

La crise du système d'évaluation scientifique

LAURENT NOTTALE

La science française traverse actuellement, et sans conteste, une crise. Mais quelle est la nature de cette crise ? S'agit-il d'un simple soubresaut qui accompagne les changements d'organisation inévitables dans l'évolution des grandes institutions, ou est-elle plus profonde ? "Déclin", "décadence", "régression", voilà les mots employés pour qualifier la situation actuelle en France.

Ce déclin semble réel et profond, mais ceci peut n'être qu'une apparence, liée en fait à un défaut de vision et de définition des priorités des différents acteurs sociaux, alors que toutes les forces nécessaires sont bel et bien présentes, mais mal identifiées. On sait qu'il est souvent arrivé de par le passé que des révolutions scientifiques majeures soient restées cachées pendant des décennies (sinon des millénaires!), du fait de l'inertie et de pesanteurs qui peuvent tourner à l'obscurantisme.

Une autre question essentielle est de savoir si cette crise est proprement spécifique de notre pays. Qu'en est-il ailleurs? Certains commentateurs considèrent que le problème est surtout français alors que d'autres pays, au premier rang desquels les Etats-Unis, garderaient une science florissante. Il me semble au contraire que la crise de la science est en fait mondiale: la science n'est-elle pas après tout l'une des toutes premières entreprises humaines qui se soit mondialisée, et sur laquelle la globalisation actuelle prend appui. La diminution du nombre d'étudiants en science est un phénomène général dans tous les pays industrialisés: nous assistons en fait à une crise commune, qui a pris plus d'ampleur en France du fait de l'inadéquation criante de la réaction de notre pays à ce constat de crise. Quand tous les autres grands pays, faisant le même constat, ont aussitôt conclu qu'il fallait augmenter de

manière significative les crédits et l'ensemble de l'effort de recherche, (ce qui inclut toutes les actions visant à rendre à la science le statut attractif qu'elle avait encore il y a quelques années), au premier rang duquel l'embauche des jeunes chercheurs, notre pays a diminué de manière drastique et les crédits de recherche et l'embauche, un ancien ministre allant même jusqu'à déclarer que "l'on n'avait qu'à s'appuyer sur la science américaine"!

Des Etats Généraux de la Recherche ont maintenant eu lieu, qui ont synthétisé un grand nombre de propositions concrètes de réforme et de réorganisation issues de la mobilisation de la communauté scientifique française. La question est maintenant celle de la mise en oeuvre effective de ces propositions, ce qui n'est pas une mince tâche face à l'inertie et aux blocages multiples du système, précisément dénoncés par toutes les analyses. "Les actes seront-ils cohérents avec les discours", se demande à juste titre le collectif "Sauvons la Recherche" (janvier 2005).

La réflexion doit donc se poursuivre. Je voudrais intervenir sur un point qui mérite, me semble-t-il, une analyse spécifique (qui n'aura rien d'exhaustif, car c'est en soi un sujet bien trop vaste): il s'agit d'un mode d'organisation qui est au coeur du fonctionnement de la science actuelle, à savoir le système d'évaluation des travaux scientifiques par les pairs.

L'évaluation par les pairs

Le système d'évaluation par les pairs est considéré à tous les niveaux, aussi bien avant les propositions de réformes que dans le cadre de celles-

ci, comme la base inaliénable et indiscutable sur laquelle doit s'appuyer toute organisation de la recherche, nationale et internationale. Qu'il soit d'emblée bien clair que j'adhère totalement à ce principe, et que ce qui va être discuté ici ne concerne aucunement le principe lui-même, mais les conditions réelles de son application.

