

Centre de Cerisy la Salle
 23-31 Août 2002
 Pontigny, Cerisy : un siècle de rencontres

**LES COLLOQUES SCIENTIFIQUES :
 VINGT ANS APRES**

Jean Petitot
 EHESS, CREA-École Polytechnique

J'aimerais saisir cette occasion du cinquantenaire pour exprimer avec un peu de solennité, en mon nom personnel et au nom de la communauté scientifique et philosophique que je puis prétendre représenter, toute ma reconnaissance envers le Centre de Cerisy et son équipe. Il s'agit plus que d'un remerciement pour un dévouement sans borne, pour l'organisation de rencontres inoubliables, pour la défense et illustration du dialogue, de la tolérance et de l'interdiscipline, pour l'aide à des idées nouvelles. Il s'agit aussi de reconnaître le rôle irremplaçable que Cerisy a tenu sur le plan scientifique avec l'organisation de certains de ses colloques dus à l'impulsion éclairée d'Edith Heurgon. De nombreuses conférences ont déjà parlé du rôle de Pontigny-Cerisy dans le domaine de la culture. L'action dans les domaines scientifiques et techniques a peut-être été moins spectaculaire, mais tout aussi profonde.

Je me souviens évidemment en particulier des colloques que j'ai moi-même organisés, ceux de 1982 *Logos et théorie des catastrophes* autour de l'œuvre de René Thom, de 1988 *Rationalité et objectivités* sur la philosophie transcendantale et le problème de l'objectivité, de 1990 *Actualité de la Critique de la Faculté de Juger* sur l'opérativité contemporaine de la philosophie kantienne, et de 1996 *Au Nom du Sens* autour de l'ensemble de l'œuvre sémiotique et romanesque d'Umberto Eco. Je me souviens aussi avec émotion de colloques que j'ai suscités comme *L'actualité de Leibniz* organisé en 1995 par Frédéric Nef et Dominique Berlioz en hommage à André Robinet ou qui ont été organisés par des proches comme celui sur *Le Labyrinthe du Continu* de 1990 dû à Hourya Sinaceur et Jean-Michel Salanskis. J'ai également participé à de nombreux autres colloques scientifiques en particulier en 1983 à *Temps et Devenir* organisé par Jean-Pierre Brans, Isabelle Stengers et Philippe Vincke à partir de l'œuvre d'Ilya Prigogine et en 1999 au colloque Hayek organisé par Robert Nadeau et Alain Leroux. Tous ces colloques présentent des affinités étroites avec d'autres grands colloques scientifiques : le colloque de 1982 sur *L'auto-organisation : du physique au politique* organisé par Jean-Pierre Dupuy et Paul Dumouchel ainsi que ceux sur *Les sciences cognitives* organisés par Daniel Andler en 1987 et 1990.

Quand on y regarde de près, on constate une convergence remarquable entre tous ces colloques. Au-delà de leurs spécificités sociologiques superficielles et des éventuels petits différends entre les ego de leurs protagonistes, ils présentent de nombreux points communs. En fait ils concernent tous un transfert massif de méthodes et de modèles scientifiques “hard” extrêmement techniques (venant des mathématiques, de l'informatique, de la physique ou de la biologie) vers des domaines considérés jusque-là comme impossibles à traiter, pour *des raisons de droit*, en termes de sciences naturelles.

Je retiendrai quatre domaines de ce type.

1. Les théories des formes et des patterns émergeant dans les substrats matériels physiques et biologiques ;
2. La théorie des structures cognitives dans un sens plus abstrait que celui de forme : structures perceptives, Gestalten, structures linguistiques et sémiotiques (constituance, syntaxe) et leur implémentation neuronale ;
3. La philosophie de l'esprit et l'intentionnalité des états mentaux dans une optique relevant aussi bien de la phénoménologie que de la philosophie analytique ;
4. Les dynamiques sociales auto-organisatrices, les structures émergeant dans des populations d'agents en interaction.

Ce qui est en jeu dans toutes ces théories est le projet d'une *naturalisation* opérationnelle des concepts philosophiques traditionnels de forme, de structure, d'organisation et de sens, et cela dans le cadre de sciences naturelles *élargies*, et même en partie *refondées*, devenues à même de se réapproprier les dimensions de l'être qu'elles avaient dû exclure pour pouvoir se constituer. Il s'agit donc de sciences naturelles susceptibles d'effectuer une question en retour — une Rückfrage au sens de Husserl — sur la coupure épistémologique fondatrice des sciences modernes, mais non pas comme chez Husserl et tant d'autres au nom d'une alternative, d'ailleurs introuvable, à la science, mais bien à partir de progrès remarquables de ces sciences elles-mêmes.

Une telle naturalisation possède nécessairement (au moins) quatre aspects.

1. le dégagement des nouveaux outils fondamentaux justifiant l'opération de naturalisation ;
2. la formulation du nouveau paradigme leur servant de cadre ;
3. les travaux d'application de ces nouveaux outils dans le cadre de ce nouveau paradigme ;
4. la réflexion philosophique rendue nécessaire par le fait que l'impossibilité d'une naturalisation avait été argumentée à fond par des philosophes aussi immenses que Leibniz, Kant ou Husserl et relayée par la plus grande majorité des penseurs (à part quelques exceptions comme celle de Paul Valéry) jusqu'à devenir l'une des principales idées reçues de la modernité.

Ces quatre aspects sont pondérés de façon variable suivant les protagonistes, mais ils sont toujours plus ou moins présents dans leur œuvre.

Historiquement parlant, le premier front de recherches s'est développé vers la fin des années 60 et les années 70. Déjà en 1975 l'*Enciclopedia Einaudi* de Ruggiero Romano paria sur l'importance et l'originalité de ces développements alors tout récents et j'ai eu le privilège de m'en occuper avec Fernando Gil, Giulio Giorello et Krisztof Pomian. Puis il y eut les grands colloques de Cerisy.

Que peut-on en dire 20 ans après ? D'abord qu'il s'agissait d'une authentique révolution scientifique. Pas plus qu'en littérature ou en philosophie, Cerisy ne s'est trompé en sciences. Ce changement qui a profondément et durablement transformé notre conception de ce que sont l'explication et la compréhension scientifiques a diffusé depuis massivement dans les milieux concernés, et y est même souvent devenu dominant. Institutionnellement prioritaire dans les grands programmes nationaux et internationaux, c'est actuellement une partie essentielle de la science "normale" pour les nouvelles générations.

Le changement de paradigme a été très bien formulé par Michaël Berry (l'un des plus éminents spécialistes du chaos quantique), dans cette salle même, la bibliothèque de Cerisy, en 1982, lors du colloque sur René Thom. Berry parlait du "dépassement interne des paradigmes de la physique classique", *interne* au sens de dépassement de la physique *par elle-même*.

Il critiquait trois "dogmes" de l'image vulgarisée de la physique classique.

(i) Le premier "dogme" est celui du déterminisme newtonien-laplacien. Il est remis en cause par le développement de la dynamique qualitative depuis Poincaré et, en particulier, par l'étude de systèmes dynamiques qui sont mathématiquement (i.e. idéalement) déterministes tout en étant physiquement (i.e. concrètement) chaotiques et imprédictibles dans tous les sens pratiques du terme. Les oppositions classiques entre le hasard et la nécessité ou entre l'ordre et le chaos sont devenues désormais obsolètes.

(ii) Le second "dogme" est celui de la différentiabilité selon lequel, aux sauts catastrophiques près, les évolutions soumises à des lois doivent être différentiables ("lisses"). Or il existe des structures "fractales" en physique classique. Par exemple celle du mouvement brownien ou celle des fluides au voisinage d'un point critique (dans ce dernier cas une hiérarchie de fluctuations possédant la propriété d'auto-similarité rend impossible, par divergence des longueurs de corrélation, la distinction entre les niveaux respectivement microscopique et macroscopique).

(iii) Le troisième "dogme" est celui du réductionnisme. Berry le dénonçait en prenant l'exemple de la théorie des caustiques. L'optique géométrique est une approximation asymptotique de l'optique ondulatoire pour des longueurs d'ondes $\lambda \rightarrow 0$. Dans un processus de focalisation de la lumière, les caustiques sont *géométriquement* les enveloppes des rayons lumineux. Leurs singularités structurellement stables sont

données par la théorie des singularités mais, au niveau *ondulatoire*, à chaque singularité est associé un pattern caractéristique appelé “catastrophe de diffraction”. Ces catastrophes de diffraction sont optiquement dominantes. Il faut les penser morphologiquement à partir de leurs singularités organisatrices plutôt que comme solutions des équations de Maxwell. C'est en ce sens que l'on peut remettre en cause le réductionnisme.

En France, sans le Centre de Cerisy, ce changement profond de l'image classique de la science aurait eu beaucoup plus de mal à diffuser au-delà de cercles scientifiques hyperspécialisés comme les séminaires de l'Institut des Hautes Études Scientifiques.

J'aimerais maintenant donner quelques exemples.

I. DYNAMIQUES NEURONALES

Mon premier exemple sera celui des dynamiques neuronales. L'histoire de ces modèles est assez bien connue. On peut y distinguer quatre périodes principales :

1. Les années 1940-1950 avec les grands fondateurs John von Neumann, Norbert Wiener, Warren McCulloch, Walter Pitts, les fameuses Conférences Macy, etc. Comme l'a montré Jean-Pierre Dupuy, ces savants cherchaient une base physicaliste des facultés cognitives, y compris de leurs composantes “synthétique a priori” au sens transcendantal.
2. À la fin des années 1960 et pendant les années 1970, la première période morphodynamique de René Thom et Christopher Zeeman (théorie des catastrophes).
3. Pendant les années 1980, la vague néo-connexionniste des “Parallel Distributed Processing”.
4. Pendant les années 1990, la synthèse entre les modèles connexionnistes et les modèles morphodynamiques. Comme y a fortement insisté Tim van Gelder, l'editor avec Bob Port de l'ouvrage de référence *Mind as Motion. Explorations in the Dynamics of Cognition* (MIT Press, 1995) :

“If connectionism was the most dramatic theoretical revolution of the 1980s, it appears that dynamics is the connectionism of the 1990s”.

Personnellement, j'ai beaucoup participé aux périodes (2) et (4). C'est Christopher Zeeman qui introduisit l'approche dynamique des relations entre neurologie et psychologie. Dans son article fondateur de 1965 *Topology of the Brain*, il introduisit l'idée que l'activité cérébrale devait être modélisée par des systèmes dynamiques X_w sur des espaces de configuration de très grande dimension $M = I^N$, où $I = [0,1]$ est le domaine d'activité d'un neurone, N étant le nombre des neurones de la population considérée, et où le flot X_w est une dynamique *interne* au réseau qui dépend de paramètres de contrôle *externes*, micro paramètres comme les poids synaptiques et macro paramètres comportementaux ou psychologiques. On identifie alors :

- (i) les états mentaux aux *attracteurs* des flots X_w ,

- (ii) le contenu des états mentaux à la structure *topologique* des attracteurs,
- (iii) le flux temporel de la conscience à une évolution temporelle “lente” des dynamiques internes “rapides” X_w .

