

## Golem numérique, vie et vie artificielle

Jean-Philippe Rennard  
GEM, 2008

---

Ce texte, à paraître dans M. Reffet (Ed.), *Le Golem, légendes, romans, films et cybernétique*, est une version expurgée des aspects techniques et mise à jour de : Rennard, J.P., Perspectives for Strong Artificial Life dans De Castro, L.N. & von Zuben F.J. (Eds), *Recent Developments in Biologically Inspired Computing*, IGP, 2004, 301-318.

---

« Le but ultime de l'étude de la vie artificielle serait de créer la "vie" au sein d'un autre support, idéalement un support *virtuel* où l'essence de la vie aurait été abstraite des détails de sa mise en œuvre sur *n'importe quel* support. On aimerait construire des modèles qui sont si semblables à la vie qu'ils cesseraient d'être des *modèles* de la vie pour en devenir des *exemples* ( [25], p. 147 ). »

Cette déclaration est à l'origine d'un vaste débat parmi les biologistes, les philosophes et les informaticiens. Elle définit ce que l'on connaît maintenant comme *l'interprétation forte de la vie artificielle*, c'est-à-dire le fait que la vie artificielle n'est pas limitée à la biologie théorique où aux artefacts biomimétiques, mais qu'elle peut s'étendre à la création de nouvelles instances de la vie, éventuellement indépendantes de tout support physique. Avec l'émergence de la vie artificielle comme champs scientifique, la question de la création de la vie par l'homme a ainsi quitté les domaines de la religion, de la croyance ou de la superstition pour rejoindre celui de la science.

Vingt ans plus tard, le débat autour de l'interprétation forte de la vie artificielle reste très actif. En dépit de réalisations remarquables —comme *Venus* ( [38] ), *Tierra* ( [41] ) ou *Cosmos* ( [54] ) — sa validité n'a pas été démontré. Le problème a été joliment introduit par un raisonnement provocateur présenté par S. Rasmussen à la seconde conférence sur la vie artificielle ( [39] ) :

1. Un ordinateur universel considéré comme une machine de Turing peut simuler tout processus physique (interprétation physique de la thèse de Church-Turing).
2. La vie est un processus physique (von Neumann).
3. Il existe des critères qui permettent de distinguer le vivant de l'inerte.

Accepter (1), (2) et (3) implique la possibilité de la vie dans un ordinateur.

4. Un organisme artificiel doit percevoir une réalité  $R_2$  qui, pour lui, est aussi réelle que notre « vraie » réalité  $R_1$  l'est pour nous.
5. Il découle de (4) que  $R_1$  et  $R_2$  ont le même statut ontologique bien que  $R_2$  soit matériellement intégrée dans  $R_1$ .
6. Si  $R_1$  et  $R_2$  ont le même statut ontologique, il est possible d'apprendre quelque chose sur les propriétés fondamentales des réalités en général, et de  $R_1$  en particulier en étudiant les détails des différents  $R_2$ . Un exemple de ce type propriété est la physique de cette réalité.

Même si les points (4) à (6) pourraient justifier la contribution potentielle des constructions artificielles, pour l'heure, le point (1) (computationalisme) d'une part et les points (2) et (3) (concept de vie) d'autre part, fondent ce que l'on peut appeler « la double impasse de l'interprétation forte de la vie artificielle. » Nous allons d'abord présenter les éléments constitutifs de cette double impasse et montrer les limites des approches purement théoriques. On montrera ensuite l'écart croissant entre la relative stagnation de la théorie et le progrès des expérimentations. On terminera enfin en examinant les conséquences de ce hiatus.

## 1 Du concept de vie

La vie est évidente; un enfant de cinq ans peut aisément distinguer l'inerte du vivant. Néanmoins, des milliers de pages ont été écrites depuis Aristote sur le concept de vie et le

problème demeure entier. Certaines définitions sont essentiellement conceptuelles et tentent de poser des bases théoriques fortes. D'autres sont avant tout empiriques et basées sur l'observation et la recherche d'invariants apparents.

## 1.1 Limites des approches conceptuelles de la vie

Selon Dawkins, le terme « vivant » a quelque chose de défini dans le monde réel ( [14], p. 39 ). Pour lui, les systèmes que nous qualifions de vivants ont émergé d'un processus cumulatif et très progressif qui favorise les répliqueurs les plus efficaces, c'est-à-dire ceux qui ont la meilleure capacité à s'embarquer dans une « machine de survie » (*surviving machine*). C'est ainsi le même processus qui est à l'œuvre depuis les premières « bulles » de la soupe primordiale jusqu'à l'émergence de répliqueurs robustes. Cela suggère une continuité qui rend extrêmement difficile la définition d'une frontière claire divisant l'inerte d'un côté et le vivant de l'autre.

Il y a maintenant un consensus croissant parmi les biologistes sur cette continuité : « Il semble plus approprié de considérer la vie comme une propriété continue des structures organisées, avec certaines plus ou moins vivantes que d'autres ( [20], p. 819 ). » « La vie n'est pas une qualité absolue qui est soudainement apparue, elle a émergé graduellement au début de l'évolution (...) il n'y a pas de frontières claires ( [11], p. 14 ). » « Vraisemblablement, la vie a émergé au cours d'un processus graduel de long terme, où des proto-organismes primitifs ont été créés dans l'envahissante soupe primordiale des mers. La vie n'est toujours pas terminée ; c'est en partie un phénomène indéfini en devenir, non un état d'être complètement déterminé( [19], p. 37 ). »

Le fait que les théories modernes dominantes tendent à considérer que la vie a progressivement émergé d'une sorte de proto-organisme fonde solidement l'hypothèse de l'existence d'une vaste frontière floue entre l'inerte et le vivant, où celui-ci s'est progressivement fondu dans celui-là. Cette continuité complique fortement la conceptualisation de la vie puisqu'il peut exister des systèmes « semi-vivants » ou « presque vivants. » Les principales conceptualisations de la vie échouent à traiter cette continuité. Examinons quelques exemples significatifs.

