant. C'est de cette présence spirituelle que dépend la qualité

des rapports avec le prochain, auquel on n'aura plus aucune raison de faire violence puisque l'on connaîtra l'origine de toute puissance et de toute détermination — une origine qu'il sera impossible de confondre avec le moi. De la vie du moi, on pourra passer à celle de l'âme : « L'âme ne possède rien, mais elle peut tout accueillir. Tout est pour elle offrande et don.<sup>54</sup> »

Ces développements, en ce qui concerne le rapport que l'on doit avoir avec soi-même et le comportement que cela rend capable d'adopter envers autrui, sont fondés sur la métaphysique de Lavelle, ils n'y dérogent en rien, cependant ils ne sont possibles que dans la mesure où leur auteur a quitté le domaine du discours philosophique proprement dit pour entrer dans celui de ce que l'on pourrait appeler un enseignement spirituel. Pierre Hadot, par exemple, a fait l'apologie de « la suite de méditations qui forment ce petit ouvrage [*L'Erreur de Narcisse*] et qui sont, chacune, une invitation à pratiquer un exercice spirituel<sup>55</sup> ». C'est que la lecture même de l'ouvrage ressemble moins à ce que l'on a l'habitude de classer dans le domaine du travail philosophique qu'à l'épreuve libératrice que constitue un examen de conscience. C'est la forme que prend, chez Lavelle, l'apprentissage de la sagesse.

### Conclusion

Il semble donc qu'un ouvrage comme *L'Erreur de Narcisse*, considéré à partir de ses fondements en métaphysique et en lien avec l'expérience qui l'a suscité et qu'il permet d'approfondir (tout en lui donnant une justification pratique), constitue une réponse possible à l'interrogation qui était la nôtre au départ concernant la possibilité, en philosophie moderne, d'une liaison réussie entre la connaissance et la sagesse. Et il semble qu'en fait pour Lavelle, connaissance et sagesse soient inextricablement liées et que quoiqu'il puisse insister, selon les exigences du discours qu'il tient, tantôt sur un de ces aspects et tantôt sur l'autre, on les rencontre toujours les deux à la fois, ne serait-ce que de manière implicite. Cela tient à notre avis au point de départ qu'il donne à la philosophie, dans l'expérience réflexive de l'être. En effet, dans cette expérience, connaissance et existence sont d'emblée unies, et elles le seront ensuite à chaque moment du

développement de cette philosophie: chaque nouveau savoir aura une implication du point de vue de l'existence, et chaque événement vécu sera susceptible d'enrichir et de compléter la théorie. Cela peut rappeler la démarche du principal interlocuteur philosophique de Lavelle, Descartes, dont l'œuvre renferme des considérations morales de nature tout à fait personnelle (par exemple dans la correspondance avec Élisabeth) et dont on a pu dire qu'il « est du nombre des penseurs de qui la doctrine et la vie sont en pleine harmonie<sup>56</sup> ». D'ailleurs, on peut déplorer que Lavelle, contrairement à Descartes, ne s'attarde que très peu à décrire les moyens de parvenir à l'expérience qu'il propose et élabore aussi longuement un discours métaphysique qui en définitive n'est que l'expression extérieure d'un savoir plus intime auquel le lecteur ne peut véritablement adhérer que par son propre travail et son propre vécu. Cela peut même rendre sa pensée, qui pourtant coule de source, étrangère et difficile d'accès pour plusieurs. Mais peut-être faut-il parler de la philosophie de Lavelle comme d'une démonstration de sagesse, et à ce titre, comme étant la preuve qu'il est toujours possible d'allier l'exercice clair de la raison à la douceur du sentiment.

- 
1. Louis Lavelle, *L'Erreur de Narcisse*, Paris, La Table ronde, 2003, p. 217.
  2. À ce propos, voir Pierre Hadot, *Qu'est-ce que la philosophie antique ?*, Paris, Gallimard, 1995, p. 353-424.
  3. Lavelle, *Manuel de méthodologie dialectique*, Paris, Presses universitaires de France, 1962, p. 3.
  4. Voir Jean-Louis Vieillard-Baron, « Spiritualisme et spiritualité », *Laval théologique et philosophique*, vol. 69, n° 1, février 2013, p. 57.
  5. Louis Lavelle, *De l'Être*, Paris, Aubier, 1947, p. 289.
  6. Louis Lavelle, Radio-dialogue avec Frédéric Lefèvre, 1934, dans Christiane D'Ainval, *Une doctrine de la présence spirituelle. La philosophie de Louis Lavelle*, Louvain, Éditions Nauwelaerts et Paris, Béatrice-Nauwelaerts, 1967, p. 366.
  7. Voir Hervé Barreau, « L'apport de Bergson à Lavelle: le réalisme spirituel », *Laval théologique et philosophique*, vol. 69, n° 1, février 2013, p. 9-21.
  8. Louis Lavelle, *De l'Être*, Paris, Aubier, 1947, p. 34.
  9. Lavelle, *La présence totale*, Paris, Aubier, 1934, p. 25.