Zeeman (1977) a très bien expliqué son fonctionnalisme dynamique *mésoscopique* dans son article de 1976 *Brain modelling* :

“What is needed for the brain is a medium-scale theory. (...) The small-scale theory is neurology: the static structure is described by the histology of neurons and synapses, etc., and the dynamic behaviour is concerned with the electrochemical activity of the nerve impulse, etc. Meanwhile the large-scale theory is psychology: the static structure is described by instinct and memory, and the dynamic behaviour is concerned with thinking, feeling, observing, experiencing, responding, remembering, deciding, acting, etc. It is difficult to bridge the gap between large and small without some medium-scale link. Of course the static structure of the medium-scale is fairly well understood, and is described by the anatomy of the main organs and main pathways in the brain. (...) But what is strikingly absent is any well developed theory of the dynamic behaviour of the medium-scale...

Question: what type of mathematics therefore should we use to describe the medium-scale dynamic? *Answer:* the most obvious feature of the brain is its oscillatory nature, and so the most obvious tool to use is differential dynamical systems.

For each organ O in the brain we model the states of O by some very high dimensional manifold M and model the activity of O by a dynamic on M (that is a vector field or flow on M). Moreover since the brain contains several hierarchies of strongly connected organs, we should expect to have to use several hierarchies of strongly coupled dynamics.”

Lorsqu’à la fin des années 1960, René Thom et Christopher Zeeman introduisirent cette thèse selon laquelle c’était l’approche qualitative des dynamiques neuronales qui devait fonder l’analyse des processus mentaux, cette affirmation fut accueillie avec réserve. Depuis, le développement fulgurant des théories dynamiques et des technologies informatiques des réseaux neuro-mimétiques ont confirmé cette idée au-delà de tout ce qu’on aurait pu imaginer. La différence majeure entre ces deux étapes concerne la question des équations explicites. Zeeman continuait en affirmant :

“such a model must necessarily remain implicit because it is much too large to measure, compute, or even describe quantitatively. Nevertheless such models are amenable in one important aspect, namely their discontinuities.”

On voit apparaître ici l’approche catastrophiste qualitative fondée sur les singularités.

Dans les modèles connexionnistes au contraire, on part des équations *explicites* et, comme elles sont effectivement très compliquées, on utilise des outils de physique statistique.

1. Réseaux neuro-mimétiques

Sous sa forme la plus simple, un réseau de neurones formels consiste en la donnée de N unités u_i dont l'état d'activation y_i varie dans un certain espace d'états S . Les cas les plus utilisés sont $S = \{0, 1\}$, $\{-1, 1\}$, $[0, 1]$. Un état global instantané du réseau est donc décrit par le vecteur $\mathbf{y} = (y_i)_{i=1, \dots, N}$ de l'espace de configurations $M = S^N$. M a le statut d'un *espace interne* au sens de Thom-Zeeman. Les unités u_i sont connectées entre elles par des connexions de poids synaptique w_{ij} . Les w_{ij} déterminent le *programme de calcul* du réseau. Les $w_{ij} > 0$ correspondent à des connexions excitatrices et les $w_{ij} < 0$ à des connexions inhibitrices. On a en général $w_{ii} = 0$.

Le réseau "calcule" de la façon suivante. Chaque neurone u_i reçoit des signaux afférents venant de ses neurones présynaptiques, "calcule" (i.e. change d'état interne en fonction d'une loi de transition) et envoie un signal efférent à ses neurones postsynaptiques.

On définit en général l'input de l'unité u_i comme la somme pondérée des signaux afférents :

$$h_i = \sum_{j=1}^{j=N} w_{ij} y_j, \text{ i.e. } \mathbf{h} = \mathbf{w}\mathbf{y}.$$

Les neurones u_i sont traités comme des automates à seuil dont l'état interne est régi par une loi de transition locale du type :

$$y_i(t+1) = g(h_i(t) - T_i), \text{ i.e. } \mathbf{y}(t+1) = g(\mathbf{h}(t) - \mathbf{T})$$

où T_i est un seuil et g une fonction gain. On a typiquement :

- $g =$ fonction de Heaviside si $S = \{0, 1\}$,
- $g =$ fonction signe si $S = \{-1, 1\}$,
- $g =$ sigmoïde $= 1/(1+e^{-x})$ si $S = [0, 1]$.

Les w_{ij} et les T_i parcourent un espace de contrôle W ayant le statut d'un *espace externe*. La dynamique globale du réseau s'obtient en agrégeant les lois de transition locales et en les itérant. Elle caractérise le réseau comme calculateur.

Dans la limite d'un temps continu, on obtient un grand système d'équations différentielles du type :

$$\dot{\mathbf{y}} = -\mathbf{y} + g(\mathbf{w}\mathbf{y} - \mathbf{T}).$$

Dans la limite d'un continuum spatial de neurones, on obtient des équations aux dérivées partielles (sur des densités) du type :

$$\frac{\partial y(x, t)}{\partial t} = -y(x, t) + g\left(\int [w(x, z)y(z, t) - T(x)] dz\right).$$

Sous l'hypothèse d'un feed-back complet (bouclage des entrées sur les sorties) ce sont les *états asymptotiques stables* du système — *ses attracteurs* — qui sont significatifs. Les attracteurs définissent les *états internes* du réseau. De tels modèles sont par conséquent des cas particuliers de modèles morphodynamiques. Le phénomène dynamique de base y est la capture asymptotique d'un état global instantané \mathbf{y}_0 par un attracteur A .

Ces réseaux neuro-mimétiques sont dotés *d'intentionnalité* au sens phénoménologique car ils définissent des états internes (les attracteurs) qui renvoient à des entités externes (transcendantes) en fonction de leurs dynamiques internes (immanentes). Ils fondent par conséquent la transcendance dans l'immanence. Ils calculent d'une façon radicalement différente de celle d'une machine de Turing : ce sont des calculateurs dynamiques bifurquant d'attracteur en attracteur.

2. La complexité dynamique des réseaux

Les dynamiques $Y_{\mathbf{w}}$ que l'on peut obtenir ainsi sont en général d'une redoutable complexité.

Par exemple, dans le cas (totalement irréaliste sur le plan neurobiologique) où les connexions sont *symétriques*, Hopfield a remarqué que, pour $S = \{-1, +1\}$ et $g =$ fonction signe, les équations du réseau sont celles d'un système de spins en interaction. L'énergie minimisée par la dynamique est alors donnée par :

$$E = -\frac{1}{2} \sum_{i \neq j} w_{ij} y_i y_j + \sum_i T_i y_i .$$

Dans la mesure où les poids synaptiques w_{ij} fonctionnent comme l'analogie de constantes de couplage et où ils sont, de façon très intriquée, à la fois > 0 et < 0 , ces systèmes — qui exemplifient le cas *le plus simple* de réseaux de neurones formels — correspondent au cas *le plus complexe* de systèmes de spins, celui des *verres de spins*. Leur énergie présente un nombre considérable de minima relatifs locaux (états métastables) et pour accéder aux minima absolus globaux (états stables) les méthodes classiques du genre descente de gradient sont inopérantes. Il faut utiliser des algorithmes sophistiqués de physique statistique comme celui du *recuit simulé* (simulated annealing).

Lorsque les poids synaptiques deviennent *asymétriques*, il n'existe plus de fonction énergie et la dynamique peut devenir d'une encore plus grande complexité. Steve Renals et Richard Rohwer ont considéré des systèmes :

$$y_i(t+1) = g \left(r \sum_{j=1}^{j=N} w_{ij} y_j(t) \right)$$

où r est la pente de la sigmoïde. Ils en ont analysé les spectres (les transformées de Fourier) :

$$P_i(k) = \frac{1}{T} \left[\sum_{t=0}^{t=T-1} y_i(t) \exp\left(-\frac{2i\pi kt}{T}\right) \right]^2, \quad k = 0, 1, \dots, T/2$$

et ont étudié les bifurcations présentées par le comportement des états d'activité y_i lorsque r varie. Ils ont retrouvé ainsi de nombreux scénarios classiques de route vers le chaos et en particulier, pour $r \in [12, 14]$, la route par doublement de période, i.e. la cascade sous-harmonique de Couillet-Feigenbaum-Tresser. Ils ont trouvé pour valeur de la constante universelle δ de Feigenbaum intervenant dans la récurrence

$$r_n = r_\infty - \text{cste} \cdot \delta^{-n} \quad (\text{i.e. } \frac{r_n - r_{n-1}}{r_{n+1} - r_n} = \delta)$$

la valeur $\delta = 4.67 \pm 0.04$, ce qui est en excellent accord avec la valeur standard $\delta = 4,6692016091029909$.¹

H. Sompolinsky, M. Samuelides et B. Tirozzi ont aussi étudié de tels systèmes lorsque N devient très grand et lorsque les poids synaptiques w_{ij} (asymétriques) sont des variables aléatoires (par exemple gaussiennes) de moyenne nulle et de variance w^2/N . Pour la valeur critique $rw = 1$, ils présentent une transition de phase d'un régime convergent vers un régime chaotique. Samuelides a en particulier étudié les routes vers le chaos dans le cas où les systèmes sont *dilués* (presque tous les $w_{ij} = 0$), où il existe des seuils T_i et où les variables aléatoires w_{ij} ne sont plus centrées.

De très nombreux résultats de cet ordre montrent qu'il est devenu désormais possible de donner un statut rigoureux à la thèse selon laquelle les contenus mentaux sont des attracteurs de systèmes dynamiques implémentés dans des réseaux de neurones et que, par conséquent, les fonctions cognitives peuvent naturellement être conçues en termes de *morphodynamique et de thermodynamique neurales*. C'est ce que disait déjà Christopher Zeeman en 1965. La différence fondamentale est que dans les travaux que nous venons d'évoquer les dynamiques internes sont explicites.

3. Catégorisation et apprentissage

Les modèles connexionnistes ont développé massivement certains aspects de cette thèse, entre autres pour les phénomènes de catégorisation et d'apprentissage. Mais ils restent pourtant encore très en deçà des propositions de Zeeman et Thom.

Par exemple en ce qui concerne les phénomènes de *catégorisation*, ceux-ci résultent dans les modèles connexionnistes de la partition de l'espace interne M en bassins d'attraction d'attracteurs qui fonctionnent quant à eux comme autant de *prototypes*. Ce que les psychologues appellent les "gradients de typicalité" s'interprètent alors comme des fonctions de Liapounov sur ces bassins. Quant à l'opposition typique /

¹ Cf. Renals, Rohwer [1990].

non typique (i.e. générique / spécial), elle se trouve géométrisée à travers l'opposition entre l'intérieur et les bords des bassins d'attraction.