### 1.1.1 Une structure dissipative

C'est au célèbre physicien E. Schrödinger que l'on doit en 1944 la plus célèbre définition thermodynamique de la vie : « Quelle est la propriété caractéristique de la vie ? Quand peut-on dire qu'un morceau de matière est vivant ? Quand il « fait quelque chose », se déplace, échange de la matière avec son environnement (...) C'est en évitant le déclin rapide vers l'état d'équilibre inerte qu'un organisme apparaît si énigmatique (...) Comment l'organisme vivant évite-t-il ce déclin ? La réponse évidente est : en mangeant, buvant, respirant et ( dans le cas des plantes ) en assimilant. Le terme technique est métabolisme( [47], pp. 168-171 ). »

Pour Schrödinger, la vie est avant tout la capacité d'un système physique à se maintenir en déséquilibre thermodynamique vis-à-vis de son environnement. Ilya Prigogine a conceptualisé de tels systèmes comme « structures dissipatives », c'est-à-dire des structures au sein desquelles les interactions non linéaires, en consommant de l'énergie et à travers des transitions de phases, sont conduites vers des structures imprévisibles plus complexes ( [32] ).

Il est maintenant largement admis que les systèmes vivants sont des structures dissipatives ; c'est en utilisant de l'énergie qu'ils sont capables de créer de l'ordre, mais cette propriété est aussi connue comme insuffisante à définir la vie. De nombreuses structures dissipatives sont clairement inertes (la flamme d'une bougie, les tourbillons du fleuve ...).

La conceptualisation des structures dissipatives s'appuie fortement sur les bifurcations, les ruptures de symétries ou les transitions de phase. L'évolution vers la vie y est donc très soudaine ; il n'y a pas d'évolution douce, mais des transformations catastrophiques. Cette approche ne peut rendre compte de la continuité de la transition de l'inerte au vivant.

### 1.1.2 Une structure autopoïétique

À la recherche d'une définition de la vie adaptée aux sciences cognitives, H. Maturana et F. Varela ont proposé une approche plus sophistiquée que la précédente, basée sur l'autonomie et l'auto-référence ( [28] ). Ils ont ainsi conceptualisé la vie comme *système autopoïétique* :

« Un système autopoïétique est organisé comme un réseau de processus de production de composants qui (*a*) régénèrent continuellement par leurs transformations et leurs interactions le réseau qui les a produits, et qui (*b*) constituent le système en tant qu'unité concrète dans l'espace où il existe, en spécifiant le domaine topologique où il se réalise comme réseau ( [58], p. 45 ). »

Auto-référence et clôture opérationnelle sont vues comme les propriétés essentielles de la construction d'une identité dynamiquement préservée, c'est-à-dire qui est capable de réagir aux perturbations de l'environnement de manière à préserver son propre équilibre interne<sup>1</sup>. L'autopoïèse est ce qui distingue les systèmes vivants : « L'autopoïèse dans un espace physique est une condition nécessaire et suffisante pour qu'un système soit vivant ( [29], p. 84 ). »

La référence à l'espace physique signifie que les systèmes vivants sont nécessairement métaboliques et permet de les distinguer de systèmes autopoïétiques mais inertes (les sociétés par exemple ( [10] )).

La définition autopoïétique de la vie est très élégante. Elle conceptualise la vie à partir d'un concept unique et fondamental qui semble fiable et conforme à l'observation. Mais, une fois encore, cette définition est insuffisante. Certains phénomènes physiques, comme les cellules de Bénard, sont des systèmes autopoïétiques dans l'espace physique, mais n'en sont pas moins clairement inertes. On a pu arguer que ce problème peut être surmonté en ajoutant une condition simple : les systèmes vivants sont des systèmes autopoïétiques dans l'espace physique qui intègrent une description d'eux-mêmes ( [22] ). Cet ajout n'est pas inutile, mais ad-hoc. En outre, du fait de la clôture opérationnelle, l'autopoïèse viole la propriété de continuité, une semi-cloture n'existe pas. L'autopoïèse est ainsi une propriété binaire, oui/non ; un système ne peut pas être semi-autopoïétique.

### 1.1.3 Des systèmes adaptatifs

Afin de proposer un concept unificateur, certains auteurs (par exemple [11], [30] ) se sont focalisés sur les propriétés évolutives du vivant. Ainsi, selon E. Mayr : « On doit considérer comme vivante toute population d'entités qui a les propriétés de reproduction, d'hérédité et de variation ( [31] ). » Dans le domaine de la vie artificielle, la plus célèbre définition évolutive est le concept « d'adaptation souple » de M. Bedau ( [3], [4] ). L'adaptation souple fait référence au fait que les systèmes vivants ont une capacité infinie à découvrir de nouvelles solutions adaptatives imprévisibles à des changements inattendus de l'environnement : « (...) l'entité qui est vivante au *premier* sens du terme est le système adaptatif souple lui-même. Les autres entités vivantes ne le sont qu'au sens *secondaire* en vertu de leur relation avec le système adaptatif souple ( [3], p. 339 ). » C'est donc la population dans son ensemble qui est vivante au sens premier. Les individus ne le sont que par leurs relations au système adaptatif formé par la population.

Une telle définition est hautement contre-intuitive ; il est difficile de considérer la vie comme une propriété de groupe. En outre, Bedau n'y répond pas réellement à l'éventualité de longues périodes de stabilité évolutive (stases). Malgré cette stabilité, les individus concernés sont clairement vivants. C'est néanmoins l'une des plus attrayantes des définitions modernes. Elle est à la fois concise et apte à gérer les cas problématiques comme les mules, les virus où même les spores, voire les gamètes congelés qui sont dormants et donc temporairement déconnectés du système vivant auxquels ils appartiennent.

Selon Bedau ( [4] ), l'adaptation souple est conforme à l'hypothèse de continuité. De la même manière que la vie est affaire de degré, l'adaptation souple peut varier selon différents degrés. Cet argument est problématique, car malgré ces écarts il demeure une distinction claire entre les systèmes adaptatifs et les autres, même si cette adaptivité peut éventuellement être proche de zéro. On a ainsi de nombreuses preuves de l'unité de la vie (par exemple l'universalité du code génétique, la chiralité des molécules biologiques ...) et la plus ancienne des structures auto-reproductrice descend de ce qui ne pouvait qu'être potentiellement une structure adaptative souple.

Ces exemples — qui sont représentatifs des grandes classes des définitions conceptuelles de la vie — montrent bien la difficulté à gérer la continuité entre le vivant et l'inerte et le fossé croissant entre les théories de la vie et ce que les biologistes pensent de son origine.