La situation est très semblable à celle de la démocratie dans l'organisation politique: les principes démocratiques et républicains, s'ils sont appliqués, garantissent le droit et la justice, la liberté et l'égalité entre les citoyens. Mais on sait très bien que, dans la réalité, ce qui fait qu'on ne s'écarte pas trop de l'idéal démocratique est une vigilance constante permettant d'identifier les tendances contraires et d'en corriger les effets. Mais il arrive régulièrement que certains Etats glissent vers le totalitarisme et la dictature, piétinant le droit et la justice et faisant fi de la séparation des pouvoirs, tout en continuant de s'arroger en parole tous les attributs apparents de la démocratie.

Un danger semblable existe dans le cas du système d'évaluation par les pairs. Qu'entend-on par "évaluation", comment définit-on un "pair", comment est-il désigné, quels recours existent en cas d'éventuelles injustices ou abus de pouvoir ?

Les dérives du système

Il se trouve que, déjà depuis quelques années, la sonnette d'alarme a été tirée par de nombreux scientifiques sur les dérives graves du système d'évaluation. Le mieux est de laisser la parole à l'un de ceux qui, au coeur même du système, semble particulièrement bien placé pour en parler. En 1995 (il y a donc déjà dix ans!), David Goodstein, professeur de Physique au *California Institute of Technology*, ancien Editeur-en-chef des plus grands journaux scientifiques et membre de nombreux comités d'évaluation, a écrit un article intitulé "L'évaluation par les pairs après le *Big Crunch*"⁽¹⁾, qui délimite le problème d'une manière claire et montre qu'il est en fait devenu général.

Il fait remarquer que depuis trois siècles, la science a connu une expansion à un taux exponentiel effrayant. Le nombre de journaux scientifiques s'est accru d'un facteur dix tous les cinquante ans et atteint actuellement (en 1995)

40000! David Goodstein suggère que le système s'est en fait écroulé dans les années 1970, mais que la plupart des scientifiques, bien que vivant déjà dans l'après "*Big Crunch*", nient encore que le problème existe. La crise se caractérise par la diminution des postes et des crédits, et donc par une compétition intense dans la course aux ressources. Il poursuit:

"Un des signes de détresse les plus essentiels concerne la question de l'honnêteté et du comportement éthique parmi les scientifiques. [...] L'un des piliers cruciaux de tout l'édifice, le jugement par les pairs, est en grave danger. [...] Les rédacteurs en chef de journaux envoient ordinairement les manuscrits qui leurs sont soumis à des "referees" qui restent anonymes pour les auteurs des articles. [...] L'évaluation par les pairs est un moyen tout à fait bon pour identifier la science valide. Elle était merveilleusement bien adaptée à l'ère précédente où le progrès en science n'était limité que par le nombre de bonnes idées disponibles. Le jugement par les pairs n'est plus du tout adapté, cependant, à décider d'une intense compétition pour de rares ressources telle que la recherche de crédits ou de pages dans des journaux prestigieux. La raison en est assez évidente. Le *referee* [...] a un conflit d'intérêt évident. Il lui faudrait un niveau éthique impossible à atteindre pour ne pas utiliser son privilège d'anonymité à son propre avantage. [...] De plus en plus de *referees* voient leur standards éthiques érodés par les avis injustes qu'ils reçoivent quand ils sont auteurs. Ainsi c'est le système dans son ensemble qui est en péril."

En 2005, avec l'effondrement du nombre d'étudiants dans toutes les disciplines scientifiques, phénomène devenu catastrophique en ce qui concerne les sciences fondamentales comme la physique et les mathématiques, et les baisses de crédit massives pour la recherche en France, il ne s'agit plus d'un péril futur, mais bel et bien d'une crise présente.

Tout est dit dans ce texte sur la spirale infernale provoquée par le *publish or perish*. Dès qu'un auteur propose une idée nouvelle, originale ou simplement sortant de l'ordinaire, il se trouve souvent confronté à des problèmes et des blocages sans fin.