Mais les modèles morphodynamiques initiaux étaient allés d'emblée beaucoup plus loin en considérant les catégorisations induites dans des espaces *externes* par le déploiement des *singularités* des dynamiques internes. De telles catégorisations externes interviennent naturellement chaque fois que des états internes dépendent de paramètres de contrôle. Tel est par exemple le cas en phonétique où les dynamiques internes sont des dynamiques acoustiques (définissant en particulier les formants, i.e. les pics du spectre continu modulant le spectre harmonique, pics qui reflètent la géométrie des résonateurs du tractus vocal) et où les contrôles externes sont des indices acoustiques comme le voisement ou des indices articulatoires comme le point d'articulation.

En ce qui concerne *l'apprentissage* c'est la problématique des espaces externes qui domine. L'apprentissage peut en effet se concevoir comme *le problème inverse* de celui qui, étant donnée la matrice \mathbf{w} des poids synaptiques, consiste à trouver les attracteurs de la dynamique $Y_{\mathbf{w}}$. Il s'agit de se donner a priori des attracteurs et de trouver un \mathbf{w} . Certains algorithmes ont été développés à cette fin, en particulier celui dit de *rétropropagation* qui consiste à partir d'une matrice initiale \mathbf{w}_0 , à calculer l'écart entre les attracteurs de $Y_{\mathbf{w}_0}$ et les attracteurs désirés et à rétropropager l'erreur en ajustant \mathbf{w}_0 . De tels algorithmes définissent des dynamiques externes *lentes* dans les espaces externes W de poids synaptiques.

Mais, là encore, le point de vue morphodynamique initial reste très en avance sur les théories connexionnistes. En effet, il met en relief le fait que dans de tels systèmes lents / rapides il existe dans W *une stratification catégorisant* les $Y_{\mathbf{w}}$ en différents types qualitatifs. Cette stratification est une partition de W par un fermé catastrophique K (un système de frontières). Les algorithmes comme la rétropropagation fournissent des dynamiques externes qui ne sont définies que dans les strates ouvertes (les composantes connexes de $W-K$, i.e. les catégories de $Y_{\mathbf{w}}$). Mais le propre d'un apprentissage est en général de complètement transformer le type qualitatif de $Y_{\mathbf{w}}$. Il faut donc comprendre comment les dynamiques de rétropropagation définies sur les différentes strates ouvertes se recollent le long de K . Ce problème est encore totalement ouvert.

4. Le problème du liage, "labeling hypothesis" et réseaux d'oscillateurs

Mais on peut aller encore beaucoup plus loin, et l'histoire a également donné raison à Zeeman en ce qui concerne les activités *oscillatoires* du cerveau.

4.1. Le problème méréologique du liage

Le problème cognitif central du *binding* ou liage est celui de la “*constituance*” et de la “*compositionalité*” des représentations mentales, par exemple des scènes perceptives ou des énoncés linguistiques. C'est l'ancien problème philosophique du tout et des parties, le problème méréologique des structures composées. Il est central pour tous les traitements cognitifs de haut niveau car ces derniers sont causalement sensibles à la structure en constituants des représentations mentales (cela est particulièrement évident dans le cas du langage). On l'appelle le “binding problem”, celui des mécanismes de liage entre les constituants. Il est facile à comprendre. Au niveau neuronal, les représentations mentales sont implémentées de façon *distribuée* sur un très grand nombre d'unités élémentaires. Comment donc éviter la “catastrophe de superposition” dissolvant les parties dans le tout ? Comment arriver à extraire des constituants et des relations entre ces constituants ? Comment coder les liens relationnels ?

L'une des hypothèses actuellement les plus discutées — qui remonte à des travaux de Christoph von der Malsburg (1981) — repose sur le *codage temporel fin* des processus mentaux. Elle est que la cohérence structurale, l'unité, des constituants d'une représentation se trouve encodée dans la dynamique de l'activité neuronale sous-jacente, dans ses corrélations temporelles et, plus précisément, dans la *synchronisation* (accrochage de fréquence et de phase) de réponses neuronales oscillatoires. L'idée est donc que la cohérence temporelle rapide (de l'ordre de la ms) code la cohérence structurale. La *phase* commune des oscillateurs synchronisés implémentant un constituant peut alors servir d'étiquette (de label) pour ce constituant dans des processus de traitement ultérieurs. D'où le nom de “labeling hypothesis”.

On voit ainsi l'hypothèse de Zeeman non seulement être confirmée mais ouvrir un énorme chantier expérimental, théorique et modélisateur.

4.2. Les résultats expérimentaux

Il existe de nombreuses confirmations expérimentales d'oscillations synchronisées (bande de fréquence autour de 40 Hz) des colonnes et hypercolonnes corticales (en particulier dans le cortex visuel primaire), la synchronisation étant sensible à la constituance des stimuli, à la cohérence de leurs constituants (travaux de Reinhard Eckhorn, Charles Gray, Wolf Singer, Peter König, Andreas Engel, 1992).

Ces résultats ont été fort débattus et sont en partie controversés. Ils sont fort délicats à obtenir et de nombreux paramètres y interfèrent. Mais ils valident néanmoins une idée directrice. Même si l'on simplifie et idéalise celle-ci, elle conduit, comme cela a été le cas avec les réseaux de neurones formels, à des problèmes mathématiques d'une redoutable difficulté.

4.3. La modélisation

En ce qui concerne la modélisation, on montre d'abord que des colonnes corticales peuvent effectivement fonctionner comme des oscillateurs élémentaires. Elles sont constituées d'un grand nombre de neurones excitateurs et inhibiteurs. En moyennant sur ces deux groupes les équations standard, on obtient un système de deux équations (équations de Wilson-Cowan). On montre alors que l'état d'équilibre subit une bifurcation de Hopf lorsque l'intensité du stimulus dépasse un certain seuil.

Les modèles dont la formulation est de simplicité maximale consistent à étudier des réseaux constitués d'un grand nombre N d'oscillateurs F_i dont la fréquence propre ω_i dépend de l'intensité du stimulus à la position i . Soient θ_i leurs phases et $\varphi_i = \theta_{i+1} - \theta_i$ leurs différences de phases. Les systèmes les plus courants sont du type :

$$\dot{\theta}_i = \omega_i - \sum_{j=1}^{j=N} K_{ij} \sin(\theta_i - \theta_j)$$

où les K_{ij} sont des constantes de couplage. Même si leur formulation est très simple, ce sont des systèmes typiquement complexes que l'on doit étudier avec des méthodes de physique statistique.

4.4. Le modèle de Kuramoto : synchronisation = transition de phase

Par exemple, dans le cas d'une seule constante de couplage et d'une totale connectivité, Y. Kuramoto (1987) a analysé en détail le système :

$$\dot{\theta}_i = \omega_i - \frac{K}{N} \sum_{j=1}^{j=N} \sin(\theta_i - \theta_j).$$

Dans ce dessein, il a introduit le *paramètre d'ordre* qu'est la phase moyenne :

$$Z(t) = |Z(t)| e^{i\theta_0(t)} = \frac{1}{N} \sum_{j=1}^N e^{i\theta_j(t)}$$

et a étudié le système équivalent :

$$\dot{\theta}_i = \omega_i - K |Z| \sin(\theta_i - \theta_0).$$

Si les fréquences ω_i sont tirées au hasard suivant une loi $g(\omega)$ représentant les régularités statistiques de l'environnement (en prenant un repère tournant on peut supposer g centrée sur 0), la synchronisation globale est une *transition de phase* s'effectuant pour la valeur critique $K_c = 2/\pi g(0)$ de la constante de couplage.

Kuramoto cherche d'abord des solutions $Z = \text{constante}$. Après avoir classé les oscillateurs en deux groupes : le S -groupe des oscillateurs pouvant se synchroniser i.e. satisfaisant

$$\dot{\theta}_i = 0 \text{ et donc } \left| \frac{\omega_i}{KZ} \right| \leq 1$$

et le D -groupe des oscillateurs ne le pouvant pas parce que

$$\left| \frac{\omega_i}{KZ} \right| > 1,$$

il montre que seul le S -groupe intervient dans la synchronisation. En écrivant que

$$Z = \int_0^{2\pi} n_0(\theta, t) e^{i\theta} d\theta$$

où $n_0(\theta, t)$ est la distribution des phases à l'équilibre au temps t et en écrivant que

$$n_0(\theta, t) d\theta = g(\omega) d\omega \text{ avec } \omega = K|Z| \sin(\theta - \theta_0),$$

il obtient une équation d'auto-consistance $Z = S(Z)$ qu'il développe au voisinage de $Z = 0$. Il obtient ainsi l'équation

$$\varepsilon Z - \beta |Z|^2 Z = 0$$

avec $\varepsilon = \frac{K - K_c}{K_c}$, $\beta = -\frac{\pi}{16} K_c^3 g''(0)$.

L'analyse de la stabilité des solutions montre que la solution $Z = 0$, qui est stable pour $K \approx 0$ (oscillateurs découplés), devient instable à la traversée de $Z = Z_c$.

Kuramoto établit ensuite, sous une hypothèse de quasi-adiabaticité, l'évolution du paramètre d'ordre Z . Il obtient une équation du type :

$$\xi \frac{dZ}{dt} |KZ|^{-1} = \varepsilon Z - \beta |Z|^2 Z .$$

Il étudie ensuite les fluctuations, en particulier au voisinage du point critique lorsqu'elles deviennent géantes et entraînent la transition de phase.

Tout cela montre que la synchronisation est un phénomène typique d'organisation collective émergente.

III. MODELES MORPHOLOGIQUES ET THEORIE DES PATTERNS

Mon second exemple portera sur la théorie des formes.

1. Le principe morphodynamique

C'est René Thom qui a défini le premier de façon à la fois mathématique et générale ce que sont une morphologie et un processus morphogénétique. L'idée fondamentale est de considérer qu'en chaque point w de l'espace W du substrat de la forme il existe une dynamique *locale*, dite dynamique interne, X_w qui définit la physique ou la chimie ou le métabolisme local du substrat. Ce régime local, cet état interne du substrat, se manifeste phénoménologiquement par des qualités sensibles (couleur, texture, etc.). Les rapports de voisinage spatial entre les différents points w induisent alors des *couplages* entre les dynamiques internes locales. Celles-ci interagissent et des *instabilités* peuvent donc se produire. Cela entraîne des bifurcations des régimes locaux, des brisures des symétries du substrat, brisures qui entraînent à leur tour des discontinuités qualitatives dans l'apparence du substrat. Et ce sont ces ruptures

d'homogénéité qui engendrent enfin les formes. L'idée principale est donc de considérer l'espace et le temps non plus comme un simple contenant pour des objets mais comme une *espace de contrôle* permettant de faire interagir des dynamiques internes locales.

Ce point de vue fournit un cadre théorique unitaire à tout un ensemble de travaux à la fois anciens et récents. Je citerai deux exemples, celui des équations de réaction-diffusion et celui des champs continus d'oscillateurs.