1. L'autopoïèse peut ainsi être interprétée comme extension de l'homéostasie.

On a pu arguer du fait que la continuité n'est pas un problème réel, puisque, par exemple la continuité entre le vert et le bleu n'empêche personne de reconnaître ces deux couleurs. De fait, cela ne prouve rien car le vert pas plus que le bleu n'existent, les deux couleurs sont été arbitrairement définies de la même manière que le LA en musique a été défini comme un son de 440 Hz. Cela ne signifie par qu'une définition conceptuelle de la vie est impossible, même si une telle définition devra inévitablement intégrer des frontières (ou des zones frontières) plus ou moins arbitraires, mais on manque toujours d'un concept apte à couvrir l'ensemble de l'espace qui va de l'interne au vivant. La continuité reste un obstacle majeur sur le chemin qui mène à une théorie de la vie.

## 1.2 La vie comme ensemble de propriétés

Confrontés aux difficultés d'une définition conceptuelle de la vie, certains chercheurs ont proposé une approche empirique basée sur l'observation et l'identification d'invariants. Les définitions connues du grand public sont en général de ce type. Ainsi, selon le Merriam-Webster, la vie est : « L'état d'un complexe physique caractérisé par sa capacité à réaliser certaines activités fonctionnelles incluant le métabolisme, la croissance, la reproduction et des formes de réactivité ou d'adaptabilité. »

La vie est vue comme le résultat d'une organisation physique complexe apte à réaliser certains tâches. Une telle définition ne dit rien à propos de cette organisation, elle considère simplement la vie comme un regroupement de propriétés (croissance, auto-reproduction ...). De telles définitions sont communes. Dans le domaine de la vie artificielle, la plus fameuse est celle de Farmer et Belin qui ont sélectionné huit critères ([20], p. 818) :

1. La vie est une structure (pattern) intégrée dans l'espace-temps.
2. L'auto-reproduction.
3. Le stockage de l'information associée à l'auto-représentation.
4. Le métabolisme.
5. Les interactions fonctionnelles avec l'environnement.
6. L'interdépendance des composants.
7. La stabilité face aux perturbations.
8. La capacité à évoluer.

Cette définition est fréquemment utilisée du fait de son efficacité. Il est difficile d'ôter une propriété et (généralement), il n'est pas nécessaire d'en ajouter de nouvelles. Elle pose néanmoins deux problèmes principaux :

- Elle ne permet pas d'identifier toutes les formes de vie. Par exemple, elle ne qualifie pas de manière efficace les virus ou les proto-organismes qui sont à l'origine de la vie.
- Ensuite et surtout, elle ne dit rien sur ce qu'est la vie, elle ne fait que noter un ensemble empirique et intuitif de propriétés sans donner d'explications quant à leur origine.

Toutes les définitions de la vie basées sur un ensemble de propriétés partagent ce second problème. Étant essentiellement empiriques elles sont incapables d'expliquer pourquoi y intégrer l'une ou l'autre. Le problème est : comment sélectionner les propriétés significatives ? Par exemple, pourquoi inclure l'auto-reproduction ou l'évolutivité ? Ces deux propriétés peuvent être vues comme qualifiant la vie, mais elles n'ont été prises en compte que relativement au seul exemple de vie que nous connaissons, rien ne prouve qu'elles qualifient la vie elle-même. La définition autopiétique par exemple, ne retient nullement le critère d'auto-reproduction qui semble pourtant si essentiel. Ce problème n'a pas de solution directe puisque toute définition empirique est nécessairement basée sur « la vie telle que nous la connaissons » et non sur « la vie telle qu'elle pourrait être. » Ceci est connu sous l'expression de « problème du petit échantillon » (Small Sample Problem) : « Le problème du petit échantillon est celui de la détermination à partir d'un petit échantillon (dans ce cas un échantillon de taille 1) quelles caractéristiques de la chose (en l'occurrence la vie) sont essentielles à la chose ([12]). » L'un des objets principaux de la science consiste à déterminer les invariants fondamentaux, comment découvrir ces invariants avec une instance unique d'un phénomène ? Ceci ne signifie pas que définir la vie est impossible, mais que, tant que l'on ne connaît que la vie terrestre, il est impossible de la baser sur les seules caractéristiques observables, on a immanquablement

besoin d'une théorie générale pour surmonter ce problème.

Plus de 2000 ans après Aristote, le concept de vie reste ainsi à définir. Les problèmes liés à la continuité et celui du petit échantillon rendent la question excessivement difficile. Du fait du petit échantillon, il est impossible de fonder solidement la définition des invariants. Du fait du problème de la continuité, il est très difficile de trouver une définition conceptuelle apte à embrasser la zone (éventuellement vaste) située entre l'inerte et le vivant (même si cela reste théoriquement possible). Le problème est si fondamental que la question est posée du statut scientifique de la vie :

« (...) la vie n'est peut être pas une catégorie scientifiquement fondée (comme l'eau ou un tigre) dont les propriétés réelles unifient et soutiennent les similitudes observées dans toutes ces choses que l'on appelle *en vie*. À la place, on pourrait avoir choisi un fouillis de caractéristiques sans unité fondamentale ([9], p. 1) ». Ainsi, le seul moyen que nous avons pour définir la vie est de trouver des processus ou des principes qui pourraient expliquer pourquoi certaines caractéristiques spécifiques sont rassemblées dans les systèmes vivants (complexité, évolutivité ...) : « Au final, la nature de la vie sera fondée par ce qui donne la meilleure explication du riche ensemble de phénomènes naturels qui semblent caractériser les systèmes vivants ([6]). » Cela signifie que les fameuses *propriétés intuitives du vivant* demeurent d'une importance centrale. Les constructions informatiques apte à produire une grande partie de ces propriétés pourraient ainsi mettre en évidence certains des principes essentiels qui gouvernent la vie.

## 2 Du computationnalisme

« [Ceux qui connaissent les automates] considéreront [le] corps comme une machine qui, ayant été faite des mains de Dieu, est incomparablement mieux ordonnée, et a en soi des mouvements plus admirables qu'aucune de celles qui peuvent être inventées par les hommes (...) s'il y avait de telles machines qui eussent les organes et la figure d'un singe ou de quelque autre animal sans raison, nous n'aurions aucun moyen pour reconnaître qu'elles ne seraient pas en tout de même nature que ces animaux ([16], pp. 69-70). »

Publiée en 1637, cette phrase montre que le computationnalisme n'est pas une idée nouvelle. La forme de computationnalisme proposée par René Descartes est très différente de la forme moderne (notamment du fait du *dualisme* qui considère que l'esprit doit être interprété de manière différente du corps), mais l'idée fondamentale est là : le corps est une machine.