Dans les faits, le système d'évaluation par les pairs est devenu, dans certaines situations,

profondément injuste. L'expert choisi par les éditeurs d'une revue à laquelle est soumis un article scientifique est appelé *referee* en anglais. A l'origine, ce mot signifie arbitre ou rapporteur, mais aussi juge, sens qui ne devrait pas avoir cours ici. Or de plus en plus de *referees* s'arrogent le droit de juger du travail de leurs collègues, en terme de critique destructive et de jugement de valeur. Il y a ainsi un glissement de plus en plus systématique de la fonction de conseil et de rapporteur telle qu'elle était initialement définie, vers une véritable fonction inquisitoriale qui s'accompagne du pouvoir de vie et de mort sur le travail des chercheurs, et donc sur la vie des idées elles-mêmes, qui n'ont d'existence sociale que sous forme de publication et de diffusion pour le plus grand nombre.

Les *referees*

Le problème est l'absence d'instance d'arbitrage indépendante. Il n'y a pas vraiment d'organisation d'une justice au sens démocratique du terme, qui soit fondée sur la séparation des pouvoirs. Les *referees* étant très souvent à la fois juges et parties, le système ne saurait mener qu'à de nombreux abus de pouvoir. Les exemples abondent où le concurrent direct, qui est sur une position alternative contradictoire avec celle de l'auteur, est choisi comme *referee* (pas forcément par volonté explicite des éditeurs, mais plutôt par ignorance et par dysfonctionnement du système lui-même) et s'empresse de refuser l'article.

L'idée d'évaluation par les pairs suppose également qu'une telle évaluation objective soit possible. Or la réalité est, là aussi, bien loin de cet idéal. Pour éviter les abus de pouvoir et les jugements injustifiés, les éditeurs ont multiplié le nombre de *referees*, beaucoup de journaux demandant deux rapports indépendants sinon plus. Qu'arrive t'il dans ce cas? Autant d'avis différents que de *referees*, tous contradictoires! Combien d'auteurs n'ont-ils pas reçu des réponses aliénantes telles que: "*Referee* 1: l'article n'est acceptable que s'il est réduit d'un facteur deux; *Referee* 2: l'article doit être publié exactement comme il est; *Referee* 3: l'article ne peut être publié que s'il est fortement augmenté..." Un autre exemple typique est celui du *referee* qui demande un changement important auquel se

conforme l'auteur; l'éditeur envoie alors l'article révisé à un autre *referee* qui le refuse en prenant précisément pour argument la partie rajoutée. De tels exemples d'incohérence pourraient être multipliés à l'infini: l'évaluation de la science n'est pas une science exacte!

Mais le plus grave est le nombre très important de cas où les "décisions" des *referees* (qui ne devraient qu'avoir valeur d'avis, mais qui dans la pratique se ramènent à des décisions, les éditeurs ne déjugant que très rarement ceux qu'ils ont choisi comme experts) ne sont motivées par aucun argument scientifique, mais reposent uniquement sur des jugements de valeur. Ce sont alors les opinions, les croyances et les choix purement personnels de "l'expert" qui sont pris pour des vérités scientifiques.

Le dysfonctionnement est tel qu'il arrive que la décision de refus soit prise sans autre argument que des déclarations insultantes pour l'auteur, que le refus de publication soit notifié à l'auteur sans que les rapports des *referees* ne lui soit transmis, ou enfin dans les cas les plus extrêmes que l'article soit refusé par l'éditeur sans même avoir été transmis à des *referees*, ce qui constitue finalement un acte de censure délibéré (je ne parle pas ici des quelques revues internationales à très grande visibilité soumises à un flux si intense qu'elles se sont vues contraintes de renvoyer 90 % des articles qui leur sont soumis sans les prendre en considération).

Pourquoi en est-on arrivé là ?