2. Équations de réaction-diffusion (Turing)

Les équations de réaction-diffusion introduites par Turing en théorie de la morphogenèse permettent de comprendre l'émergence de motifs morphologiques macroscopiques dans les réactions chimiques. Elles couplent des équations cinétiques de réaction décrivant des interactions moléculaires locales et des équations de diffusion décrivant des phénomènes de transport. La diffusion produit de l'uniformisation, elle homogénéise. C'est par excellence un processus destructeur de morphologies. Mais si le milieu est le siège de réactions chimiques avec catalyse et autocatalyse (les équations différentielles de la cinétique chimique exprimant l'évolution temporelle des concentrations des espèces chimiques sont alors non linéaires) et s'il est loin de l'équilibre thermodynamique (système ouvert) alors il peut y avoir des morphologies spatio-temporelles complexes qui émergent de façon stationnaire et qui sont engendrées par des processus d'auto-organisation. Le caractère explosif de l'autocatalyse se trouve inhibé par d'autres réactifs et, suivant les vitesses de diffusion relatives des produits de la réaction, les morphologies peuvent être très différentes.

Par exemple si A est un activateur auto-catalytique et si H est un inhibiteur dont la synthèse est catalysée par A , alors à partir d'une situation initiale homogène on peut obtenir des motifs périodiques. Une petite fluctuation de A produit par autocatalyse un pic local de A . Mais cela amplifie aussi la concentration de H localement. Mais si H diffuse plus vite que A , la formation de A ne sera inhibée par H que latéralement et non pas au centre du pic. D'où un pic de A bordé par un manque de A .

Un exemple de système d'équations non linéaires modélisant un tel système sont par exemple :

$$\begin{cases} \frac{\partial a}{\partial t} = \rho \frac{a^2}{h} - \mu_a a + D_a \frac{\partial^2 a}{\partial x^2} + \sigma_a \\ \frac{\partial h}{\partial t} = \rho a^2 - \mu_h h + D_h \frac{\partial^2 h}{\partial x^2} + \sigma_h \end{cases}$$

où $a(x, t)$ et $h(x, t)$ sont les concentrations respectives de l'activateur A et de l'inhibiteur H , où les termes non linéaires en a^2 expriment l'autocatalyse de A et la catalyse de H par A , où le terme en $1/h$ exprime l'inhibition de la production de A par H , où les termes linéaires $-\mu_a a$ et $-\mu_h h$ sont des termes de dégradation (les constantes μ sont des durées

de vie de molécules et $\mu_a < \mu_h$: H se dégrade plus vite que A), où les termes $D_a \frac{\partial^2 a}{\partial x^2}$ et $D_h \frac{\partial^2 h}{\partial x^2}$ sont des termes de diffusion avec $D_a \ll D_h$ (H diffuse plus vite que A), et où enfin les termes constants > 0 σ_a et σ_b garantissent que les espèces chimiques A et H restent toujours présentes.

On peut obtenir ainsi des morphologies complexes, par exemple des structures en bandes (structures localement simples mais globalement complexes avec des défauts, points d'arrêt, dislocations, etc. comme dans les cristaux liquides) (cf. figure 1).

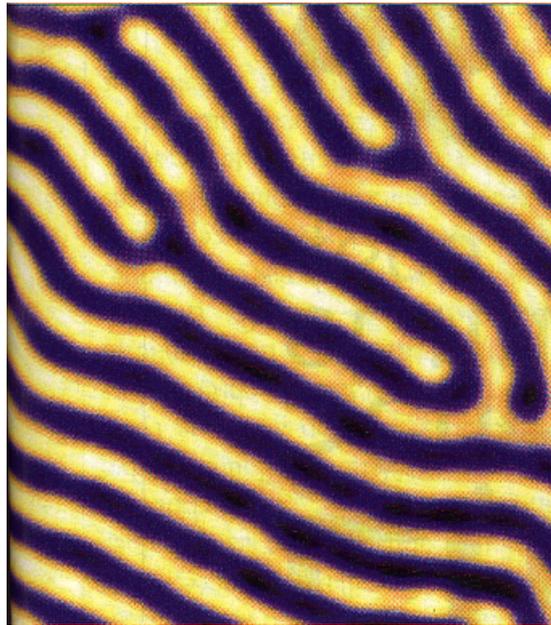


Figure 1. *Bandes parallèles produites par un processus chimique de réaction-diffusion. On remarquera l'existence de défauts. (D'après De Kepper [1998]).*

3. Réseaux d'oscillateurs (Coulet)

En analysant les instabilités de *champs continus* d'oscillateurs, Pierre Coulet a montré comment on pouvait engendrer un nombre considérable de formes de type différents. On considère par exemple des oscillateurs faiblement couplés par leurs relations topologiques de voisinage et soumis à un forcing externe possédant une fréquence voisine du double de leur fréquence propre. La variable locale observée peut être l'amplitude ou la phase de l'oscillateur. L'amplitude de la modulation et l'écart à la résonance sont des paramètres.

En passant à la limite d'un continuum d'oscillateurs dont le paramètre d'ordre (la phase moyenne) Z dépend de la position spatiale, on obtient des équations du type :

$$\frac{\partial Z}{\partial t} = \lambda Z - \mu |Z|^2 Z + \gamma_n \bar{Z}^{n-1} + \nu \Delta Z,$$

où λ , μ et ν sont des paramètres complexes et γ_n un paramètre réel. Ces équations complexifient celle de Kuramoto en introduisant un terme de *diffusion spatiale* et le terme de coefficient γ_n .

Ces oscillateurs peuvent se synchroniser et se désynchroniser localement. En introduisant de la diffusion, on obtient une très riche variété de patterns spatiaux : turbulence développée, défauts, ondes spirales, cellules hexagonales, réseaux de bandes, etc. (cf. figures 2 et 3).

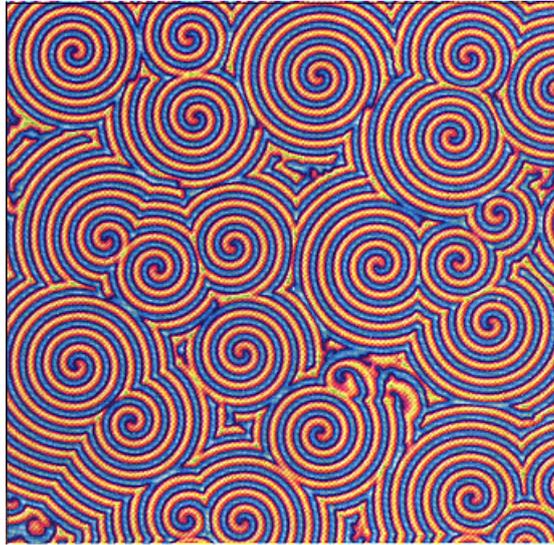


Figure 2. Ondes spirales induites dans un champ continu d'oscillateurs. (D'après Coulet-Emilsson [1992]).

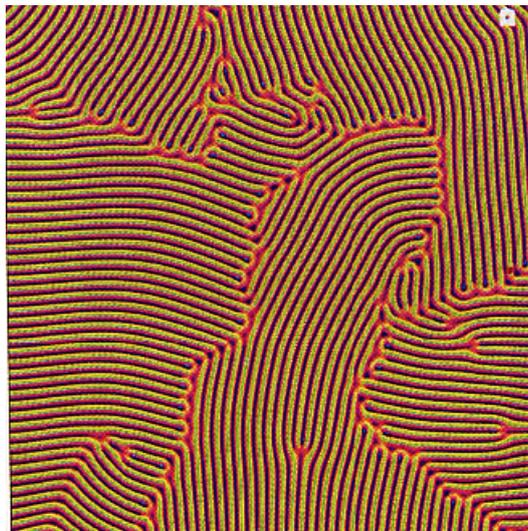


Figure 3. Textures en bandes parallèles induites dans un champ continu d'oscillateurs. (D'après Coulet-Emilsson [1992]).

4. Modèles de patterns

Hans Meinhardt du Max Planck Institut a développé des modèles pour des motifs morphologiques comme ceux des coquilles. La croissance d'une coquille se fait par couches successives d'accrétion de matériau calcifié le long du bord du manteau. L'état de pigmentation d'une cellule est déterminé par la cellule sous-jacente et l'état des cellules voisines. Une coquille peut donc être considérée comme un diagramme "position \times temps de développement". Par exemple dans l'espèce *Conus marmoreus*, on peut supposer que l'activateur produisant la pigmentation "noir" se déclenche aléatoirement, s'autocatalyse et diffuse lentement. D'où la formation de triangles noirs. Mais quand la production a duré assez longtemps (donc après un certain délai), l'inhibiteur se déclenche et diffuse rapidement. D'où l'arrêt brutal de la diffusion et "l'extinction" des bases de ces triangles. Mais l'activateur reste actif aux bords de ces intervalles car il a été tardivement déclenché. D'où de nouveaux triangles de diffusion. On obtient ainsi des cascades caractéristiques (cf. figure 4).

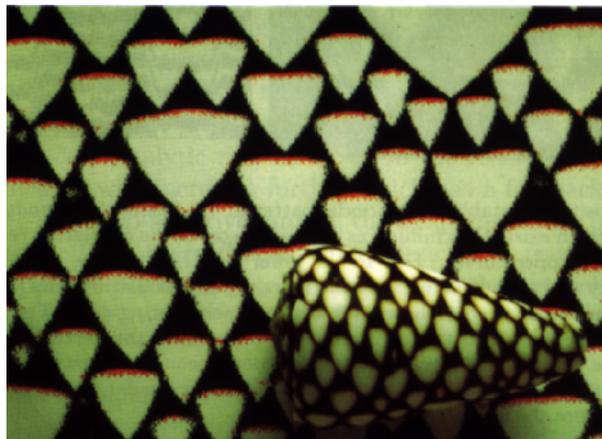


Figure 4. Simulation des motifs d'une coquille de l'espèce *Conus marmoreus*. La coquille est photographiée sur le fond de la simulation. (D'après Meinhardt [1995].)

IV. MODELES DE PERCEPTION

Disons maintenant un mot des modèles de perception.

1. La difficulté du problème

Dans le moindre des problèmes de vision, il faut passer d'une matrice de pixels (activités des photorécepteurs) à une description de la scène de nature conceptuelle et linguistique. Pour ce faire, il faut au minimum résoudre les problèmes suivants :

- (i) Segmentation des images locales 2D ; établissement de bords et de discontinuités qualitatives dans les zones où il y a une forte variation des qualités sensibles comme l'intensité lumineuse, la couleur, la texture, etc.
- (ii) Recollement des images 2D en une scène globale (saccades visuelles) et interprétation de certains bords comme des contours apparents d'objets 3D (stéréopsie).
- (iii) Reconnaissance de formes et résolution du problème métréologique des relations tout / parties. Coopération des textures, ombres, etc. pour la résolution de ce problème.
- (iv) Utilisation des mouvements locaux pour reconstituer, par recollement et intégration, des mouvements globaux (en général différents) d'objets 3D (problème particulièrement difficile).
- (v) Reconnaissance d'interactions types entre objets.

