

University of Aix-Marseille

From the Selected Works of Sacha Raoult

Summer 2014

L'évaluation du risque de récidive : l'expert, le politique et la production du chiffre

Sacha Raoult



Available at: <https://works.bepress.com/sacharaoult/32/>

DOCTRINE

487

Violations graves des droits de l'homme
et crimes internationaux.

*Du jus (non) puniendi de l'État nation à un
deber puniendi impératif tiré du jus cogens*

John A. E. Vervaele

RSC

Revue de science criminelle
et de droit pénal comparé

CHRONIQUES

612

Cyberdélinquance et provocations policières

Jacques Francillon

611

Les peines extensibles de la loi du 15 août 2014

Pierrette Poncela

655

L'évaluation du risque de récidive : l'expert,
le politique et la production du « chiffre »

Sacha Raoult

669

L'encadrement du Parlement dans
la fabrication de la loi pénale

Christine Lazerges

ref : 551403



DALLOZ

Chronique de criminologie

Sacha Raoult

Maître de conférences
à Aix-Marseille Université,
Andrew W. Mellon Fellow,
University of Chicago

L'évaluation du risque de récidive : l'expert, le politique et la production du « chiffre »

Le rôle de l'expertise scientifique dans la mise en place d'une « nouvelle pénologie » appuyée sur la gestion du risque et des flux, une nouvelle pénologie en concurrence avec le vieil idéal de réinsertion, a été étudié dans différents contextes depuis la mise en lumière du phénomène par Feeley et Simon en 1992 ¹. En France, la nouvelle place donnée à l'expertise a même connu un certain nombre de dérives qui ont fait l'objet d'études récentes, avec l'apparition de « criminologues de plateau de télévision », de « marchands de peur » ² et d'associations de défense des victimes qui emploient de pseudo « experts criminologues » prétendument neutres ³.

Ma connaissance du domaine de l'expertise est différente de cette perspective habituelle, car je m'intéresse non pas à ceux qui communiquent directement avec le politique mais à la production de la littérature empirique anglophone dans les revues à comité de lecture. Cette production scientifique est un des éléments essentiels dans « l'envers du décor »,

c'est le lieu de fabrication des « chiffres », des « opinions » des « évaluations » qui sont ensuite réutilisés à divers niveaux du débat public. Nous avons établi avec Bryce Peppers et les autres membres du projet *Social Position / Knowledge Production* une banque de données regroupant l'entièreté des publications sur la récidive en langue anglaise et nous y étudions la diversité des opinions dans le monde académique. L'objectif du présent article est d'élaborer à partir de ce travail en cours trois remarques concernant cette production qui détermine en partie les possibilités et les limites du champ expertal actuel (pour la question de l'évaluation du risque de récidive). J'exposerai ces remarques en substance en introduction avant de les détailler dans le reste de l'article.

La première remarque concerne la tendance de certains spécialistes à minimiser les difficultés fondamentales liées à la recherche empirique dans le domaine de la récidive. C'est-à-dire qu'il existe, pour caricaturer, dans la littérature que

655

- (1) M. M. Feeley, and J. Simon, *New Penology: Notes on the Emerging Strategy of Corrections and Its Implications*, *Criminology* 30.4 (1992) : 449-474.
- (2) M. Rigouste, *Les marchands de peur: la bande à Bauer et l'idéologie sécuritaire*. Libertia, 2011.
- (3) L. Mucchielli, *Criminologie et lobby sécuritaire. Une controverse française*, La Dispute, 2014.

l'on a examiné deux groupes de chercheurs. Le premier groupe produit une recherche sur des questions fondamentales : peut-on prédire la récidive aujourd'hui et si non le pourra-t-on un jour ? Que peut-on espérer d'une telle prédiction ? Quelle sont les conséquences souhaitables et néfastes de tels mécanismes dans notre système pénal ? Cette recherche tend, depuis très longtemps, à aboutir à une conclusion récurrente : la prédiction de la récidive n'est ni politiquement utile ni scientifiquement faisable. Trop d'obstacles, trop d'inconnues, trop peu de distance, trop de politisation s'emparent de la question « récidive » pour qu'on puisse en tirer un « outil », littéralement quelque chose d'utilisable, sans trahir dans le même temps le souci de rigueur scientifique.