Dans les sciences contemporaines, le computationnalisme est formellement basé sur ce que l'on appelle l'interprétation physique de la thèse de Church-Turing, qui, dans sa version forte, peut être présentée ainsi : « (...) tout processus physique peut être vu comme un calcul, en conséquence tout processus physique peut être recréé au sein d'un medium computationnel ([21], p. 75). » Sur ces bases, le fonctionnalisme (qui est une version étendue du computationnalisme, non limité aux espaces calculatoires) pose le problème de la *réalisabilité multiple*. De la même manière qu'il y a différentes façons de réaliser, par exemple des moyens de transport (voiture, avion, rollers ...), il y a différentes façons de réaliser un esprit ou un être vivant. L'élément important n'est pas la réalisation physique, mais bien les propriétés du système. La conséquence ultime de cette thèse dans le domaine de la vie artificielle est que « (...) la vie artificielle est possible précisément parce que les *organismes* vivants eux-mêmes sont un type de *machines* qui peuvent s'auto-reproduire. Et, puisque les fonctions d'une machine — sa logique — peuvent en principe toujours être imitées dans d'autres constructions (qu'elles utilisent le même substrat ou non n'est pas important), il s'ensuit que la vie elle-même, la machine organique peut être construite ([19], p. 49). » En conséquence émerge « (...) l'idée selon laquelle les ordinateurs sont des instances des processus biologiques ([51], p. 362). »

L'idée qu'un système calculatoire discret puisse réaliser tout processus physique a récemment été promue au rang de « Principe » par Stephen Wolfram avec son *Principle of Computational Equivalence* ([63]) qui considère que tout processus non trivial, naturel ou artificiel, peut être vu comme un calcul et que tout ces calculs sont d'un niveau de complexité équivalent.

Malgré des bases mathématiques et philosophiques apparemment solides, le computationnalisme (et le fonctionnalisme) reste vigoureusement débattu. Certaines objections sont essentiellement formelles. Par exemple, il n'a jamais été prouvé qu'une ordinateur numérique est équivalent à un système analogique. Au contraire, on a pu affirmer que certains types d'ordinateurs analogiques pourraient avoir des capacités « super-Turing ». H. Siegelmann ([49]) a montré par exemple que des réseaux neuronaux récurrents analogiques pourraient réaliser des opérations au-delà des machines de Turing. De manière plus générale, Copeland et Sylvan ont montré que «... l'ensemble des calculs bien défini n'est pas épuisé par les calculs que peut traiter une machine de Turing ([13]) », mais ces preuves sont controversées et les machines considérées semblent être impossibles à construire physiquement.

D'autres objections sont plus philosophiques et basées sur des notions de représentations ou de sémantique, la plus célèbre étant la « chambre chinoise » de Searle (très rapidement : si vous avez un ensemble de règles appropriées, vous pouvez écrire une réponse en chinois à une question en chinois sans comprendre un mot de chinois ([48])). Même si on a pu arguer que cet argument ne s'applique pas à la vie artificielle ([2]), il n'en reste pas moins l'une des objections les plus discutées.

Le problème de l'implémentation demeure également très ouvert. Pousser l'idée d'une correspondance entre les états physiques et le calcul à ses limites tend même à vider le computationnalisme : « (...) si on pouvait montrer que pour toute description computationnelle et tout système physique, on peut trouver des « états physiques » de ce système qui peuvent être mis en correspondance avec les systèmes computationnels et qui, en outre, sont appropriés (...) alors l'explication computationnelle serait en danger : tout système pourrait alors être vu comme calculatoire ! En d'autres termes, le computationnalisme serait vide si tout système physique pouvait être vu comme implémentant chaque calcul ([46]). »

Le débat entre les tenants du computationnalisme et leurs critiques est sans fin ; chaque argument trouve systématiquement son opposé. Fondamentalement, le computationnalisme reste une question de foi puisque, peu importe la précision de la mesure et du résultat, les détracteurs peuvent toujours arguer qu'une mesure plus fine montrerait un continuum et les supporters qu'elle montrerait des processus discrets ([52]). En conséquence, et aussi longtemps qu'une conception plus approfondie du statut ontologique des réalisations computationnelles n'aura pas émergé, le computationnalisme ne pourra pas plus être prouvé que réfuté ([15]).

### 3 Des réalisations

La double impasse de l'interprétation forte de la vie artificielle n'empêche pas la multiplication des expériences qui présentent un nombre croissant des propriétés de la vie telle que nous la connaissons.

Vingt ans après sa présentation, l'expérience la plus fameuse reste le Tierra de Tom Ray ([41]), une sorte d'univers artificiel où la matière est représentée par les instructions processeurs, l'énergie par le temps processeur, l'espace par la mémoire vive, et où des créatures algorithmiques mutent et se reproduisent. Ray a résolu le problème de la définition de la vie en déclarant que, pour être vivant, il suffit qu'un système soit capable d'auto-reproduction et « d'évolution ouverte. » L'évolution ouverte se réfère à : « (...) un système dont les composants continuent d'évoluer vers de nouvelles formes de manière continue, plutôt que s'arrêtant sur un quelconque optimum ou sur un quelconque point stable ([55], p. 34). »

Les résultats de Ray ont été assez impressionnants pour donner lieu à des commentaires comme : « D'un strict point de vue logique, la barrière entre la vie et la vie artificielle semble être tombée : l'universalité de la vie a été démontrée ([1], p. 49). » C. Emmeche a examiné les créatures de Ray à la lumière des huit critères de Farmer et Belin ([19], p. 43-46) :

1. Les créatures de Ray sont des structures informationnelles plutôt que des objets matériels.
2. Elles sont capables de s'auto-répliquer.
3. Elles disposent d'une représentation d'elles-mêmes.
4. Elles ont une forme de métabolisme puisqu'elles redistribuent de l'énergie électrique de l'ordinateur.

5. Elles ont des interactions fonctionnelles avec leur environnement.
6. Leurs composants sont mutuellement interdépendants et elles peuvent mourir.
7. Elles sont stables dans leur environnement.
8. Elles peuvent évoluer.