Résoudre ces problèmes exige des algorithmes extrêmement sophistiqués. Pourquoi sont-ils si difficiles à résoudre ? Essentiellement parce que la détection de structures organisées (de patterns) à partir d'inputs bruités et ambigus constitue un *problème mal posé*. Comme l'explique David Mumford (Médaille Fields 1974 en géométrie algébrique qui, depuis plusieurs années, se consacre à la vision computationnelle et aux neurosciences) :

- (i) Dans la moindre scène perceptive, les patterns observés permettent d'inférer de l'information sur les causes qui les produisent. Ces causes sont des *variables cachées*.
- (ii) Il y a beaucoup trop de variables cachées en jeu et les situations sont beaucoup trop compliquées pour être modélisées de façon déterministe. La possibilité de faire des inférences présuppose donc des modèles *stochastiques*.
- (iii) Mais en même temps, ces modèles stochastiques doivent respecter les patterns. C'est la contrainte la plus difficile à réaliser. Ils doivent, qui plus est, pouvoir être *appris à partir des données* — il s'agit là d'une autre énorme contrainte — et validables par exemplification (sampling). Les inférences sont alors effectuées en utilisant par exemple la règle de Bayes.

Le problème est que les signaux (les données, les inputs) ont une variabilité très grande et les variables cachées y sont très subtilement encodées. Si l'on veut inférer un état du monde S (i.e. les valeurs des variables définissant notre représentation du monde, par exemple la position, le mouvement, la forme, le type d'interaction spatio-temporelle d'objets individuels dans l'espace externe global 3D) à partir des observations I (des données acquises et mesurées par certains capteurs, par exemple l'état d'activité des photorécepteurs de la rétine), on doit calculer la probabilité conditionnelle a posteriori $p(S|I)$. L'égalité des probabilités

$$p(S, I) = p(S|I)p(I) = p(I, S) = p(I|S)p(S)$$

implique le théorème de Bayes :

$$p(S|I) = \frac{p(I|S)p(S)}{p(I)} = \frac{p(I|S)p(S)}{\sum_S p(I|S)p(S)}$$

La probabilité conditionnelle $p(I|S)$ correspond au *problème direct* (simple) : si l'on connaît l'état du monde (par exemple les objets 3D, les réflectances, les sources de lumière), alors on peut reconstruire I (l'image sensorielle 2D). Toutefois, à cause du bruit et de l'imprécision du modèle, on ne peut pas accéder exactement à I mais seulement à $p(I|S)$. La probabilité conditionnelle $p(S|I)$ correspond au contraire au *problème inverse*. Il est incroyablement difficile car la reconstruction du monde 3D à partir d'images 2D n'est pas, en tant que problème inverse, un problème bien posé.

Tout tourne autour de $p(S)$, ce que l'on appelle le *prior model* encodant nos connaissances *préalables* sur le monde. Comme l'explique également David Mumford :

“In general the image alone is not sufficient to determine the scene and, consequently, the choice of priors becomes critically important. They *embody the knowledge* of the patterns of the world that the visual system use to make valid 3D inferences”.

Sans prior model, il est impossible d'apprendre et de faire des inférences.

2. L'exemple du modèle de Mumford-Shah

Un exemple est le modèle de segmentation des images dû à David Mumford lui-même. Dans un article de synthèse “*Bayesian rationale for the variational formulation*”, Mumford explique que

“one of the primary goals of low-level vision is to segment the domain W of an image I into the parts W_i on which distinct surface patches, belonging to distinct objects in the scene, are visible”

et que l'approche mathématique de base à ce problème de segmentation consiste à utiliser les différentes sources d'information de bas niveau

“for *splitting* and *merging* different parts of the domain W ”

de façon optimale. Il propose alors une approche *variationnelle* qui consiste à minimiser une fonctionnelle “énergie” $E(u, K)$ où K est une segmentation de W partitionnant W en domaines ouverts W_i (les composantes connexes de $W - K$) et u une approximation de I qui est régulière sur les W_i tout en pouvant présenter des discontinuités le long de K .

Ces modèles variationnels peuvent être interprétés comme des modèles probabilistes à partir de l'équivalence $E(u, K) = -\text{Log}(p(u, K))$, p étant une probabilité définie sur l'espace des segmentations (u, K) possibles.

Dans ces modèles bayesiens, il existe une partie a priori (le prior model) et une partie concernant les données (le data model). Le prior model consiste à prendre comme a priori la description phénoménologique qui était à la base des modèles *morphologiques* de Thom. Cela signifie que l'on cherche à approximer I par une

fonction u différentiable par morceaux sur (W, K) en imposant a priori que u varie le moins possible dans les zones homogènes W_i et que le bord K ne soit pas trop compliqué et irrégulier. L'approximante u est une interprétation morphologique du signal I , la meilleure interprétation contrainte par l'a priori synthétique du "merging and splitting".

On obtient ainsi le modèle dit de Mumford-Shah (1989) :

$$E(u, K) = \int_{W-K} |\nabla u|^2 dx + \int_W (u-I)^2 dx + \int_K d\sigma$$

On peut le rendre *multi-échelle* en introduisant différents poids pour les différents termes.

La minimisation de E est un compromis entre :

- (i) l'homogénéité des composantes connexes de $W - K$: si $u = \text{constante}$ alors $\nabla u = 0$ et $\int_{W-K} |\nabla u|^2 dx = 0$; la minimisation du premier terme force donc u à être la plus constante possible sur les domaines W_i ;
- (ii) l'approximation de I par u : si $u = I$ alors $\int_W (u-I)^2 dx = 0$; la minimisation du deuxième terme force donc u à rester proche de I ;
- (iii) la parcimonie et la régularité des bords : ils sont mesurés par la longueur globale L de K , $L = \int_K d\sigma$; la minimisation du troisième terme force donc les bords à être peu nombreux et le plus réguliers possible.

Un tel algorithme variationnel optimise la façon dont on peut fusionner des pixels voisins en domaines homogènes séparés par des discontinuités qualitatives. Il est très difficile à résoudre car les trois termes sont en compétition et portent sur des entités de dimensions *différentes*. De nombreux travaux lui sont actuellement consacrés.

Alessandro Sarti et Giovanna Citti ont récemment démontré le beau théorème qu'un modèle de synchronisation d'oscillateurs généralisant le modèle de Kuramoto (les fréquences propres des oscillateurs dépendent de l'intensité du stimulus I au point considéré, les oscillateurs sont couplés *localement* et la synchronisation est donc une homogénéisation locale en régions séparées par des lignes catastrophiques de désynchronisation) *converge* vers le modèle de Mumford-Shah. Ce dernier possède par conséquent une forte plausibilité neurophysiologique.

V. MODELES FORMELS D'ORGANISATION ET DE COOPERATION

Mon dernier exemple concernera les modèles de coopération dans une population d'agents. Il est en résonance avec le Colloque de Cerisy de 1982 sur l'auto-organisation et le Colloque sur Hayek organisé à Cerisy en 1999 par Alain Leroux et Robert Nadeau.

1. Hayek et les normes du sens commun

Il s'agit de formaliser des règles du *sens commun* qui sont le résultat d'une évolution culturelle et peuvent donc être assimilées à un apprentissage collectif sélectionnant des stratégies qui sont extrêmement performantes et pratiquement impossibles à trouver de façon planifiée. La maîtrise de la complexité de tels processus évolutionnaires permet d'engendrer par synthèse computationnelle des évolutions culturelles *virtuelles* et d'enrichir expérimentalement le domaine des stratégies héritées du sens commun.

Il s'agit là d'un énorme domaine concernant la *naturalisation du sens commun*. Le développement de telles “sciences du sens commun” est fondamental car il permet de dépasser certains aspects négatifs du rationalisme classique. Hayek critiquait dans le constructivisme planificateur politique le fait

“que c’est uniquement ce qui est rationnellement justifiable, démontrable par l’expérimentation, observable et susceptible de faire l’objet d’un rapport qui peut susciter une adhésion (...) et que tout le reste doit être rejeté. Les règles traditionnelles *qui ne peuvent être démontrées* doivent être rejetées.”²

Mais tout change si l’on peut montrer que des règles traditionnelles sont en fait justifiables rationnellement grâce à de nouveaux modèles et démontrables par l’observation grâce à de nouvelles méthodes (computationnelles) d’expérimentation virtuelle.

Chez Hayek, la critique du rationalisme constructiviste en matière de politique, de droit et de morale repose sur le constat que constructivisme planificateur nie la complexité organisationnelle et élimine l'intelligence collective résultant de l'évolution historique. Or, à cause de l'irréductibilité de la complexité endogène des systèmes organisés, la planification et la prévision sont négatives non seulement en fait mais en droit. Dès qu'elles cessent d'être régulatrices pour devenir normatives et déterminantes, elles font chuter la complexité interne et trivialisent les dynamiques auto-organisationnelles. Elles paralysent la main invisible de Smith.

Les structures organisationnelles ne sont pas récapitulables dans une intelligence individuelle. Il est cognitivement impossible de posséder une connaissance complète des causes et des effets des actions. L'information nécessaire sur la société serait tellement énorme et complexe qu'elle serait inintégrable par les agents. Il y faudrait une omniscience laplacienne ou un intellect “archétypique” au sens de Kant et non pas les intellects “ectypiques” de responsables politiques. À cause de la complexité même de leurs interactions et du caractère distribué des connaissances associées, on ne peut pas aller au-delà de leur coordination cohérente.

² Hayek [1988], p. 85.

Sur le plan cognitif (individuel et social), il existe une origine évolutionniste des règles de perception et de conduite, des conventions et des normes. Ces patterns d'action sont le résultat d'une sélection culturelle — donc d'un *apprentissage* collectif-historique — fonctionnant comme un processus concurrentiel ayant avantage les groupes les ayant adoptés. Ils permettent d'agir sans devoir à chaque fois récapituler toutes les expériences permettant d'agir. Le *sens commun* est lui-même un ensemble de connaissances tacites et de schèmes pratiques permettant, en *schématisant* (et donc en simplifiant) l'expérience de notre environnement et en la ramenant à des situations *génériques* valables par défaut, d'y agir et d'y prévoir sans nous laisser submerger par le haut flux d'informations non pertinentes (non significatives) charriées par sa complexité. Pour Hayek, les *normes* du sens commun ne sont pas des contraintes mais possèdent une valeur *cognitive*. Les traditions sont des “savoirs incorporés” d'origine “phylogénétique” (au sens de l'évolution culturelle) et il est par conséquent rationnel de s'y conformer “ontogénétiquement”.