Pourquoi en est-on arrivé là ? Comme le fait remarquer David Goodstein, une des raisons principales en est la croissance démesurée du nombre d'articles soumis aux diverses revues scientifiques. Ces revues ne cessent de croître en taille, se découpent en sous-sections qui elles-mêmes se développent sans fin. Le nombre de ces revues approche les 50000, et le nombre d'articles publiés, toutes sciences confondues, est de plusieurs millions par an. Les comités éditoriaux tentent de faire barrage à cette inflation galopante par tous les moyens, et en particulier par le système des *referees*. Le résultat finit malheureusement par être exactement contraire au but souhaité: les articles standard, clairement reconnaissables car sans idée

nouvelle particulière mais au contraire en accord avec le cadre généralement admis, seront acceptés sans problème par les experts et publiés. Les authentiques découvertes scientifiques, impliquant des concepts, outils, ou instruments totalement nouveaux donc non identifiables à partir du bien connu (et en particulier certainement pas par les experts de la science en question, qui sont par définition des spécialistes des méthodes acceptées), sont amenées à être systématiquement refusées et extrêmement difficiles à publier dans des revues à forte visibilité, ce qui empêche leur large diffusion.

Soumises à cette pression du nombre d'articles à gérer, les revues scientifiques ont été amenées à automatiser les tâches. Il est devenu impossible aux éditeurs de maîtriser l'ensemble des domaines couverts par une revue ni de lire eux mêmes en détail l'ensemble des articles qui leurs sont soumis (à part quand elles sont extrêmement spécialisées), ce qui est sans doute une des raisons pour lesquelles ils se reposent sur les *referees*. Ceux-ci sont en fait passés du statut de conseil à celui de décideurs: ainsi la plupart des revues demandent explicitement aux referees, non pas seulement de s'en tenir à des remarques sur le contenu scientifique, mais d'indiquer si l'article doit être accepté ou refusé. C'est ainsi qu'un pouvoir absolu sur le travail des chercheurs est donné à l'un de leurs collègues, souvent concurrent, dont les critères de choix et l'identité sont la plupart du temps couverts par l'anonymat (sans compter les cas où l'auteur refusé a la surprise de découvrir ses idées dans une autre publication rapidement acceptée quelques mois plus tard...).

On pourrait penser que ceci n'est qu'un problème secondaire interne au monde des scientifiques et qu'on est bien loin de la crise de la science. On touche là au contraire à l'une des principales clés de cette crise: ce n'est pas pour rien que Goodstein parle de "*Big Crunch*". En effet, les publications sont autant l'aboutissement que la racine du travail du chercheur. C'est sur elles que vont ensuite se fonder les agences de moyens et de financement de la recherche, ainsi que les comités d'attribution de bourses et de postes.

Or ce deuxième niveau d'évaluation entérine sans aucune critique ni analyse les décisions déjà prises par les revues scientifiques (et donc en fin de compte

par quelques experts eux-mêmes juges et parties, choisis par ces revues), qui se trouvent donc investies du vrai pouvoir de décision. En effet, dans la plupart des cas, c'est le nombre d'articles publiés dans les revues scientifiques prestigieuses, l'importance des revues, la visibilité du travail qui servira d'argument, et non pas la nature effective de ce travail. Des résultats scientifiques majeurs non publiés, ou publiés dans des revues mineures (ou considérées comme telles), seront ainsi considérés comme nuls et non avenus. Le Comité d'éthique pour les sciences du CNRS (COMETS) s'en est ému, écrivant ⁽²⁾: "Ces dangers de dérive sont renforcés par l'utilisation des publications comme critère essentiel de l'évaluation individuelle et par la pratique qui tend à privilégier la quantité par rapport au contenu scientifique."

