

Au fur* et à la mesure** : à propos de l'évaluation numérisée de la production scientifique

Patrick Vinay, Gabriel Baverel

Tout semblerait plus simple si les commissions scientifiques qui attribuent bourses et subventions étaient remplacées par des ordinateurs. Justifiées par un ensemble de règles incontestables, les décisions tomberaient, irrévocables dans leur froide objectivité, comme le couperet de la guillotine. Personne ne pourrait se plaindre et l'objectivité automatique et absolue ne pourrait souffrir de contestation ! Mais les règles les mieux prévues pour rendre la justice sont fatalement injustes lorsque l'élément humain, original et imprévisible, intervient. A croire que ce dernier a la passion du grain de sable dans l'engrenage, qu'il refuse de se conformer aux généralisations, qu'il s'identifie de préférence à l'individualité, qu'il impose sans cesse le cas particulier. Cela rend nécessaire la poursuite du travail des commissions.

Pourtant, la tentation de simplifier et de « numériser » la décision en matière de jury scientifique apparaît aujourd'hui. Sous la façade de la justice et de l'efficacité, on nous propose une entreprise d'évaluation « objective » de la performance des scientifiques. Cette méthode, d'un intérêt certain lorsqu'elle est prise comme un élément d'évaluation parmi d'autres, expose les chercheurs à de grands dangers si elle

devient un élément dominant dans le processus d'évaluation. Nous ne disposons pas d'un grand recul, mais nous pouvons déjà voir quelques écueils graves menacer l'évolution libre de la créativité scientifique surgissant à l'appel de cette nouvelle pensée normative.

Le jugement par les pairs : entre le meilleur et le pire

Un scientifique produit un travail, et le publie dans un journal. Cette publication est le produit fini de son labeur, la communication à ses pairs, l'apport dans la mêlée de mesures nouvelles, d'hypothèses originales, de quelques idées qui épaulent, raffinent ou révolutionnent le cadre ou le contenu des connaissances acquises. La qualité du travail de ce scientifique sera facilement appréciée de ceux qui œuvrent dans le même domaine que lui, parfois une communauté internationale très restreinte. Mais, tôt ou tard, elle devra être jugée aussi par des pairs qui ne sont pas forcément des experts du domaine examiné : au moment où il publie, lorsqu'il lui faut faire des demandes de crédit ou justifier son incorporation au corps constitué de recherche ou d'enseignement, etc. De ce jugement dépendra la poursuite ou non de ses activités de recherche, le

développement de ses idées, l'offre d'un contrat universitaire... Le jeune scientifique qui désire entreprendre une carrière se verra alors jugé presque uniquement sur ses publications à un moment où des décisions importantes pour sa vie seront prises par les commissions scientifiques désignées.

Il faut dire que la communauté scientifique s'est toujours prêtée à cet exercice avec énormément de générosité. Nombre de scientifiques passent une fraction importante de leur temps à évaluer des demandes de bourses ou de subventions, parfois l'équivalent d'un mois de travail bénévole par année, souvent effectué le soir ou le dimanche ! Au Québec, quelque 1 500 évaluations sont effectuées chaque année pour les seules sciences de la Santé (données du FRSQ). Chaque commission comporte de cinq à vingt membres qui ont tous le statut de « pairs », c'est-à-dire qui sont des chercheurs actifs et subventionnés par des organismes reconnus. Comme aucune commission ne peut comporter des spécialistes dans tous les domaines, on fait également appel, pour chaque dossier, à des experts qui témoi-

* Fur : petit tonneau servant de mesure de volume (grain, pommes, vin).

** Mesure : ruban ou tige graduée servant à mesurer les longueurs (tissus, bois...).

gnent par écrit de leur analyse et ce jugement vient étayer le travail de la commission. Que de personnes mises à contribution, parfois pour des concours dotés de maigres ressources !

Mais l'exercice n'est pas simple : les pairs doivent lire les demandes, comparer les hypothèses avec ce qui est déjà connu, évaluer les risques et les chances de succès de l'entreprise, la profondeur de la compréhension du champ scientifique, l'à-propos des méthodes proposées, etc. La qualité de la production antérieure doit aussi être évaluée et cela ne peut être bien fait qu'en lisant des articles représentatifs du travail du demandeur.

Les grands ennemis ici, sont le jugement qualitatif et rapide, l'impression non fondée, la compétition déloyale ou le conflit d'intérêt. Il faut s'assurer que la plus grande objectivité possible a bien été observée. Malgré des précautions multiples, les commissions se trompent parfois, surtout pour le jugement des entreprises les plus hardies, les plus originales qui bousculent le cadre rassurant des vérités acceptées du moment. L'histoire relate des prix Nobel qui ont vu leur contribution magistrale rejetée ou mal évaluée par un jury qui pourtant faisait de son mieux. Les commissions se trompent surtout dans le jugement des candidats les plus jeunes, c'est-à-dire de ceux qui présentent une base objective minimale au jugement des pairs. Le jugement demeure bien notre plus difficile exercice !