10. Louis Lavelle, *De l'Être*, Paris, Aubier, 1947, p. 12.
11. Louis Lavelle, *De l'Acte*, Paris, Aubier, 1946, p. 10.
12. *Ibid.*, p. 10.
13. À propos du rapport de la philosophie de Lavelle à celle de Fichte, voir Bruno Pinchard, « Métamorphose de la *philosophia perennis* selon Louis Lavelle », *Revue des sciences philosophiques et théologiques*, tome 88, février 2004, p. 263-265.
14. Louis Lavelle, *De l'Acte*, Paris, Aubier, 1946, p. 34.
15. Louis Lavelle, *La Présence totale*, Paris, Aubier, 1934, p. 44.
16. La connaissance, selon Lavelle, est analytique ; voir section 2.1.
17. *Ibid.*, p. 44.
18. Louis Lavelle, *De l'Acte*, Paris, Aubier, 1946, p. 28.
19. Louis Lavelle, *De l'Être*, Paris, Aubier, 1947, p. 203.
20. Louis Lavelle, *Manuel de méthodologie dialectique*, Paris, Presses universitaires de France, 1962, p. 16.
21. René Descartes, *Les Principes de la philosophie*, art. 1, AT IXb-24.
22. Sébastien Robert, *La philosophie de Louis Lavelle. Liberté et participation*, Paris, L'Harmattan, 2007, p. 25.
23. Louis Lavelle, *De l'Acte*, Paris, Aubier, 1946, p. 32
24. *Ibid.*, p. 33.
25. *Ibid.*, p. 34.
26. Lavelle n'a pas pu terminer cet ouvrage avant sa mort.
27. Louis Lavelle, *De l'Acte*, Paris, Aubier, 1946, p. 163.
28. Il faut en effet parler de « déduction » puisque la forme de l'intervalle n'a rien d'arbitraire, mais est conditionnée par le rapport du moi à l'être en tant qu'il constitue le champ d'action nécessaire à l'exercice de la liberté.
29. Louis Lavelle, *De l'Acte*, Paris, Aubier, 1946, p. 250.
30. *Ibid.*, p. 253.
31. *Idem* ; nous soulignons.
32. *Idem*.
33. *Ibid.*, p. 174.
34. Ici, Lavelle se situe délibérément aux antipodes d'Octave Hamelin, qui dans son *Essai sur les éléments principaux de la représentation* (1907) faisait de la synthèse le seul moyen de parvenir à une connaissance vraie (dans la relation thèse-antithèse-synthèse). La notion d'analyse, dont la filiation avec le concept kantien ne fait pas de doute, est mise de l'avant par Lavelle de manière à éviter tout formalisme logique : elle doit témoigner de la primauté de l'être, qui ne se réduit pas aux processus constitutifs de la pensée.

35. Lavelle, *De l'Être*, Paris, Aubier, 1947, p. 145.
36. Louis Lavelle, *De l'Acte*, Paris, Aubier, 1946, p. 42-43.
37. *Ibid.*, p. 180.
38. Sébastien Robert, *La philosophie de Louis Lavelle. Liberté et participation*, Paris, L'Harmattan, 2007, p. 36-37.
39. Louis Lavelle, *Du temps et de l'éternité*, Paris, Aubier, 1945, p. 164.
40. *Ibid.*, p. 256.
41. *Ibid.*, p. 109.
42. Louis Lavelle, *De l'Acte*, Paris, Aubier, 1946, p. 325.
43. *Ibid.*, p. 326.
44. *Ibid.*, p. 164.
45. Louis Lavelle, *La Présence totale*, Paris, Aubier, 1934, p. 14; l'auteur souligne.
46. Jean-Louis Vieillard-Baron, «Préface», dans Louis Lavelle, *L'Erreur de Narcisse*, Paris, La Table ronde, 2003, p. 8.
47. Louis Lavelle, *L'Erreur de Narcisse*, Paris, La Table ronde, 2003, p. 48.
48. *Ibid.*, p. 59.
49. Jean-Louis Vieillard Baron, «Préface», dans Louis Lavelle, *L'Erreur de Narcisse*, Paris, La Table ronde, 2003, p. 26-27.
50. Louis Lavelle, *L'Erreur de Narcisse*, Paris, La Table ronde, 2003, p. 148.
51. *Ibid.*, p. 153.
52. *Ibid.*, p. 58.
53. *Ibid.*, p. 194.
54. *Ibid.*, p. 256.
55. Pierre Hadot, *La Philosophie comme manière de vivre*, cité par Jean-Louis Vieillard-Baron, «Préface», dans Louis Lavelle, *L'Erreur de Narcisse*, Paris, La Table ronde, 2003, p. 9
56. Ernest Cassirer, *Descartes Corneille, Christine de Suède*, Paris, Vrin, 1981, p. 41.

## Politique éditoriale de la revue *Phares*

Tous les textes reçus font l'objet d'une évaluation anonyme par les membres du comité de rédaction selon les critères suivants : clarté de la langue, qualité de l'argumentation ou de la réflexion philosophique et accessibilité du propos. En outre, il est important pour les membres du comité que les textes soumis pour un dossier prennent en charge la question proposée de manière effective. La longueur maximale des textes recherchés est de 7 000 mots, **notes comprises**. Les textes soumis à la revue doivent respecter certaines consignes de mise en page spécifiées sur le site Internet de la revue, au [www.revuephares.com](http://www.revuephares.com). Un texte qui ne respecte pas suffisamment les consignes de mise en page de la revue ou qui comporte trop de fautes d'orthographe sera renvoyé à son auteur avant d'être évalué, pour que celui-ci effectue les modifications requises. Les auteurs seront informés par courriel du résultat de l'évaluation.

Les propositions doivent être acheminées par courriel à l'adresse électronique [revue.phares@fp.ulaval.ca](mailto:revue.phares@fp.ulaval.ca) avant la prochaine date de tombée. Toute question concernant la politique éditoriale de la revue *Phares* peut être acheminée au comité de rédaction à cette même adresse. Les auteurs sont incités à consulter le site Internet de la revue pour être au fait des précisions supplémentaires et des mises à jour éventuellement apportées à la politique éditoriale en cours d'année.