Un frein au remises en question

Le COMETS a reconnu l'existence de telles dérives du système d'évaluation et tiré la sonnette d'alarme. Il remarque que "les chercheurs sont alternativement juge et partie (évalués de façon fréquente, mais aussi souvent évaluateurs) et s'inscrivent dans des réseaux relationnels dont on ne saurait sous-estimer l'importance ⁽³⁾". "Les découvertes les plus importantes sont souvent des "surprises" qui conduisent à des révisions (quelquefois déchirantes) de certitudes antérieures, bousculant les processus de recherche déjà lancés et éventuellement les hiérarchies établies. Garantir que ces phénomènes rares, mais si importants pour le renouveau de la recherche, puissent se produire est l'un des aspects auxquels il convient de porter une attention particulière dans l'organisation de l'évaluation". "Ces dysfonctionnements et dérives sont liés en particulier à un ensemble de pratiques qui viennent fausser ou infléchir le jeu, normal et sain, de la concurrence entre chercheurs et entre équipes. [] C'est ainsi que l'évaluation, souvent à l'insu de ceux qui la font, tend progressivement: à faire une place plus ou moins large aux phénomènes de réputation; [] à se servir d'experts qui [] soit couvrent rarement la totalité du champ disciplinaire concerné soit s'inscrivent dans des réseaux de connaissances et d'affinités qui ne sont pas seulement scientifiques." Il rappelle enfin que "les

conséquences de l'évaluation pour les chercheurs ou les équipes dont les dossiers n'auraient pas été traités de manière satisfaisante sont considérables et quelquefois dramatiques."

Le résultat d'une telle situation est évident: les scientifiques, dans leur très grande majorité, sont amenés à ne travailler que sur des sujets publiables et directement reconnaissables par leur communauté, sans réel enjeu fondamental, car c'est à une menace mortelle sur la possibilité même de poursuivre leur recherche (moyens et/ou salaires) qu'ils sont soumis par la dérive de ce système. Les réflexions profondes et à long terme, les recherches sur des thèmes vraiment fondamentaux, les remises en question sans garantie de résultat à court ou moyen terme, les innovations imprédictibles sont exclues dès l'origine dans ce type d'organisation. Autrement dit, c'est le fondement même de la création scientifique qui est miné: "par création, on entend ici le fait de développer de manière originale des idées, des méthodes ou des dispositifs, d'effectuer de nouvelles observations, de produire des résultats nouveaux ou inattendus, ou bien de contribuer à définir ou réorganiser un champ de recherche" (4).

Bien que le problème soit général à l'ensemble de la communauté scientifique mondiale, il n'est pas impossible que la science française en ait plus particulièrement souffert. L'esprit critique cartésien, qualité nécessaire et particulièrement bien développée par les études scientifiques en France, peut se retourner et devenir un défaut s'il se transforme en critique destructive. C'est ainsi qu'on a pu voir dans plusieurs domaines, ces trente dernières années, un nombre très important d'idées, de projets, de propositions et de résultats préliminaires étouffés dans l'oeuf en France par les refus des comités de programme et d'attribution de crédits, qui, cinq ou dix ans plus tard, sont devenus les découvertes scientifiques majeures... américaines de ces domaines de recherche.

Que faire ?

Que faire? Une réforme du système des publications scientifiques n'est pas une mince affaire. Elle touche et implique l'ensemble de la communauté mondiale, qui est elle-même en évolution constante et très rapide dans ce domaine

(je n'ai pas évoqué les changements essentiels qui mériteraient une analyse spécifique, comme celui des publications et des archives électroniques qui révolutionnent actuellement les modes de publication et de diffusion de l'information scientifique). Je me contenterai ici de faire quelques propositions qui pourraient sembler évidentes, mais qui ne le sont plus étant donné la dérive du système.

Le premier point est que ce n'est pas aux *referees* de décider de l'acceptation ou du refus des articles, ce qui mène à des abus de pouvoir caractérisés de plus en plus nombreux. C'est aux éditeurs de prendre leurs responsabilités, sur la base des remarques des *referees*.

Le deuxième point est que le travail des *referees* (par ailleurs très long, lourd, difficile et sans contrepartie) soit accompagné de règles éthiques claires qui doivent leur être transmises (ce qui est le cas pour certains journaux).