Comme la démocratie, ce système est retenu comme le moins mauvais parce que l'on n'en connaît pas de meilleur ! Bien sûr, ce système inspire d'autant plus confiance qu'il implique le travail bénévole de scientifiques intègres.

Alors vint l'ordinateur

On propose aujourd'hui d'accroître l'objectivité du jugement des pairs en évaluant systématiquement la qualité des publications non sur leur contenu mais, chose étrange, sur leur contenant ! La nouvelle propo-

sition veut utiliser une procédure « objective » pour pondérer la qualité des articles scientifiques produits en fonction des revues qui les ont publiés. Cette nouvelle façon de faire émane de l'ère de l'ordinateur. En effet, il est maintenant possible de compter le nombre de citations atteint par une revue ou un journal dans la littérature scientifique sur une période récente, trois ans par exemple. Cette analyse de fréquence permet de classer les journaux cités et de dégager ceux qui jouent un rôle important de forum scientifique, dans les pages desquels la Science s'écrit et dont les articles font l'opinion. La publication dans ces journaux procure aux auteurs une image de qualité. Ces journaux deviennent donc très sollicités et leurs comités de lecture deviennent, par force, très sévères. Un système de hiérarchie s'est d'ores et déjà mis en place parmi les journaux scientifiques. Ainsi, un article dans *Annual Review of Biochemistry* vaudra 35.5 points*, un autre dans *Journal of Biological Chemistry* 6 points, et dans *La Dépêche Médicale du Sud-Ouest...* seulement 0,012 point, etc. Un système de cote est alors proposé et mis à jour annuellement en fonction du relevé informatique des citations. Tenant compte de ces cotes, on peut « mesurer » la production scientifique des individus : il suffit d'additionner, sur une période de temps donnée, les cotes des articles qu'ils ont publiés en fonction des journaux impliqués. Le plus haut total indique évidemment la meilleure production.

On peut aller plus loin dans la quantification de la production scientifique en tenant compte de la place tenue dans la liste des signataires de chaque article. Une première place est censée refléter la production directe de l'individu (pondération de 0,8 par exemple), une cinquième celle de la dernière roue du carrosse (pondération de 0,1) et la dernière, celle de l'autorité scientifique sous les auspices

duquel le travail a été conduit (pondération de 0,3). On pourrait ainsi pondérer (cote du Journal x facteur de place d'auteur) la valeur de chaque publication dans le dossier d'un chercheur et faciliter plus avant le processus de hiérarchisation de la production individuelle qui est au cœur de l'exercice des concours scientifiques.

Cette nouvelle cote intègre le jugement des comités de lecture des journaux et jouit apparemment (à juste titre ?) de la rigueur et de la crédibilité de ce processus. Mais, en poussant un peu, les commissions n'auraient plus besoin de lire les œuvres de leurs candidats : les comités de lecture des journaux ont déjà fait le travail ! La cote calculée de façon automatique dispenserait les arbitres de se forger leur propre opinion sur la démarche qu'on leur demande d'évaluer ! Il faut craindre une solution de facilité, une dérobade qui court-circuiterait le nécessaire exercice de jugement et qui stériliserait une partie de la créativité scientifique.

Une critique de l'automate

La cote attribuée à chaque journal est essentiellement basée sur le nombre de citations qui réfèrent à ces revues. La communauté scientifique qui cite le plus est la communauté scientifiquement la plus active parce que la plus riche, c'est-à-dire la communauté américaine. Les journaux américains seront donc forcément souvent les plus cités, les mieux cotés, même si plusieurs journaux européens prestigieux de langue anglaise sont bien recensés. Cela donne aux journaux américains une valeur prédominante dans ce système qui reflète autant la culture du moment que la qualité intrinsèque des revues en cause. Le jugement automatique défavorise alors les scientifiques qui n'appartiennent pas à la culture scientifique « dominante » (allemands, chinois, français), et impose certaines revues américaines au détriment de publications locales souvent excellentes (par exemple la revue centenaire

* *Science Citation Index; Journal citation report 1992.*

Hoppe Seyler's d'Allemagne, *médecine/sciences* en France), mais étrangères à la culture scientifique monolingue dominante. Dès lors, les cotes ne font pas justice aux contributions parues dans ces journaux. Certains jurys retirent même parfois des points aux auteurs qui ont publié dans des journaux dits secondaires, défavorisant clairement ceux qui veulent préserver une contribution à leur contexte culturel.

La qualité de la production scientifique devient sournoisement identifiée à la publication dans certaines revues. On refuse ici de voir que la Science progresse aussi de proche en proche, de confirmation en confirmation, au moyen de petites démonstrations qui sont essentielles, même si elles ne sont pas spectaculaires, et qui ne trouvent pas toujours de place dans les « grandes revues ».