Selon Emmeche, seules les propriétés 2 (l'auto-réplication est essentiellement formelle, elle ne consomme pas de « matière »), 4 (est-il raisonnable de considérer l'altération des états électromagnétiques comme un métabolisme?) et 7 (la stabilité évoquée est faible) ne sont pas pleinement satisfaites. La question n'est pas ici de dire que les créatures de Ray sont « presque vivantes » mais d'insister sur le fait que de telles constructions satisfont nombre de propriétés intuitives de la vie. En dépit de critiques fortes, Tierra a inspiré de nombreux travaux dont Avida ([1]) ou Cosmos ([54]). Les « créatures » de type Tierra sont probablement l'une des réalisations les plus étonnantes, mais il est très difficile de les considérer comme vivantes. Elles ne satisfont même pas la définition de Ray de la vie. Il a en effet été montré que ces constructions ne satisfont pas réellement l'hypothèse d'évolution ouverte. À partir d'indicateurs sur l'activité évolutionnaire, Bedau et al. ([7]) ont montré que le processus évolutionnaire lié à ces constructions est quantitativement et qualitativement différent de celui qui caractérise la biosphère. Dans la réalité, le processus évolutionnaire semble ne jamais stagner; la biosphère propose toujours de nouveaux environnements, de nouvelles niches. Au contraire avec Evita (un environnement de type Tierra — voir [7]) ou Bugs (voir [34]), après une phase de croissance, le succès adaptatif tend à rapidement stagner. Des résultats similaires ont été obtenus ([60]) avec Tierra et Echo ([23]).

On insiste généralement sur deux points principaux pour expliquer l'incapacité des constructions de type Tierra à générer un processus évolutionnaire ouvert :

1. L'environnement est petit est faiblement diversifié.
2. Les lois de la « biochimie » sont déterminées par le créateur, elles sont simples et fixes. À la différence du monde réel, il n'y a pas de possibilité de sélection d'un ensemble efficace de processus chimiques éventuellement apte à supporter une évolution ouverte.

Certaines constructions récentes proposent de dépasser ces limites. L'espace disponible peut être grandement élargi en utilisant Internet; c'était l'idée de Tom Ray quand il a proposé NetTierra ([42]) qui a été implémenté en 1997 entre différentes universités à travers le monde. Ray pense que comme en Amazonie où la diversité repose plus sur la multiplication des interactions entre espèces que sur la spécificité du substrat écologique, la complexité des organismes numériques et de leurs interactions dans NetTierra pourrait conduire à une dynamique évolutionnaire auto-entretenu ([43]). Le problème de la construction d'un environnement évolutionnaire endosémantique est plus complexe. Pargellis a proposé Amobea-II ([36]), un environnement conçu pour tester l'émergence « spontanée » de « créatures » auto-répliquantes. À la différence d'Amobea I ([35]), l'assembleur d'Amobea II est Turing-universel et permet, au moins partiellement, au système de définir lui-même son propre code génétique. Avec ce travail, Pargellis est proche du champ de la chimie artificielle qui essaie de construire des systèmes similaires au système chimique réel et d'analyser l'origine des unités évolutionnaires ([17]).

Pour l'heure, les chimies artificielles, en dépit de grands succès dans des champs variés, ont échouées à produire un système évolutif ouvert, mais des développements récents tendent à montrer que cela pourrait être possible. Les nouvelles approches tentent d'intégrer les créatures artificielles dans leur environnement, c'est-à-dire essaient de ne pas différencier les constituants de l'environnement et de la créature<sup>2</sup>. « [Pour construire un système évolutif ouvert] aucune distinction dans les capacités de représentation ne doit exister entre le phénotype et l'environnement abiotique. La distinction représentative importante est entre le génotype (vue comme structure symbolique relativement inerte) et les phénotypes plus l'environnement abiotique (un système dynamique) ([57], p. 12). »

On tente alors de réaliser la clôture sémantique de Pattee ([37], voir aussi [44]) : « (...) l'auto-référence ayant un potentiel d'évolution ouverte est une clôture autonome entre les dynamiques des aspects matériels (les lois physiques) et les contraintes des aspects symboliques

<sup>2</sup> On note que c'est là l'une des directions suggérées pour le computationnalisme de « prochaine génération » — (voir [46]).

(règles syntaxiques) d'une organisation physique. J'ai appelé cette relation auto-référente clôture sémantique (. . .). » Une telle construction pourrait conduire à une « évolution créative » ([56]), autorisant l'apparition de mécanismes complexes comme l'exaptation (modification de l'usage d'un caractère). Des premiers résultats ont été obtenus avec EvoCA ([57]), où l'environnement est représenté par une couche d'automates cellulaires (lois physiques) et le génotype par une seconde couche (règles syntaxiques). Chaque génotype contrôle une cellule donnée dans la première couche (en utilisant également des préconditions spécifiées par le voisinage) et — dans la version initiale — évolue grâce à un algorithme génétique. Dans un premier temps, Taylor a seulement démontré les capacités de sa construction à supporter un processus évolutif. La prochaine étape suppose la suppression des objectifs externes et l'utilisation d'une sélection endogène. Les individus devront survivre et se reproduire ([57]). Les règles syntaxiques devront trouver le bon compromis avec la dynamique de l'automate cellulaire. En définissant l'adaptation à travers cette dynamique, le système doit tenter d'approcher la clôture sémantique. Selon Taylor, en cas de succès, ces travaux pourraient mener à des systèmes capables de supporter des structures autopoïétiques autorépliquatrices. EvoCA doit encore prouver ses capacités et le système reste loin d'un système réellement endogène. Néanmoins c'est une tentative prometteuse d'unifier le phénotype et l'environnement abiotique, ce qui est un prérequis essentiel sur le chemin qui mène à des structures autopoïétiques capables d'une dynamique évolutive ouverte.

Des progrès significatifs ont également été réalisés dans la construction de systèmes auto-répliquatifs multi-cellulaires. D. Mange et al. ont proposé l'algorithme *Tom Thumb* de construction de boucles auto-répliquatrices ([27]). Les auteurs ont montré que cet algorithme est universel, il serait ainsi possible de construire des boucles auto-répliquatrices d'une complexité quelconque. Ceci est d'une importance fondamentale (voir par exemple [18] et [45]). Les résultats sont si impressionnants que les auteurs ont essayé d'en tester la conformité à la checklist de l'autopoïèse de Valera et al. ([59]) :

- Les cellules ont des frontières identifiables.
- Elles sont constituées d'un ensemble de parties.
- Le fonctionnement dépend d'interactions complexes entre les composants.
- C'est le même processus qui produit les frontières et la cellule elle-même.
- Tous les constituants sont produits par d'autres constituants.