Mais surtout le troisième point essentiel est qu'il y ait un contrôle par les comités éditoriaux du suivi de ces règles. Un rapport de *referee* ne devrait ni être pris en compte, ni être transmis à l'auteur, ni servir d'aucune façon à la décision de publication ou de non-publication dès lors que ces règles seraient violées, et ceci quelle que soit la personne ayant été choisie comme *referee* (l'argument d'autorité est souvent celui qui prévaut implicitement). Le grand nombre de témoignages chez les chercheurs quant à la réception d'avis clairement illégitimes et de décisions néanmoins prises sur la base de ces avis montre qu'un tel contrôle est actuellement pauvre ou purement et simplement inexistant. Dans l'idéal, il devrait être exercé par d'autres instances indépendantes qui restent à définir, car il y a là encore un problème de séparation des pouvoirs: il n'existe pas jusqu'à présent de cours de justice dans le domaine scientifique. Le COMETS déclare ainsi : "il est souhaitable que les instances à l'origine de l'évaluation ne se contentent pas d'appels à la responsabilité de chaque évaluateur, mais rendent fiables les conditions dans lesquelles s'exerce cette responsabilité. Pour cela, il est nécessaire de mettre en place des dispositifs de détection d'éventuelle dérives []. Une évaluation ne doit jamais être demandée à une personne pour qui l'élimination ou la sélection d'un candidat offre un intérêt direct. Pour la même raison elle doit être refusée par l'évaluateur

qui serait dans la même situation (5)."

Certaines des règles d'éthique et certains devoirs fondamentaux auxquels toute évaluation devrait s'astreindre ont été rappelés par le COMETS, en particulier:

- "L'interdiction d'exploiter personnellement et de transmettre à d'autres les informations qui sont communiquées pour l'évaluation."

- "Présenter explicitement les critères et les raisons selon lesquels les candidats ont été éliminés, recrutés ou promus, les articles ou les moyens refusés."

- "L'évaluateur doit justifier ses conclusions []. Il doit les motiver de façon précise et argumentée."

Mais dans le cas de l'évaluation proprement scientifique dans le cadre de la publication des travaux de recherche, qui est discutée ici, cela n'est pas suffisant: certains évaluateurs donnent leurs critères, justifient longuement leurs conclusions et les motivent de façon argumentée,... sans qu'aucun des arguments utilisés n'aient quoi que ce soit de scientifique.

Si "la validité de ces développements doit pouvoir être évaluée par la vérification de prédictions ou d'hypothèses, par l'expérimentation ou par des constructions logiques (6)", inversement le rejet par un évaluateur ne devrait pas reposer sur ses propres choix, préférences et orientation, mais uniquement sur la détection explicite et argumentée d'une contradiction patente avec une expérience ou d'une évidente erreur interne.

Un renouveau de la science n'est possible que si les idées nouvelles, même choquantes en première impression ou semblant *a priori* incroyables, sont de nouveau les bienvenues et sont publiées dès lors qu'elles ne sont invalidées par aucun argument scientifique décisif et explicite. Peut-être quelques articles reconnus ensuite comme absurdes seront-ils du coup publiés, mais ils se perdront sans problème dans l'océan des cinq millions d'articles qui paraissent tous les ans. En revanche, les scientifiques pourront peut-être ainsi se réinvestir dans la recherche originale, fondamentale et à long terme, si bien que le bébé (des découvertes créatives) ne sera plus jeté avec l'eau du bain, condition sine qua non d'une reprise de croissance... (7)

(1) David Goodstein, *American Scientist*, vol. 83, p.401, septembre-octobre 1995.

(2) "Recherche et création: aspects éthiques", *Les cahiers du Comité d'éthique pour les sciences*, n° 4, septembre 2001, p.1.

(3) "Ethique et évaluation", *Les cahiers du Comité d'éthique pour les sciences*, n° 4, septembre 2001, p. 11.

(4) "Recherche et création: aspects éthiques", *op.cit.*

(5) "Ethique et évaluation", *op.cit.*

(6) "Recherche et création: aspects éthiques", *op.cit.*

(7) Je remercie Charles Alunni pour ses utiles remarques sur ce texte.