2. L'exemple des jeux évolutionnistes

Depuis les travaux fondateurs de Robert Axelrod, on a beaucoup travaillé sur certains systèmes complexes adaptatifs “sociaux” pour lesquels on sait analyser les mécanismes causaux sous-jacents, à savoir les jeux évolutionnistes. Un exemple typique en est celui du *dilemme du prisonnier itéré* (IPD). On en trouvera une excellente présentation synthétique dans le numéro spécial de *Pour la Science* sur les *Mathématiques Sociales* (juillet 1999) où Jean-Paul Delahaye (Université de Lille) résume des travaux prolongeant ceux de Robert Axelrod, William Poundstone, M.artin Nowak et Karl Sigmund.

2.1. Le dilemme du prisonnier

Il y a deux joueurs *A* et *B*. Pour chaque comportement possible des joueurs, à savoir *d* = défection (trahir) et *c* = coopération, la matrice du jeu donne les gains (payoffs) des joueurs (en haut à droite de chaque case pour le joueur colonne *A*, en bas à gauche pour le joueur ligne *B*). La matrice fait intervenir 4 gains :

$T = (d, c) =$ Temptation,

$S = (c, d) =$ Sucker

$R = (c, c) =$ Reward,

$P = (d, d) =$ Punishment.

Pour que le jeu soit intéressant, il faut que les gains satisfassent les conditions :

$$T > R > P > S \text{ et } (T + S)/2 < R.$$

En voici un exemple typique :

	$A(c)$	$A(d)$
$B(c)$	$R = 3$	$T = 5$
	$R = 3$	$S = 0$
$B(d)$	$S = 0$	$P = 1$
	$T = 5$	$P = 1$

Comportements :

$d =$ défection (trahir), $c =$ coopération

Gains :

$$T = (d, c) = 5, S = (c, d) = 0$$

$$R = (c, c) = 3, P = (d, d) = 1$$

Conditions :

$$T = 5 > R = 3 > P = 1 > S = 0$$

$$(T + S)/2 = 5/2 < R = 3$$

Ce jeu très simple représente une situation où la rationalité individuelle entre en conflit avec la rationalité collective. En effet

1. Si le joueur colonne A joue c , alors le joueur ligne B gagne R s'il joue c et T s'il joue d . Comme $T = 5 > R = 3$, B a donc intérêt à jouer d .
2. Si le joueur colonne A joue d , alors le joueur ligne B gagne S s'il joue c et P s'il joue d . Comme $P = 1 > S = 0$, B a donc intérêt à jouer d .
3. Si B est rationnel, il jouera donc d quel que soit le comportement de A . On dit que la stratégie d domine strictement la stratégie c : d fait mieux que c quel que soit le comportement de l'autre joueur.
4. Il en va de même pour A par symétrie.
5. Le résultat du jeu est donc $(d, d) = (\text{lose}, \text{lose})$, non coopération qui conduit au mauvais gain collectif ($P = 1, P = 1$).
6. Or clairement, la coopération $(c, c) = (\text{win}, \text{win})$ conduisant au gain collectif ($R = 3, R = 3$) aurait été une stratégie bien supérieure.

Avec une telle matrice de gain, la double stratégie (d, d) est le seul *équilibre de Nash* du jeu, i.e. la double stratégie telle que chaque joueur fait moins bien s'il change de stratégie de façon unilatérale.

On peut généraliser cet exemple de multiples façons, par exemple en introduisant des asymétries, des inégalités larges, un comportement neutre (refus de jouer), des joueurs multiples, des probabilités, etc. Mais le phénomène que nous venons de décrire se révèle *robuste* pour les parties à un seul coup. Comment expliquer alors à partir de ce constat initial la façon dont la coopération peut être sélectionnée évolutivement.

Remarque.

La condition $(T + S)/2 < R$ a pour fonction de rendre le quadrilatère des payoffs associés à $(c, d) - (c, c) - (d, c) - (d, d) - (c, d)$ convexe. Cela est important pour les stratégies

mixtes où le joueur colonne A joue c avec la probabilité p et le joueur ligne B joue c avec la probabilité q . L'espérance de payoff de B

$$(1-p)qT + pqR + (1-p)(1-q)P + p(1-q)S$$

est alors interne au quadrilatère.

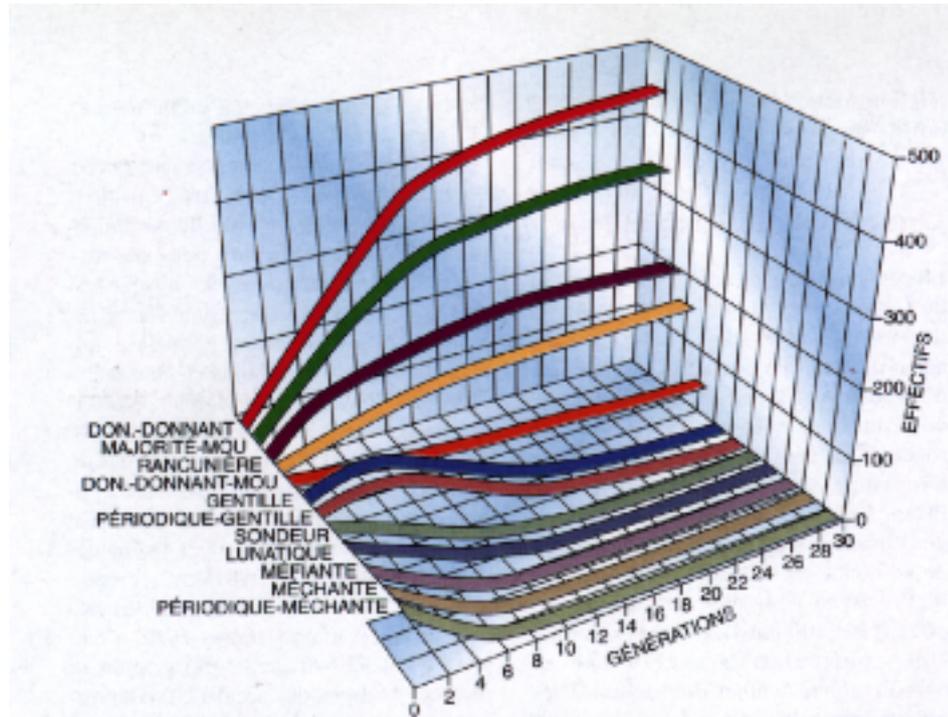
2.2 Le dilemme du prisonnier itéré (IPD)

La situation change du tout au tout lorsque l'on *itère* le jeu car la trahison peut alors être sanctionnée et la coopération récompensée. On peut dans ce cas introduire de véritables *stratégies*. On suppose le nombre de coups indéterminé pour éviter la *backward induction* (la possibilité de définir une stratégie en remontant à partir des résultats voulus au dernier coup) qui redonne le comportement (d, d) et l'on teste des stratégies du genre (cf. l'article de Jean-Paul Delahaye) : G = “gentille” = toujours c ; M = “méchante” = toujours d ; TFT = “donnant-donnant” (tit for tat) = d'abord c puis jouer ce que l'autre a joué à la partie précédente ; R = “rancunière” = c mais toujours d dès que l'autre a trahi une fois, etc. On confronte ces stratégies sur un grand nombre de parties (par exemple 1000) et l'on étudie leurs scores. La notion d'équilibre de Nash (EN) doit être renforcée car la stratégie (Id, Id) qui itère l'EN (d, d) des parties à un coup reste un EN. Mais comme beaucoup de stratégies donnent le même résultat que Id en jouant contre Id , il y a trop d'EN. D'où le concept plus contraint de “subgame perfect equilibrium” qui est un EN pour tout sous-jeu.

On constate que pour un pool de stratégies simples il y a une supériorité très nette de la stratégie TFT qui ne gagne pas toujours mais est toujours très bien placée (voir la figure ci-dessous). De façon générale, les simulations montrent qu'il y a une supériorité des stratégies coopératives (“nice”), vite réactives aux trahisons (“retaliatory”), pardonnant rapidement (“forgiving”, non rancunières) et simples (“clear”, sans ruse).

Si l'on introduit des stratégies imparfaites avec des erreurs, il faut des stratégies un peu plus coopératives. C'est le cas du GTFT (G = “generous”) étudié par Nowak et Sigmund. Molander a montré qu'en réponse à d la meilleure réponse est c avec la probabilité

$$\text{Min}\left(1 - \frac{T - R}{R - S}, \frac{R - P}{T - P}\right)$$



Compétition entre stratégies (d'après J-P. Delahaye).
 Population de 1200 agents. 12 stratégies représentées par 100 agents chacune. À la stabilisation, il n'y a plus que des stratégies coopératives et TFT domine. (Majorité-mou = jouer ce qu'a joué la majorité et si 50%-50% jouer c ; Méfiante = TFT avec d initial).

2.3. Les jeux évolutionnistes

La théorie des jeux évolutionnistes consiste à considérer des *populations* polymorphes d'individus utilisant différentes stratégies et à définir les nouvelles générations à partir des scores obtenus dans une confrontation généralisée. Les stratégies avec bons scores augmentent leurs représentants alors que celles avec mauvais scores disparaissent progressivement. C'est une théorie plus réaliste que les théories classiques de la rationalité individuelle. Elle permet de comprendre par quelles *dynamiques* les agents peuvent atteindre des équilibres.

Soient $\{s_i\}$ les stratégies et $\{p_i\}$ leur probabilité (i.e. la proportion de la population les jouant). On peut supposer que la taille N de la population reste constante. Dans le cas où il n'y a que les 2 stratégies c (avec probabilité $= p$) et d (avec probabilité $= 1-p$), on calcule facilement les espérances de gains ou utilités $U_c(p)$ et $U_d(p)$ de chaque stratégie comme fonction du paramètre p . Rappelons que $T = (d, c)$, $S = (c, d)$, $R = (c, c)$, $P = (d, d)$. Si un agent joue c , la probabilité est p qu'il joue avec un agent c et il gagne alors $(c, c) = R$, tandis que la probabilité est $1-p$ qu'il joue avec un agent d et il gagne alors $(c, d) = S$. Si en revanche l'agent joue d , la probabilité est p qu'il joue avec un agent c et il gagne alors $(d, c) = T$, tandis que la probabilité est $1-p$ qu'il joue avec un agent d et il gagne alors $(d, d) = P$. On obtient ainsi :

$$\begin{cases} U_c(p) = pR + (1-p)S \\ U_d(p) = pT + (1-p)P \end{cases}$$

Le gain moyen de la population est par conséquent donné par la formule :

$$U(p) = pU_c(p) + (1-p)U_d(p) = p^2R + p(1-p)S + (1-p)pT + (1-p)^2P$$

soit

$$U(p) = p^2R + p(1-p)(S+T) + (1-p)^2P$$

Quant à l'évolution des probabilités p_i , elle est donnée par des *dynamiques de répliation*

$$p' = p(U_c(p) - U(p))$$

2.4. La stratégie du "tit for tat" : du sens commun aux modèles

Dans ces modèles, les agents sont considérés comme des "phénotypes" exprimant des stratégies "génotypes" et leurs stratégies "micro" influent sur la dynamique "macro" de leur population. Les simulations (synthèse computationnelle) de cette dynamique fournissent des résultats extrêmement intéressants. Pour des stratégies simples comme ci-dessus Axelrod a montré :

- (i) Il y a élimination des stratégies anti-coopératives et la coopération s'installe et se stabilise.
- (ii) C'est la stratégie *TFT* qui domine, mais dans les cas où il peut exister des mutants elle est *fragile* car les mutants "gentils" *Ic* ont le même comportement que *TFT* dans un environnement *TFT* et peuvent donc se substituer progressivement au *TFT* ; mais alors des mutants "méchants" *Id* peuvent facilement les déstabiliser et envahir le système.
- (iii) Pour une stratégie, la réactivité aux trahisons est une condition pour être collectivement stable, c'est-à-dire ne pas pouvoir être déstabilisée par un mutant.
- (iv) Si l'on introduit des stratégies complexes, alors il peut se produire énormément de phénomènes subtils différents dès qu'il existe au moins trois stratégies en interaction. Par exemple une stratégie non coopérative peut en utiliser une autre pour éliminer les stratégies coopératives et l'éliminer ensuite à son tour. Ou encore le désordre permet à des stratégies non coopératives de survivre et même de gagner, etc.
- (v) La maîtrise computationnelle de ces situations permet alors de montrer qu'il existe d'autres stratégies que *TFT* qui sont maximalelement performantes dans ces contextes élargis. Autrement dit, on peut raffiner le *TFT* sélectionné par le sens commun.