Le second groupe de chercheurs fait, il me semble, les choses dans l'autre sens. D'abord, ce groupe produit une recherche empirique sur l'élaboration, la classification et l'efficacité des outils de prédiction. C'est une recherche lourde, qui nécessite généralement une entrée avec l'administration pénitentiaire ou des cliniques psychiatriques pour les besoins de la recherche de terrain (accès aux dossiers, entretiens avec les personnels et éventuellement les personnes concernées). Quand ces chercheurs sont par la suite conduits à justifier l'utilité politique et la faisabilité technique de leur travail, ils évacuent généralement ces questions le plus rapidement possible. Dans un premier temps, on détaillera cette tendance de la recherche empirique sur l'évaluation du risque de récidive à minimiser les difficultés fondamentales soulevées par la recherche critique (Section 1).

Cette « division du travail » entre la recherche critique et la recherche empirique pose une seconde difficulté. Dans certains domaines sur lesquels j'ai travaillé, les chercheurs les plus critiques participent également à la recherche empirique, ce qui conduit cette recherche empirique à rendre compte de l'entière diversité des opinions sur la question.

Pour diverses raisons, ce n'est pas le cas en matière de prédiction de la récidive, où la majorité de la recherche empirique est conduite par des personnes qui sont *a priori* convaincues que ce type de démarche fait sens. À mon sens, cela conduit le pouvoir de prédiction des outils actuariels actuels à être surestimé dans la littérature scientifique, car un nombre important de ceux qui travaillent sur cette question ont tendance à voir le verre à moitié plein (Section 2).

Enfin, j'aimerais montrer qu'une partie de la tâche confiée aux scientifiques dans le contexte actuel revient au final à déléguer à des techniciens des questions éthiques dont ne nous ne voulons (et ne pouvons ?) pas assumer la responsabilité. Très concrètement, au cœur de la question de la prédiction de la récidive se trouve la question du risque et du sacrifice : combien somme-nous prêts à sacrifier de personnes qui ne récidiveront pas pour repérer une personne qui récidivera. Nous n'avons pas de questionnement de ce type dans le débat public, mais les scientifiques sont contraints d'employer quelque chose qui ressemble à un *ratio* de ce type pour déterminer quel est le meilleur outil à leur disposition. En fait, une majeure partie de leur recherche consiste à comparer de tels *ratios*. Par un étrange hasard de l'histoire des sciences, le *ratio* qu'ils emploient est plus ou moins déterminé en amont par une technique d'évaluation des signaux de radar et conduit *in fine* des algorithmes à décider à notre place de cette douloureuse question (Section 3).

I - La minimisation des difficultés fondamentales

En un siècle, la prédiction de la récidive a régulièrement purgé son vocabulaire en transformant les mots associés à ses concepts opératoires. « Dangerosité » est devenue « risque de récidive », puis « risque de violence ». « Diagnostic » est

devenu « prédiction », puis « évaluation ». L'observateur attentif aura vu poindre depuis quelques années le nouveau *challenger* dans la compétition du *criminologiquement correct*, et il s'est vu confié un observatoire à son nom : la « désistance », c'est-à-dire, en fait, la non-dangereuse, le non-risque.

Cette fuite en avant a joué, j'ai pu l'observer auprès des personnels pénitentiaires dans mon expérience personnelle, la même fonction que celle attendue de tous les changements forcés de vocabulaire : se débarrasser des connotations négatives attachées aux vieux mots. Pour la récidive, la purge du vocabulaire a cependant ses limites. En pratique, il ne faut pas longtemps pour que les connotations négatives reviennent. Des critiques fondamentales, qui en somme sont toujours de même nature, resurgissent régulièrement et le *diagnostic/prédiction/évaluation* de la *dangereuse/récidive/risque* redevient à nouveau une vilaine chose : et pour cause, il n'est pas certain qu'on ait tant progressé que cela sur la question. On toujours peut penser se débarrasser de Michel Foucault en abandonnant la maison, mais il va bien falloir répondre à Kelly Hannah-Moffat.

Ces critiques récurrentes sont de deux ordres et elles se sont remarquablement affinées ces dernières années.