« Notre boucle auto-répliquatrice avec des capacités de construction universelle est donc une cellule autopoïétique dans l'espace au sein duquel ses molécules existent. Selon Varela et al., une telle cellule autopoïétique a la phénoménologie d'un système vivant ([27]). » Les auteurs ne considèrent évidemment pas leur construction comme vivante, mais ils montrent qu'elle intègre certaines des propriétés spécifiques des systèmes vivants, notamment une forme de dynamique autopoïétique.

EvoCA et l'algorithme *Tom Thumb* ne sont que des exemples illustratifs, mais ils montrent clairement que certains développements mènent à des constructions intégrant un ensemble toujours croissant des propriétés intuitives de la vie. L'universalité de l'algorithme *Tom Thumb* pourrait conduire à des constructions auto-répliquatrices d'une complexité quelconque. À terme, des descendants de EvoCA pourraient amener à des structures évolutives satisfaisant les conditions de clôture opérationnelle, intimement intégrées dans un environnement dynamique (ayant donc des capacités métaboliques), capables d'auto-réplication et d'auto-entretien.

## 4 Conclusion

S. Rasmussen et al. travaillent sur les proto-organismes ([40]). Ils considèrent comme vivants ceux qui ont la capacité d'évoluer, de s'auto-répliquer, de métaboliser, de s'adapter aux changements de leur environnement et de mourir. Une fois encore, nous sommes encore loin de l'objectif, mais, du fait de leur matérialité, quiconque examinerait un tel proto-organisme le considérerait sans doute comme vivant. Même s'ils ne sont pas basés sur l'ADN, aucune propriété essentielle ne différencierait ces formes de vie artificielle de la vie telle que nous la connaissons. Qu'en sera-t-il des réalisations informatiques futures intégrant également ces propriétés? Selon Olson « Il me semble que les organismes générés par ordinateur sont problématiques pour la même raison que les montagnes générées par

ordinateur ([33]). » Cet argument est classique, mais difficilement tenable. Une simulation est contrôlée par ses concepteurs. Au contraire, la dynamique auto-organisatrice des mondes artificiels — par exemple les transitions complexes des automates cellulaires de classe IV ([62]) — émerge de lois physiques locales. De même, les propriétés des systèmes potentiellement vivants émergeront d'interactions entre les règles sémantiques et les lois physiques. Les propriétés caractéristiques de ces constructions seront la conséquence de la dynamique propre de l'environnement computationnel, non des décisions du concepteur (qui dans tous les cas est incapable de prévoir l'évolution de ses constructions). Ces émergences montrent que des propriétés spécifiques comme l'auto-réplication ou l'évolutivité peuvent être réalisées de manière multiple. Le processus est donc ontologiquement différent des simulations. On peut toujours argumenter que même si certaines constructions ne sont pas de simples simulations, il est possible de considérer que la réalité de second ordre n'a pas le même statut ontologique que la réalité de premier ordre. Sullins par exemple, considère que les points IV et V de Rasmussen présentés plus haut sont circulaires : « [Rasmussen] déclare que la réalité artificielle créée dans l'ordinateur est capable de capturer toutes les qualités essentielles de notre réalité (R1 est égal à R2) aussi longtemps que les agents vivants interagissent avec le système, mais la réalité artificielle doit déjà être ontologiquement équivalente à notre réalité pour produire des formes de vie réellement vivantes ([52]). » Cet argument est très intéressant, mais « pourquoi l'un devrait-il avoir la priorité sur l'autre ? ([39]) » et qu'en est-il de notre réalité (dont en outre on ne sait pas si elle est de premier ordre) où la vie existe sans conteste ? Dans ce contexte, les bases philosophiques permettant de considérer un proto-organisme comme vivant et un organisme virtuel (donc appartenant à une réalité de second ordre) ayant les mêmes propriétés émergentes comme une simple simulation, nous semblent faibles.

« La frontière entre le vivant et l'inerte, entre le naturel et l'artificiel est de plus en plus floue. » ([50], p. 222). Une vingtaine d'année après la naissance officielle de la vie artificielle, et malgré un énorme volume de travaux philosophiques de haut niveau, le problème de l'interprétation forte de la vie artificielle reste entier. On a pu dire que la vie artificielle est de fait l'un des symptômes de la « fin de la science » (J. Horgan dans [53]), mais, à la différence des travaux purement philosophiques, les constructions informatiques ne cessent de progresser. Ces progrès dans les réalisations liées à la vie artificielle vont nécessairement contribuer à l'évolution de la pensée philosophique, au moins de deux façons :

- Elles vont contribuer à valider les hypothèses et les expériences de pensée. J. Sullins ([53]) considère les simulations informatiques comme un outil pour ce qu'il appelle « le nouveau processus de découverte assisté par ordinateur » ; le meilleur exemple étant l'automate cellulaire auto-réplicateur de von Neumann qui, en montrant la nécessaire dualité interprétation / duplication des structures de codage, a anticipé le processus biologique découvert quelques années plus tard. Bedau insiste également fortement sur la capacité de la vie artificielle à favoriser l'émergence et la validation d'hypothèses philosophiques : « Les simulations de vie artificielle sont en effet des expériences de pensée ; mais des expériences de pensée émergentes. (...) Ce qui est particulier dans ces expériences émergentes est le fait que ce qu'elles révèlent ne peut être discerné que par la simulation (...) Synthétiser des expériences émergentes par la pensée est une nouvelle technique que les philosophes peuvent adapter de la vie artificielle ([5], p. 144). » De nouvelles simulations associées à de nouvelles hypothèses sur l'origine d'une propriété donnée de la vie (plus encore sur un ensemble donné de propriétés) vont nécessairement se développer dans les prochaines années. Les comparaisons entre les systèmes artificiels et réels vont inévitablement donner des clés pour la compréhension de la vie.
- De la même manière que l'étude des automates cellulaires et leur classification a montré que la complexité ne peut apparaître que dans un ensemble restreint de lois universelles ([26]), la multiplication de simulations de la vie de plus en plus précises — même en l'absence de nouvelles hypothèses — peut participer à la découverte de processus et de connexions sous-jacents.