Bref on constate "expérimentalement" (au sens de la synthèse computationnelle) une extrême complexité des dynamiques possibles.

Nous rencontrons ici un exemple typique de *modèle du sens commun* :

- (i) Les simulations confirment un sens commun politique, social, pédagogique qui constitue un savoir incorporé et permet d'agir de façon efficace sans avoir à répéter indéfiniment les mêmes expériences décevantes.
- (ii) Mais dans le même temps elles permettent de dépasser le sens commun, non pas en lui donnant tort, mais en sélectionnant des règles en quelque sorte d'"hyper" sens commun dans le cadre *experimental* (au sens de la synthèse computationnelle) d'évolutions culturelles *virtuelles*.

Les jeux évolutionnistes sont intéressants dans la mesure où ils remplacent une intelligence rationnelle individuelle surhumaine (irréalisable) par l'intelligence collective (réalisable) d'une population d'agents en interaction dont la rationalité et les ressources cognitives sont limitées. L'optimisation n'est plus individuelle mais collective et peut être obtenue sans hypothèse de rationalité individuelle forte.

2.5. Les généralisations de Sigmund et Nowak

Un certain nombre d'auteurs ont étudié des facteurs qui favorisent la coopération dans l'IPD lorsque l'on change l'espace des stratégies, les processus d'interaction et les processus d'adaptation (i.e. les changements de stratégie des agents par apprentissage). En particulier, l'introduction de relations "topologiques" de "voisinage" entre agents autorise pour chaque agent un apprentissage imitant le voisin qui a fait le meilleur score. L'évolution des stratégies par algorithmes génétiques a également des effets très importants.

On peut ainsi modéliser de très nombreux systèmes. Considérons par exemple des stratégies simples (i, p, q) où :

i = probabilité initiale de coopération,

p = probabilité de coopération au coup suivant si l'autre coopère,

q = probabilité de coopération au coup suivant si l'autre trahit.

On a trivialement $Ic = (1,1,1)$, $Id = (0,0,0)$, $TFT = (1,1,0)$, $c_p = (p, p, p)$ (toujours c avec la probabilité p). On a aussi $GTFT = (1,1, \text{Min}\left(1 - \frac{T-R}{R-S}, \frac{R-P}{T-P}\right))$.

Sigmund et Nowak ont montré que les Id peuvent gagner au début. Mais les TFT résistent. Une fois que les "gentils" Ic ont été décimés, les exploiters ne peuvent plus les exploiter et les stratégies coopératives de type TFT s'imposent. Après avoir permis l'émergence de la coopération, elles sont elles-mêmes dépassées par des $GTFT$. Mais la stratégie $GTFT$ est fragile et permet le retour des Id . Une stratégie qui résiste bien à Id est la stratégie "pavlovienne" de Kraines : c après R ou T , d après P ou S .

2.6. Les IPD spatialisés de Nowak et May

Pour les IPD *spatiaux*, il existe une "topologie" telle que chaque agent possède certains voisins avec qui il interagit. Pour un voisinage à quatre voisins *renouvelés* à chaque période et pour un processus d'évolution consistant à imiter le voisin ayant fait

le meilleur score, Axelrod obtient les résultats suivants : pour des probabilités (i, p, q) initiales de $(0.5, 0.5, 0.5)$ distribuées aléatoirement, ceux qui coopèrent presque toujours (les “gentils”, les “suckers” S) sont éliminés par ceux qui trahissent presque toujours (les “méchants”, les “meanies” M) dont la stratégie se propage.

Dans un tel système, l'émergence de la coopération est impossible. En revanche si les voisins sont *fixes* (au lieu de changer à chaque période), alors les stratégies défectives ne peuvent plus envahir le système. C'est au contraire la stratégie *TFT* qui domine et se généralise car si deux agents *TFT* apparaissent par mutation et se rencontrent, ils font aussitôt école et leur stratégie se propage jusqu'à envahir le système. Par exemple, dans un tel système un S avec trois voisins *TFT* et un voisin M issu d'une mutation est éliminé par le M qui gagne. Mais ensuite les M doivent interagir entre eux avec uniquement des voisins *TFT*. Comme ce sont alors les *TFT* qui gagnent, les M se convertissent à *TFT*. Autrement dit, les fluctuations M sont *récessives*. C'est la base des propriétés de *stabilité*, dans ce contexte, des stratégies évolutionnairement stables comme *TFT* qui ne peuvent pas être déstabilisées par des envahisseurs mutants.

Nowak et May ont en particulier étudié les systèmes définis sur un réseau carré à 8 voisins par la matrice:

	$A(c)$	$A(d)$
$B(c)$	$R = 1$	$T = b$
	$R = 1$	$S = 0$
$B(d)$	$S = 0$	$P = 0$
	$T = b$	$P = 0$

Comportements :

d = défection (trahir), c = coopération

Gains :

$T = (d, c)$ = Temptation, $S = (c, d)$ =

Sucker, $R = (c, c)$ = Reward, $P =$

(d, d) = Punishment

Conditions

$$T = b > R = 1 > P = 0 = S = 0$$

avec par exemple une configuration aléatoire initiale de 50% de c et de 50% de d . b est le paramètre du système. On compare les scores (la somme des gains du site et de ses 8 voisins) et le site adopte la stratégie du site de son voisinage (lui + ses 8 voisins) qui a obtenu le meilleur résultat. On trouve :

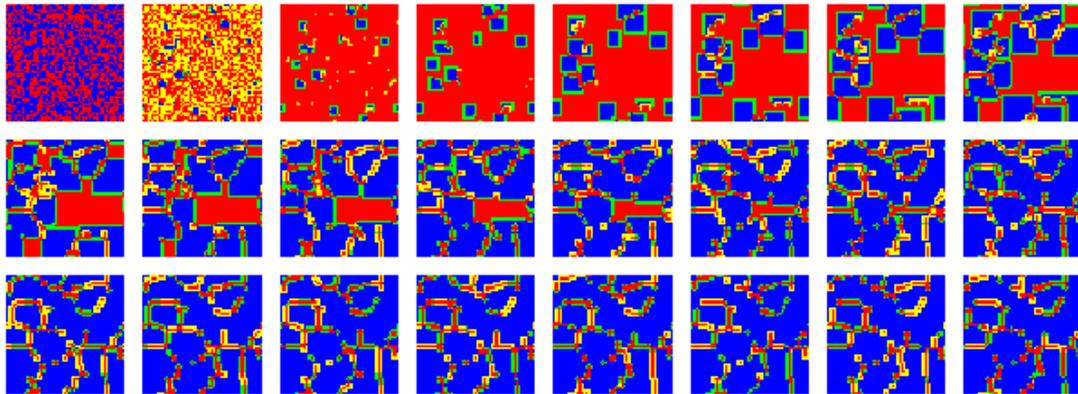
- (i) pour $b < 1.8$, c domine ;
- (ii) pour $b > 2$, d domine ;
- (iii) pour $b \in B_c = [1.8, 2]$ (intervalle critique), Zhen Cao et Rudolph Hwa ont montré qu'il y a une transition *critique* $c \rightarrow d$, avec des clusters multi-échelle emboîtés de c et de d .

Voici un exemple de ce phénomène obtenu avec une implémentation *Mathematica*TM due à Richard Gaylord et Kazume Nishidate. Le code couleur des stratégies est :

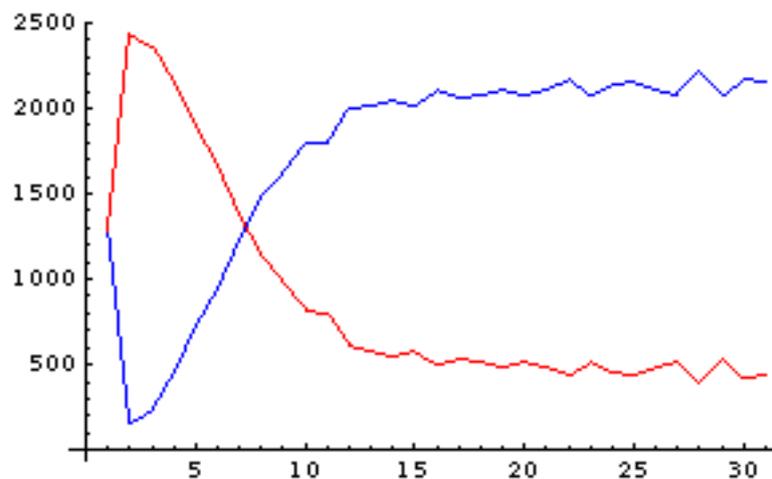
c puis $c = c - c =$ bleu; d puis $d = d - d =$ rouge;

c puis $d = c - d =$ jaune; d puis $c = d - c =$ vert.³

Pour $b = 1.5$ et une configuration initiale “InitConfig” 50%-50%, on voit que $c - c$ domine assez vite, mais uniquement à travers un processus d'extension de noyaux ayant résisté à une phase initiale catastrophique d'élimination. La domination laisse d'ailleurs subsister des lignes de fracture $d - d$ sur lesquelles les comportements oscillent.