1. Première critique. Lorsque la conférence de consensus sur la récidive s'est tenue le 14 février 2013, Pierrette Poncela et Bernard Harcourt ont organisé une contre-conférence « de dissensus » afin de s'interroger sur ce processus « de production de vérité »⁴. Dans leurs

introductions respectives, ces deux chercheurs s'interrogent sur l'intérêt de prendre la « récidive » comme point de départ d'une discussion de la réforme pénale. En effet, il ne va pas de soi que le fait que le système pénal retombe régulièrement sur ses habitués soit le problème prioritaire, voir un problème tout court, dans notre société actuelle. La montée incessante de l'incarcération, la primauté des indicateurs socio-économiques et du casier judiciaire sur la détermination de la peine au dépend de la gravité de l'acte sont tout autant de points de départ peut-être plus valide pour penser une éventuelle réforme. On identifie là un premier type de problème, celui de *l'utilité politique d'une recherche axée sur la prédiction de la récidive*. C'est un problème scientifique, dans le sens où l'érudition, la connaissance du terrain et l'articulation logique aide à poser de meilleures questions, et notamment à prendre de la distance avec les impératifs du quotidien politico-médiatique. Aux États-Unis, où la croissance carcérale a été plus démesurée que la nôtre, la prestigieuse *National Academy of Science* a commandé pour 2014 un rapport sur les causes et les conséquences de cette sur-incarcération⁵. Chez nous, de plus en plus de chercheurs et d'associations se posent la question de la façon dont nous rendons la justice d'urgence⁶. Il s'agit là de points de départ de réformes à mon sens plus prometteurs que la récidive. Le problème de la récidive, c'est qu'on ne parviendra jamais à un système satisfaisant pour ceux qui voient un échec dans chaque fait divers atroce. Cette fameuse « obsession pour la récidive » dont parlait Bernard Schnapper⁷ n'est pas nouvelle, elle a

(4) La conférence est disponible en audio sur le site internet des Carceral Notebooks, Vol. 8, 2012 : <http://www.the-carceral.org/journal-vol8.html>.

(5) W. Travis et Reburn (dir.) *Growth of Incarceration in the United States: Exploring Causes and Consequences*, National Academies Press, 2014.

(6) A. Christin, *Comparutions immédiates: enquête sur une pratique judiciaire*. Découverte, 2008; Daniel Welzer-Lang, Patrick Castex (dir.), *Comparutions immédiates: quelle justice? Regards citoyens sur une justice du quotidien*, Toulouse, Erès, coll. Société, 2012, L. Mucchielli et E. Raquet, *Les comparutions immédiates au TGI de Nice, ou la prison comme unique réponse à une délinquance de misère*, RSC 2014, 207.

(7) B. Schnapper, *La récidive, une obsession créatrice au XIX^e siècle*, in *Le récidivisme. XX^e congrès de l'Association française de criminologie* (1983) : 25-64.

plus d'un siècle et, à moins d'envisager la réclusion perpétuelle pour tous les condamnés, penser le pénal en terme de récidive est un puit sans fond d'insatisfactions et de frustration.

2. Deuxième critique. Quelques années avant la conférence de consensus et la contre-conférence de dissensus, Philippe Robert et Renée Zauberman, qui sont sans doute parmi ceux qui maîtrisent en France le mieux la question des « chiffres » du crime, faisaient un point sur le malaise que représente depuis longtemps la question de la récidive chez les statisticiens et les sociologues de l'institution pénale⁸. Ce malaise a une raison simple : « lutte contre la récidive » appuyée sur l'état actuel des savoirs n'aurait guère de sens, au vu du décalage entre ce que apporte concrètement l'étude savante et les préoccupations politiques. Ce second problème, c'est celui de la *faisabilité d'une recherche axée sur la prédiction de la récidive*.

Les arguments contre la faisabilité d'une telle recherche sont nombreux, et il est nécessaire au moins d'en rappeler ici les principaux. D'abord, la qualité des données pour travailler sur la « récidive » est un problème en soi qui n'est pas vraiment dépassé. Le mot « récidive » dans la recherche ne veut jamais vraiment dire « récidive » au sens commun (ni récidive en Droit d'ailleurs), mais souvent ré-arrestation ou re-condamnation. Selon le contexte et les études, le « récidiviste » peut-être un libéré conditionnel violant les conditions de son suivi, un « pactolien » rattrapé par son passé⁹. Il est sans doute choquant pour le profane d'apprendre qu'une expression terre-à-terre comme « retour en prison » ne veut pas dans un contexte scientifique dire « retour en prison »

mais généralement « nouvelle condamnation à une peine de prison ferme », qu'elle soit ou non mise à exécution. On peut imaginer mille configurations, milles paramétrages différents pour définir le « récidiviste », et d'une étude à l'autre, le même personnage l'est ou ne l'est pas selon la définition *ad hoc*. Tout cela est sans rapport avec la question qui suscite l'engouement des médias et donc la demande politique. Dans certaines études empiriques est nommé « récidiviste » le détenu pour homicide, libéré et condamné pour possession de stupéfiants ou pour outrage, mais est-ce ce qu'on entend par « récidive » dans la presse ? Est-ce là la récidive qui intéresse le politique ?