Dans ce contexte, la vie artificielle est un nouvel outil de découverte, au même titre que le microscope ou le télescope l'ont été au 17<sup>ème</sup> siècle. Chacun sait, au moins depuis T. Kuhn ([24]), que l'écart croissant entre théorie et observation est très courant dans l'histoire des sciences et que c'est l'une des principales sources de rupture de paradigme. De la même manière que les outils d'observation de l'espace ont progressivement conduit de la

description (T. Brahé) à une forme de conceptualisation empirique (J. Kepler) et enfin à une théorie générale (I. Newton), ou que les progrès de l'étude des fossiles ont mené de Cuvier à Lamarck, puis Darwin, les simulations informatiques de la vie sont un outil unique pour aider à la construction d'une théorie générale de la vie. Tôt ou tard, des expériences (informatiques et physiques) seront si proches de la vie que l'interprétation philosophique de la vie en sera inévitablement fortement influencée. Cette évolution, c'est-à-dire le fait que des simulations informatiques satisferont un ensemble croissant des propriétés intuitives de la vie, contribuera inévitablement à renverser la double impasse de l'interprétation forte de la vie artificielle.

Le problème de la capacité d'un système informatique à reproduire tout processus physique ne sera plus une impasse. Dans ce contexte, le problème n'est pas de montrer la validité du computationnalisme, mais seulement de montrer qu'un système informatique est capable d'intégrer des constructions ayant les propriétés des systèmes vivants. La démonstration de la faisabilité de telles constructions reste évidemment à faire, mais les expériences que nous avons présentées — et bien d'autres encore — montrent que nous en sommes sur le chemin.

« [Le concept de vie] est un artefact historique qui varie selon les cultures et qui change quand nos croyances et nos préconceptions évoluent. » (Bedau dans [61], p. 88 ). Les constructions informatiques basées sur des principes capables d'unifier les propriétés intuitives de la vie vont faciliter l'élaboration d'une définition scientifique de la vie, mais elles vont également contribuer à l'évolution de l'interprétation culturelle et sociale de ce qu'est la vie. En réussissant, cette approche résoudra le problème de l'interprétation forte de la vie artificielle puisque la vie elle-même pourrait être redéfinie (dans une direction ou une autre) en fonction des résultats liés aux constructions informatiques.

Le grand physicien Anrei Linde a récemment déclaré : « On a toujours utilisé la science pour améliorer les conditions de vie. À partir de maintenant, nous devons faire de la science un outil pour comprendre la vie, la conscience, et par dessus tout nous-mêmes ( [8] ). » La vie artificielle et les tentatives pour démontrer l'éventuelle validité de son interprétation forte sont des outils sans précédents pour aider la science à aller dans le sens des souhaits de Linde.

## Références

- [1] Chris Adami. *Introduction to Artificial Life*. Springer-Verlag, New York, 1998.
- [2] D. Anderson and B.J. Copeland. Artificial life and the chinese room argument. *Artificial Life*, 8(4) :371–378, 2003.
- [3] Mark A. Bedau. The nature of life. In Margaret A. Boden, editor, *The Philosophy of Artificial Life*, pages 332–355. Oxford University Press, Oxford, 1996.
- [4] Mark A. Bedau. Four puzzles about life. *Artificial Life*, 4(2) :125–140, 1998.
- [5] Mark A. Bedau. Philosophical content and method of artificial life. In T.W. Bynum and J.H. Moor, editors, *The Digital Phoenix : How Computers are Changing Philosophy*, pages 135–152. Blackwell Publishers, Oxford, 1998.
- [6] Mark A. Bedau. Artificial life. In Luciano Floridi, editor, *Blackwell Guide to the Philosophy of Computing and Information*. Blackwell Publishers, Oxford, 2003.
- [7] Mark A. Bedau, Emile Snyder, Charles Titus Brown, and Norman H. Packard. A comparison of evolutionary activity in artificial systems and in the biosphere. In Phil Husbands and Inman Harvey, editors, *Proceedings of the fourth European Conference on Artificial Life, ECAL'97*, pages 125–134. MIT Press, Brighton, 1997.
- [8] R Benkirane. *La complexité, vertiges et promesses*. Le Pommier, Paris, 2002.
- [9] Margaret A. Boden, editor. *The Philosophy of Artificial Life*. Oxford Readings in Philosophy. Oxford University Press, Oxford, 1996.
- [10] Margaret A. Boden. Autopoiesis and life. *Cognitive Science Quarterly*, 1 :117–145, 2000.
- [11] A.G. Cairns-Smith. *L'énigme de la vie*. Odile Jacob, Paris, 1990.
- [12] E. Cameron. *Teleology in Aristotle and Contemporary Philosophy of Biology : An Account of the Nature of Life*. Ph.d. thesis, University of Colorado, Boulder, 2000.
- [13] B.J. Copeland and R. Sylvan. Beyond the universal turing machine. *Australasian Journal of Philosophy*, 77 :46–66, 1999.
- [14] R. Dawkins. *Le gène égoïste*. Odile Jacob, Paris, 1996.
- [15] Jean-Paul Delahaye. *L'intelligence et le calcul*. Pour la Science. Belin, Paris, 2002.
- [16] René Descartes. *Le discours de la méthode*. Libro, Paris, 1999.
- [17] Petter Dittrich, Jens Ziegler, and Wolfgang Banzhaff. Artificial chemistries : A review. *Artificial Life*, 7(3) :222–275, 2001.
- [18] Karl Eric Drexler. *Engines of Creation. The Coming Era of Nanotechnology*. Doubleday, New York, 1987.
- [19] Claus Emmeche. *The Garden in the Machine. The Emerging Science of Artificial Life*. Princeton University Press, Princeton, 1994.
- [20] J. Farmer and A. Belin. Artificial life : the coming evolution. In Christopher G. Langton, editor, *Artificial Life II*, Santa Fe Institute Studies in the Science of Complexity, pages 815–840. Addison-Wesley, Redwood City, 1992.
- [21] Stefan Helmreich. *Silicon Second Nature. Culturing Artificial Life in a Digital World*. University of California Press, Berkeley, 1998.
- [22] Jean-Claude Heudin. *La vie artificielle*. Hermès, Paris, 1994.
- [23] J.H. Holland. *Adaptation in Natural and Artificial Systems : An Introductory Analysis with Applications to Biology, Control, and Artificial Intelligence*. MIT Press/Bradford Books, Cambridge, 1992.
- [24] Thomas S. Kuhn. *The Structure of Scientific Revolutions*. University of Chicago Press, Chicago, 1996.
- [25] Christopher G. Langton. Studying artificial life with cellular automata. *Physica D*, 22 :120–149, 1986.
- [26] Christopher G. Langton. Computation at the edge of chaos : Phase transition and emergent computation. *Physica D*, 42 :12–37, 1990.
- [27] Daniel Mange, André Stauffer, Enrico Petraglio, and Enrico Tempesti. Artificial cell division. *BioSystems*, 2003.