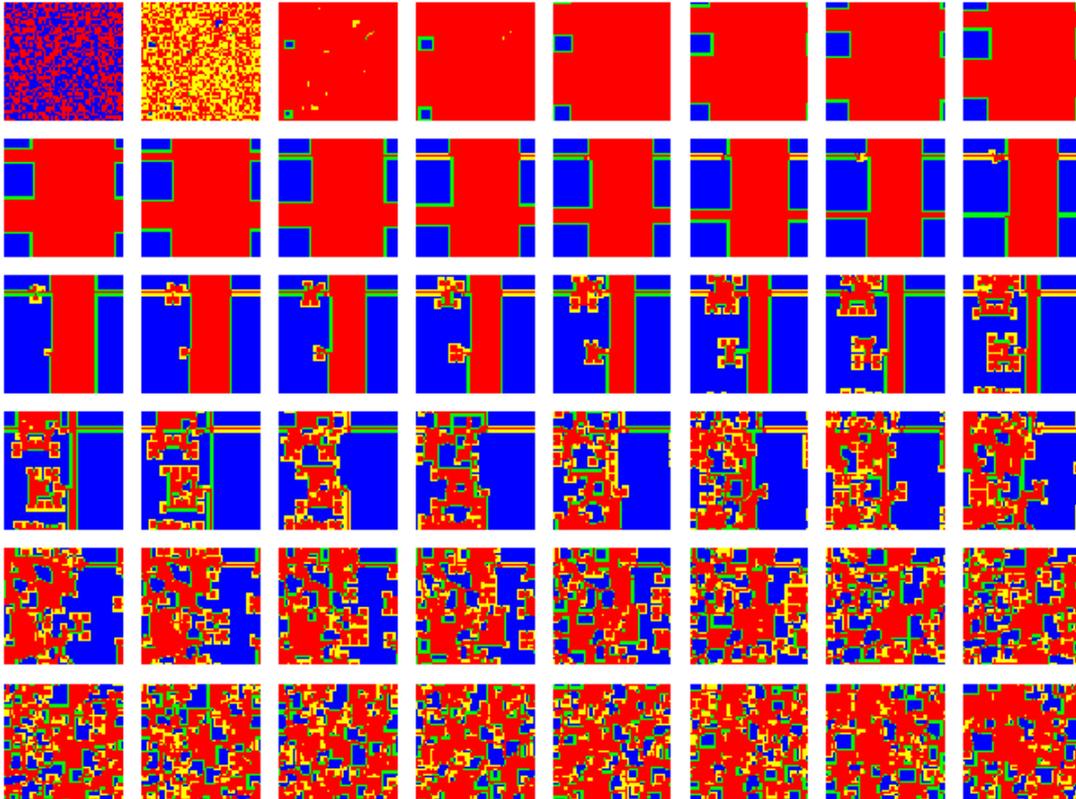
Si l'on représente l'évolution au cours du temps des sous-populations $c - c$ et $d - d$ on voit très nettement ces phénomènes de décimation initiale suivi d'une reconquête et d'oscillation.



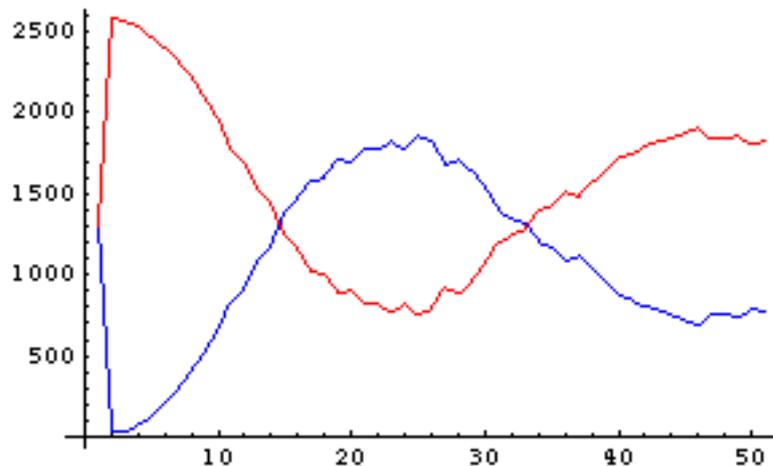
³ Il ne faut pas confondre les notations de type (c, d) disant que deux joueurs différents jouent respectivement c et d et les notations de type $c - d$ disant que le même joueur joue d'abord c puis d .

Pour $b = 2.1$ et une InitConfig 50%-50%, d domine encore plus vite et cette fois directement et sans partage.

Pour une valeur du paramètre $b = 1.85$ appartenant à l'intervalle critique et une InitConfig 50%-50%, on constate que $d - d$ commence par dominer, puis que $c - c$ commence à reconquérir du terrain à partir de noyaux ayant résisté à la décimation initiale, mais que contrairement à ce qui se passait pour l'exemple $b = 1.5$, se constituent ensuite des clusters multi-échelle de c et de d .



Cette dynamique se lit très bien sur les courbes d'évolution où les courbes $c - c$ et $d - d$ subissent, en plus de leurs petites oscillations, des oscillations de grande échelle qui les font se croiser et recroiser.



CONCLUSION

A partir de quelques exemples, nous avons indiqué la façon dont avaient évolué quelques-unes des grandes idées présentées dans cette salle lors de certains des colloques scientifiques de Cerisy. Cerisy ne s'était pas trompé dans ses choix. Ce sont des pans entiers de la recherche contemporaine qui y ont été reconnus non seulement dans leur pertinence scientifique mais aussi dans leur valeur *culturelle*. Et c'est la conclusion que j'aimerais tirer de cette aventure. Lorsqu'elles sont véritablement une *connaissance*, lorsqu'elles innovent de façon suffisamment élaborée théoriquement et surtout *techniquement* — mathématiquement, i.e. au-delà du concept — pour renouveler des problématiques à haute teneur métaphysique, les sciences accomplissent la mission culturelle que leur ont assignée les Lumières. Ce qui était en jeu dans les colloques scientifiques de Cerisy était l'actualité de cette Idée de la Raison.

BIBLIOGRAPHIE

- Amit, D., 1989. *Modeling Brain Function*, Cambridge University Press.
- Atiya, A., Baldi, P., 1989. "Oscillations and Synchronisation in Neural Networks: an Exploration of the Labeling Hypothesis", *International Journal of Neural Systems*, 1, 2 : 103-124.
- Axelrod R., Cohen M., Rislo, R., 1998. "The Emergence of Social Organization in the Prisoner's Dilemma: How Context Preservation and other Factors Promote Cooperation", Santa Fe Institute Working Paper 1999-01-002.
- Berry, M., 1988. "Breaking the paradigms of classical physics from within", *LTC* 1988, 106-117.
- Binmore K., 1994. *Playing Fair*, Cambridge, MIT Press.
- Cao, Z., Hwa, R., 1999. "Phase transition in evolutionary games", *International Journal of Modern Physics A*, 14, 10, 1551-1559.

- Coullet, P., Emilsson, K., 1992. "Strong resonances of spatially distributed oscillators: a laboratory to study patterns and defects", *Physica D*, 61, 119-131.
- Daido, H., 1990. "Intrinsic Fluctuations and a Phase Transition in a Class of Large Populations of Interacting Oscillators", *Journal of Statistical Physics*, 60, 5/6 : 753-800.
- De Kepper, P. *et al.* 1998. "Taches, rayures et labyrinthes", *La Recherche*, 305, 84-87.
- Delahaye J-P., Mathieu P., 1999. "Des surprises dans le monde de la coopération", *Les Mathématiques sociales, Pour la Science*, 58-66.
- Engel, A., König, P., Gray, C., Singer, W., 1992. "Temporal Coding by Coherent Oscillations as a Potential Solution to the Binding Problem: Physiological Evidence", *Non Linear Dynamics and Neural Networks* (H. Schuster ed.), Berlin : Springer.
- Hayek, F. von, 1988. *The Fatal Conceit. The Errors of Socialism*, London-New York, Routledge.
- Hofbauer J., Sigmund K., 1988. *The Theory of Evolution and Dynamical Systems*, Cambridge University Press.
- Kirman A., 1998. "La pensée évolutionniste dans la théorie économique néoclassique", *Philosophiques*, XXV, 2, 219-237.
- Kuramoto, Y., Nishikawa, I., 1987. "Statistical Macrodynamics of Large Dynamical Systems. Case of a Phase Transition in Oscillator Communities", *Journal of Statistical Physics*, 49, 3/4, 569-605.
- LTC, 1988. *Logos et Théorie des Catastrophes*, (J. Petitot ed.), Colloque de Cerisy à partir de l'Œuvre de René Thom, Éditions Patiño, Genève.
- Meinhardt, H., 1995. *The Algorithmic Beauty of Seashells*, Berlin, Springer.
- Mumford D., 1994. "Bayesian rationale for the variational formulation", *Geometry-Driven Diffusion in Computer Vision*, Dordrecht, Kluwer.
- Nadeau R., 1998. "L'évolutionnisme économique de Friedrich Hayek", *Philosophiques*, XXV, 2, 257-279.
- Nemo P., 1988. *La société de droit selon F.A. Hayek*, Paris, Presses Universitaires de France.
- NP, 1999. *Naturalizing Phenomenology: Issues in Contemporary Phenomenology and Cognitive Science*, (J. Petitot, F. J. Varela, J.-M. Roy, B. Pachoud eds.), Stanford University Press.
- Petitot J., 1994. "La sémiophysique : de la physique qualitative aux sciences cognitives", *Passion des Formes*, à René Thom (M. Porte éd.), 499-545, Éditions de l'École Normale Supérieure de Fontenay-Saint Cloud.
- Petitot J., 1995. "Morphodynamics and Attractor Syntax. Dynamical and morphological models for constituency in visual perception and cognitive grammar", *Mind as Motion*, (T. van Gelder, R. Port eds.), 227-281, Cambridge, MIT Press.

- Poundstone W., 1993. *Prisoners Dilemma*, Oxford University Press.
- Renals, S., Rohwer, R., 1990. "A Study of Network Dynamics", *Journal of Statistical Physics*, 58, 5/6, 825-848.
- Samuelson L., 1997. *Evolutionary Games and Equilibrium Selection*, Cambridge, MIT Press.
- Sompolinsky, H., Crisanti, A., Sommers, H.-J., 1988. "Chaos in Random Neural Networks", *Physical Review Letters*, 61, 259-262.
- Thom, R., 1972. *Stabilité Structurelle et Morphogénèse*, New York, Benjamin, Paris, Édiscience.
- Thom, R., 1980. *Modèles mathématiques de la Morphogénèse*, Paris, Christian Bourgeois.
- Thom, R., 1988. *Esquisse d'une Sémiophysique*, Paris, InterÉditions.
- Tirozzi, B., Tsodkys, M., 1991. "Chaos in Highly Diluted Neural Networks", *Europhysics Letters*, 14, 727-732.
- Turing, A., 1952. "The Chemical Basis of Morphogenesis", *Collected Works*, 4, 1-36, North-Holland, 1992.
- Weibull J., 1996. *Evolutionary Game Theory*, Cambridge, MIT Press.
- Zeeman, C., 1977. *Catastrophe Theory. Selected Papers, 1972-1977*, Addison-Wesley, Reading, Mass.