Il y a au final un écart important entre la *demande* (repérer ceux qui, parmi les auteurs de crimes relativement graves, sont susceptibles de commettre un second crime relativement grave) et *l'offre*, notamment car schéma-type qui intéresse tellement les politiques est, dans les faits, assez rare. Cela nous entraîne sur le deuxième point : les modèles prédictifs de la récidive masquent une réalité gênante, à savoir le caractère faible et disparate de la récidive pour crime sexuel ou violent. Selon le contexte et les études, ce taux a récemment varié entre 0 % et 33 % dans les dix ans, ce qui est bien en deçà de ce que croient en général les profanes¹⁰.

Troisième critique, l'attribution des différences interindividuelles en matière de « récidive » à des caractéristiques statiques et dynamiques des auteurs tels que leur célibat, leur âge, leurs antécédents, part du principe que ces différences de retour en prison (par ex.) ne traduisent que des différences de « criminalité » et non des différences de

(8) J.-P. Alline et M. Soula (dir.), *Les Récidivistes. Représentations et traitements de la récidive xix^e-xx^e siècle*, Rennes, Presses universitaires de Rennes, 2010.

(9) P. Colin, *La multirécidive pénitentiaire : analyse sociologique des contextes de la multirécidive pénitentiaire chez des hommes condamnés à de courtes peines pour atteintes aux biens*, Thèse, Strasbourg 2, 1998.

(10) L. R. Helmus et al., Absolute Recidivism Rates Predicted By Static-99R and Static-2002R Sex Offender Risk Assessment Tools Vary Across Samples A Meta-Analysis, *Criminal Justice and Behavior* 39, n° 9, 2012, p. 1148-1171.

« pénalité ». Autrement dit, on imagine que ces caractéristiques auxquelles on attribue différents taux de « récidive » ont *tout à voir* avec le fait de commettre un nouvel acte délinquant et *rien à voir* avec le fait, une fois cet acte commis, d'être soupçonné, identifié, poursuivi, condamné. Si par exemple, une étude fait le lien entre un quotient intellectuel plus élevé et un taux plus bas de « récidive », on en conclura généralement que le quotient intellectuel est un prédicteur de « désistance », alors qu'on pourrait tout à fait supposer qu'il est un prédicteur de traitement favorable par la machine pénale. Cet argument peut paraître tiré par les cheveux, mais des cas récents ont montré à quel point il est problématique : les lois de notification du voisinage prises aux États-Unis contre les délinquants sexuels ont entraîné une augmentation de 50 % de la « récidive » des personnes à qui on les a appliqué¹¹, doit-on y voir un échec (effet criminogène de la mesure) ou au contraire un succès (plus grande facilité de détection) ?

Enfin, le quatrième problème dans la construction de ces modèles prédictifs est le biais de sélection qui affecte les prisonniers sujets de telles études¹². Toute l'attribution différentielle de taux de récidive à des « profils » qui – de toute évidence – sont traités différemment par le système pénal suppose que ce traitement différentiel n'est ni criminogène, ni resocialisant, et que seul le profil détermine le taux de récidive. Evidemment, on sait se doute bien que cela doit être un peu faux, et les chercheurs essayent de corriger un certain nombre de différences de structure entre les deux populations, mais il est évident qu'elles ne peuvent être toutes corrigées.

C'est l'envers d'une critique que l'on fait régulièrement aux études favorables à la

libération conditionnelle¹³. De nombreuses études concluent que la libération conditionnelle « fonctionne bien » car ceux qui en bénéficient ont un taux de récidive plus bas. Le problème c'est que les candidats à une telle libération ne sont pas choisis au hasard, et on pourrait tout autant prendre ses chiffres comme une « validation » *a posteriori* des choix de ceux qui ont pris la décisions ou y ont contribué. Cette critique est connue. Mais il ne faut pas s'arrêter là. Si l'on ne peut attribuer à un aménagement de peine le faible taux de récidive de ceux qui en ont bénéficié, on ne peut non plus attribuer le fort taux de récidive de ceux qui n'en ont pas bénéficié à leur seul profil. *On ne peut pas avec les mêmes chiffres dire que les détenus « au bon profil » qui bénéficient d'un aménagement récidivent moins grâce à l'aménagement et que les détenus « au mauvais profil » qui font des sorties sèches récidivent plus à cause de leur profil.* Il faudrait savoir en amont quel poids attribuer au « profil » et quel poids attribuer au « traitement », et très peu de travaux (nous en avons recensé une poignée) proposent des approches innovantes pour se saisir entièrement de ce problème.