Mais il y a plus grave. Cette méthode propose une vue biaisée de la Science. En effet, le nombre de citations, et donc les cotes du système, reflète forcément l'engouement du moment pour un domaine particulier : les domaines originaux qui couvent sous braises ne peuvent être reconnus précocement pour leur originalité ou leur potentiel scientifique par cette méthode. Un jugement qui ne tiendrait pas compte du potentiel de nouveauté d'une démarche, parce qu'elle n'est pas encore largement acceptée, stériliserait l'effort scientifique en le transformant en une longue entreprise de répétitions ou d'exploitations de données déjà connues.

Différents journaux spécialisés et pointus deviennent alors des références incontournables pour certains domaines de la Science. Soit ! Mais les journaux les mieux cotés seront toujours ceux des domaines les plus fréquentés. Un journal pourra alors être de première importance pour une spécialité, et injustement évalué par un système qui reflète l'opinion de la majorité, seulement parce qu'il témoigne d'une activité quantitativement « minoritaire ». A court terme, ce système va écarter les scientifiques qui œuvrent dans des domaines con-

sidérés comme périphériques. C'est nier le fait que la Science est constamment régénérée par des progrès qui viennent justement de la frange de ses activités.

Dénonçons aussi le système de pondération lié au rang du chercheur dans la liste des signataires. Si l'on veut tuer toute entreprise en collaboration, on ne peut mieux s'y prendre ! Allez expliquer à un collègue que ses compétences et son savoir-faire, pourtant indispensables à la réalisation d'une étude en collaboration, ne valent que 0,2 point ! Allez dire à un jeune chercheur imaginaire que son travail est moins bon parce qu'il a inclus le nom d'un collaborateur senior parmi les signataires de son papier ! A l'heure où les équipes pluridisciplinaires deviennent quasi essentielles au progrès de la Science, cela ne peut être envisagé ! Il faudrait, au contraire, tout mettre en œuvre pour favoriser la concertation et pour récompenser les regroupements productifs.

Ce système implique enfin que les contributions de chacun soient correctement citées. En fait, chacun sait qu'il n'en est rien. On sait que les scientifiques incorporent volontiers le travail des autres, dans leurs techniques par exemple, le modifient de façon mineure et citent dorénavant leurs propres publications, frustrant l'auteur original d'un élément de reconnaissance auquel il aurait droit. Par ailleurs, on citera plus fréquemment un article qui établit ou modifie une nouvelle technique de mesure, même s'il ne propose pas de développement conceptuel important : la cote de citation lui donnera une importance exagérée. On citera souvent une technique incorrecte, ou une observation mal interprétée, pour faire contraste... On sait enfin la propension à l'autocitation.

On trouvera parfois des homonymes. Par exemple J.M. Bishop, prix Nobel pour son travail sur les oncogènes, a un homonyme qui publie presque autant que lui. Il en va de même pour P.H. Sharp. Est-ce que l'ordinateur saura faire la différence ? Enfin, certains scientifiques exigent de devenir coauteurs sous

différents prétextes : ils exigent parfois d'être présents dans toutes les publications qui utilisent des molécules qu'ils ont synthétisées ou des cellules qu'ils ont conservées dans l'azote liquide et dont ils n'ont plus besoin pour leurs propres recherches, accaparant ainsi un crédit qu'ils n'ont pas réellement mérité ! A l'inverse, on ne cite plus le bagage scientifique lorsqu'il devient tellement courant qu'il est considéré comme allant de soi. Qu'on pense aux multiples publications qui impliquent le cycle des acides tricarboxyliques sans citer le Pr H.A. Krebs, ou qui utilisent le *Southern blot* sans reconnaître qu'il tire son nom du Dr Southern !

Sans rejeter l'utilisation de ces nouveaux outils que l'informatique met à notre disposition, ni nier l'intérêt des indices qu'ils fournissent, il faut les replacer dans leur contexte réel et ne pas leur attribuer une importance qu'ils n'ont pas. Il faut les critiquer en même temps qu'on les considère ; il faut surtout rejeter la tentation qu'ils véhiculent de l'automatisme dans l'évaluation.

Il faut bien le dire : seule l'opinion des scientifiques mérite notre respect et entraîne une crédibilité suffisante pour justifier notre acceptation du processus souvent difficile du jugement par les pairs. Pour perpétuer ce système, il faudra encore des hommes et des femmes disponibles, capables de considérer tout ce qui ne peut entrer dans l'ordinateur et désireux de prendre des décisions sages. Des personnes qui comprennent, évaluent et pondèrent non comme des machines mais comme des scientifiques responsables ■

Patrick Vinay

Professeur à l'université de Montréal et hôpital Notre-Dame, laboratoire de néphrologie, 1560 Sherbrooke est, Montréal, QC, H2L 4M1, Canada.

Gabriel Baverel

Professeur à la faculté de médecine A. Carrel, 12, Guillaume-Paradin, 69372 Lyon Cedex 08, France.

TIRÉS A PART

P. Vinay.