- [28] Humberto R. Maturana and François Varela. Autopoiesis : The organization of living systems, its characterization and a model. *BioSystems*, 5 :187–196, 1974.
- [29] Humberto R. Maturana and François Varela. *Autopoiesis and Cognition : The Realization of the Living*, volume 42. D. Reidel Pub Co, Boston, 1980.
- [30] John Maynard Smith and Eors Szathmary. *The Major Transitions in Evolution*. D Reidel Pub Co, Boston, 1995.
- [31] Ernst Mayr. *Evolution and the Diversity of Life*. Harvard University Press, Cambridge, 1975.
- [32] G. Nicolis and Ilya Prigogine. *Self-Organization in Non-Equilibrium Systems : From Dissipatives Structures to Order through Fluctuations*. J. Wiley and Sons, New York, 1977.
- [33] Eric T. Olson. Prospects for computer-generated life, 1995.
- [34] Norman H. Packard. Intrinsic adaptation in a simple model for evolution. In Christopher G. Langton, editor, *Artificial Life I*, pages 141–155. Addison-Wesley, Redwood City, 1989.
- [35] Andrew N. Pargellis. The spontaneous generation of digital life. *Physica D*, 91 :89–196, 1996.
- [36] Andrew N. Pargellis. Digital life behavior in the amoeba world. *Artificial Life*, 7(1) :63–75, 2001.
- [37] Howard H. Pattee. Evolving self-reference : Matter, symbols and semantic closure. *Communication and Cognition -Artificial Intelligence*, 12 :9–27, 1995.
- [38] Steen Rasmussen. The coreworld : Emergence and evolution of cooperative structures in a computational chemistry. *Physica D*, 42 :111–134, 1990.
- [39] Steen Rasmussen. Aspects of information, life, reality, and physics. In Christopher G. Langton, editor, *Artificial Life II*, pages 767–773. Addison Wesley, Redwood City, 1992.
- [40] Steen Rasmussen, Liaohai Chen, Martin Nilsson, and Shigeaki Abe. Bridging nonliving and living matter. *Artificial Life*, 9(3) :269–316, 2003.
- [41] Thomas S. Ray. An approach to the synthesis of life. In Christopher G. Langton, editor, *Artificial Life II*, pages 371–408. Addison Wesley, Redwood City, 1992.
- [42] Thomas S. Ray. A proposal to create two biodiversity reserves : One digital and one organic. Technical report, ATR, 1995.
- [43] Thomas S. Ray. Artificial life. In Renato Dulbecco, David Baltimore, François Jacob, and Rita Levi-Montalcini, editors, *Frontiers of Life*. Academic Press, 2001.
- [44] Luis M. Rocha. The physics and evolution of symbols and codes : Reflections on the work of howard pattee. *BioSystems*, 60(1-3) :1–4, 2001.
- [45] Mihail C. Roco and William Sims Bainbridge, editors. *Converging Technologies for Improving Human Performance. Nanotechnology, Biotechnology and Cognitive Science*. NFS/DOC - sponsored report, Arlington, VA, 2002.
- [46] Matthias Scheutz, editor. *Computationalism. New Directions*. MIT Press, Cambridge, 2002.
- [47] Erwin Schrödinger. *Qu'est-ce que la vie ?* Christian Bourgois, Paris, 1986.
- [48] John R. Searle. Is the brain's mind a computer program? *Scientific American*, 262(1) :20–25, 1990.
- [49] Hava T. Siegelmann. Computation beyond the turing limit. *Science*, 268 :545–548, 1995.
- [50] Moshe Sipper. *Machine Nature. The Coming Age of Bio-Inspired Computing*. McGraw-Hill, 2002.
- [51] Elliot Sober. Learning from functionalism : Prospects for strong artificial life. In Margaret A. Boden, editor, *The Philosophy of Artificial Life*, pages 361–378. Oxford University Press, Oxford, 1996.
- [52] John P. Sullins. Gödel's incompleteness theorems and artificial life. *Techné : Journal of the Society for Philosophy and Technology*, 2(3-4) :141–157, 1997.
- [53] John P. Sullins. Knowing life : Possible solutions to the practical epistemological limits in the study of artificial life. *Journal of Experimental and Theoretical Artificial Intelligence*, 13(4) :397–408, 2001.

- [54] Timothy J. Taylor. The cosmos artificial life system. Informatics Research Report 263, Dpt. of Artificial Intelligence, University of Edinburg, 1999.
- [55] Timothy J. Taylor. *From Artificial Evolution to Artificial Life*. Ph.d. thesis, University of Edinburg, 1999.
- [56] Timothy J. Taylor. An alternative approach to the synthesis of life, 2002.
- [57] Timothy J. Taylor. The control of dynamical systems by evolved constraints : A new perspective on modelling life. Informatics Research Report EDI-INF-RR-0148, Division of Informatics, University of Edinburgh, September 2002 2002.
- [58] F. Varela. *Autonomie et connaissance. Essai sur le vivant*. Seuil, Paris, 1989.
- [59] Franois Varela, Humberto R. Maturana, and R. Uribe. Autopoiesis : the organization of living systems, its characterization and a model. *BioSystems*, 5 :187–196, 1974.
- [60] Mark Ward. *Virtual Organisms. The Startling World of Artificial Life*. Thomas Dunne Books. St. Martin Press, New York, 2000.
- [61] Michael Wheeler, Seth Bullock, Ezequiel Di Paolo, Jason Noble, Mark A. Bedau, Phil Husbands, Simon Kirby, and Anil Seth. The view from elsewhere : Perspectives on alife modeling. *Artificial Life*, 8(1) :87–100, 2002.
- [62] Stephen Wolfram. Universality and complexity in cellular automata. *Physica D*, 10 :1–35, 1984.
- [63] Stephen Wolfram. *A New Kind of Science*. Wolfram Media Inc., Champaign, 2002.