3. Tendance des experts à minimiser ces critiques. Ainsi, un scientifique qui participe au débat sur la récidive peut être amené à se prononcer sur trois questions :
- a. une prédiction de la récidive est-elle politiquement utile ? [première question préliminaire],
- b. une prédiction de la récidive est-elle techniquement faisable ? [deuxième question préliminaire],
- c. quelle est la meilleure méthode de prédiction de la récidive ? [question empirique].

Chaque question a, b ou c correspond à une demande politique au sujet de

¹¹ S. Raoult, Chronique de la recherche empirique internationale en sciences criminelles, *Problèmes actuels de sciences criminelles*, 24, 2013.
¹² S. Bushway and J. Smith, Sentencing using statistical treatment rules: what we don't know can hurt us, *Journal of Quantitative Criminology* 23.4 (2007) : 377-387.
¹³ Par ex., v. P. Combessie, *Sociologie de la prison ?* Repères, 2010.

laquelle un scientifique a de toute évidence, quelque chose à apporter, en articulant l'état actuel de la connaissance et son jugement de valeur éclairé. Il s'agit à chaque fois d'évaluer un savoir produit dans les coulisses de l'université et des centres de recherche. Néanmoins, ces questions ne sont pas indépendantes. En effet, les questions *a* et *b* sont les conditions de possibilité de la question *c*. Autrement dit, c'est seulement après avoir répondu « oui » aux questions préliminaires que la question empirique peut être posée.

Les travaux de Pierre Bourdieu sur l'*homo academicus*¹⁴ suggèrent que cette relation hiérarchique entre les trois propositions suscite un conflit interne chez les scientifiques qui travaillent exclusivement sur la question empirique. Produisant un travail qui représente un investissement lourd en termes de ressources et de temps, l'œuvre de cette sous-catégorie de chercheurs n'est justifiée que si les réponses aux questions préliminaires sont affirmatives. Une fois que le scientifique a investi une partie de sa carrière dans la question empirique il s'est créé ce que Bourdieu appelle un *capital académique* qui lui permet de bénéficier d'une position d'expertise sur cette question, d'un certain prestige social ainsi de ressources politiques et financières. Cela peut aller d'un avantage presque anecdotique (un poste pour un étudiant, une entrée de terrain pour la prochaine recherche) à la direction d'un institut. On peut avoir pour hypothèse que le développement de ce capital académique limite la capacité de celui qui tente de répondre à la question empirique à changer d'avis sur les questions préliminaires, au risque d'invalidier le sens

même de ses recherches précédentes et d'annihiler ces ressources durement acquises. Le chercheur qui doit à un point ultérieur de sa carrière répondre à nouveau aux questions préliminaires, par exemple du fait de nouveaux arguments ou de nouvelles données, est dans une position inconfortable.

Il y a deux façons de mesurer l'ampleur de ce phénomène. La première est quantitative. On se rend compte que, sur de nombreuses questions, la recherche empirique d'une équipe a tendance à confirmer les résultats passés de la même équipe, comme si les chercheurs validaient constamment leur travail antérieur dans leur nouveau travail. Cette tendance a été clairement identifiée au sujet des recherches sur la peine de mort, qui sont sans doute très idéologiques¹⁵. On pourrait s'attendre à ce qu'elle soit moins forte au sujet des recherches sur la récidive, qui sont sans doute moins politisées, mais cela n'est pas le cas. Comme on peut le voir au graphique 1, il y a la même tendance dans les deux domaines à « l'auto-confirmation »¹⁶.

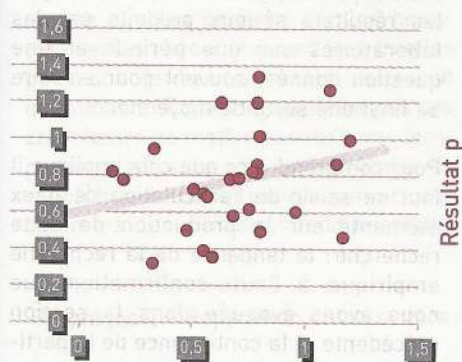
Ce graphique met en relation les résultats de chaque publication (*p*) réalisée par une équipe de chercheurs avec le résultat de la publication suivante (*p*+1) réalisée par la même équipe. Sans surprise, il y a un lien entre les deux résultats, c'est-à-dire que les équipes ayant obtenu un premier résultat *p* très favorable à un outil de prédiction ont tendance à trouver un résultat tout aussi favorable la fois d'après. Ce lien entre *p* et *p*+1 existe pour le pouvoir de prédiction des outils actuariels et pour la peine de mort, indiquant une même tendance à la « reproduction » de la recherche.

(14) P. Bourdieu, *Homo academicus*, Éditions de Minuit, 1984.

(15) S. Raoult, *Des méthodes et des hommes. La production sociale du savoir sur l'efficacité de la peine de mort*, Déviance et Société, à paraître.

(16) Les analyses statistiques faites dans le présent article s'appuient sur les chiffres produits par Yang et Lester pour la peine de mort (B. Yang et D. Lester, *The deterrent effect of executions: A meta-analysis thirty years after Ehrlich*, *Journal of Criminal Justice* 36.5 (2008) : 453-460) et par W. Yang et J. Coid pour la récidive (W. Yang, S. CP Wong et J. Coid, *The efficacy of violence prediction: a meta-analytic comparison of nine risk assessment tools*, *Psychological bulletin* 136.5 (2010) : 740.)

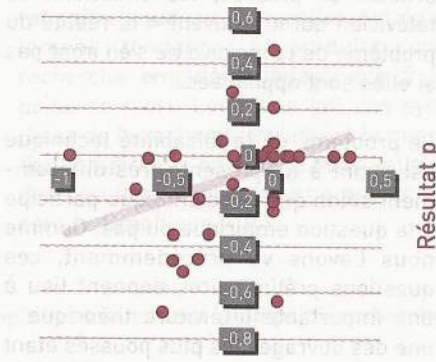
Graphique 1 - Corrélation entre le résultat (p) produit par une équipe
et son résultat suivant (p+1)



Résultat p+1
A. POUVOIR DE PRÉDICTION
DES OUTILS ACTUARIELS

D'un point de vue plus qualitatif, l'examen de la littérature scientifique, et notamment des introductions et conclusions des publications, expose cette tendance de ceux qui participent à la production de la recherche empirique à minimiser l'importance des questions préliminaires.

Dans ces articles, l'intérêt politique de la question de la récidive est souvent présenté comme « évident », des chercheurs vont parfois jusqu'à citer des sources déconcertantes : la presse, la télévision, des discours d'hommes publics ou encore des sondages sont mis en bas de page comme autant d'exemples de la réalité du « problème ». En même temps, on mentionne rarement les voix académiques dissonantes, et jamais ce qui est relatif à l'artificialité de cette « opinion publique » et à la fabrication du problème. La perception du problème est en quelque sorte une « preuve » de sa réalité, ce qui n'est pas très rigoureux. Il n'y a en fait pas tellement de différences entre l'introduction de travaux empiriques publiés dans cer-



Résultat p+1
B. CORRÉLATION
EXÉCUTION-HOMICIDES

taines revues à comité de lecture sur l'évaluation du risque de récidive et l'appel à l'émotion figurant dans les premières pages du rapport Lamanda, qui conclue son paragraphe sur les statistiques en remarquant que « Jamais les chiffres ne parviendront, en effet, à restituer l'intensité du traumatisme que subissent les victimes et l'impact du crime sur tous ceux qui peuvent s'identifier à elle »¹⁷. Il se peut bien qu'il y ait un problème politique dès lors que des personnes peuvent s'identifier à une victime, et il se peut bien que les « statistiques » n'importent pas, mais alors on voit mal quelle peut-être la contribution à la résolution de ce problème de méthodes de prédiction de la récidive appuyée sur des connaissances quantitatives. Les méthodes actuellement discutées dans les revues de médecine, de psychologie et de criminologie prétendent au mieux réduire « les chiffres de la récidive » de quelques points, elles ne se proposent sûrement pas d'empêcher qu'à l'avenir il y ait à nouveau de nouvelles victimes gravement traumatisées et des personnes qui s'identifient à elles (une

(17) V. Lamanda, *Amorindir les risque de récidive criminelle des condamnés dangereux*, rapport pour le président de la République, 30 mai 2008, p. 14

moindre couverture médiatique pourrait réduire cette seconde difficulté]. Les articles de presses, les émissions de télévision qui « prouvent » la réalité du problème de la récidive ne s'en iront pas si elles sont appliquées.

Le problème de la faisabilité technique est quant à lui présenté très différemment selon que le scientifique participe à la question empirique ou pas. Comme nous l'avons vu précédemment, ces questions préliminaires donnent lieu à une importante littérature théorique – une des ouvrages les plus poussés étant sans doute *Against Prediction* (2008) de Bernard Harcourt, qui expose les conséquences néfastes de la prédiction quand bien même elle serait théoriquement efficace. Ce type de problème est régulièrement évoqué dans la littérature empirique, également en introduction. Mais il est intéressant de remarquer que l'évocation est toujours très sélective (faute de place, supposera-t-on) et que la réponse usuelle est de déplorer ces difficultés mais de procéder « malgré tout », ne pouvant faire autrement. Souvent on minimise également l'impact que ces difficultés peuvent réellement avoir.

II – La surestimation du pouvoir de prédiction des outils actuariels

Ces dernières années, il se publie selon notre comptage autour de cinquante études originales par an dans la recherche anglophone rien que sur les méthodes actuarielles de prédictions du risque de violence, ce qui est bien supérieur à la plupart des domaines de recherche en sciences criminelles (à titre de comparaison, il y a à peu près vingt publications par an sur l'effet dissuasif des peines, trois sur l'efficacité de la peine de mort et deux sur le lien entre violence et armes à feu). Comment les chercheurs parviennent-ils à faire la synthèse d'un champ si productif ? La méthode privilégiée actuellement dans

tous les domaines de la recherche quantitative est de faire la méta-analyse des travaux publiés, c'est-à-dire de quantifier les résultats sérieux produits par les laboratoires sur une période et une question donnée, souvent pour en faire au final une sorte de moyenne.

Pour comprendre ce que cela implique il faut se saisir de l'articulation de deux éléments sur la production de cette recherche : la tendance de la recherche empirique à l'auto-confirmation que nous avons évoquée dans la section précédente et la contingence de la participation des scientifiques ayant les *a priori* les plus critiques à la production de la recherche empirique.

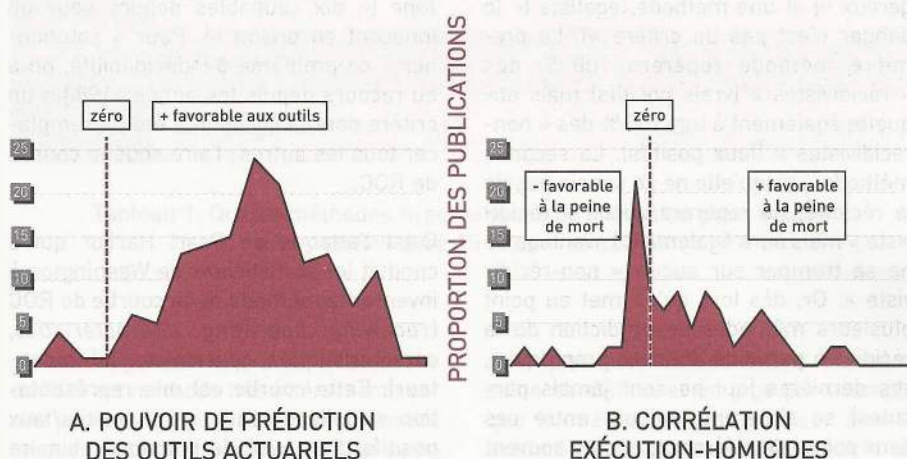
1. Tendance à la confirmation. Plus un domaine est dominé par l'auto-confirmation (plus, pour le dire autrement, la « production scientifique » devient une « reproduction scientifique »), plus la moyenne des travaux sur un domaine donné correspond en réalité à la *moyenne des opinions des équipes qui travaillent sur ce domaine, pondérée par leur productivité*. Autrement dit, on procède à une méta-analyse des recherches, pensant faire la moyenne des « études », mais on fait plutôt la moyenne des « points de vue », le point de vue de l'équipe ayant réalisé six études pesant six fois plus que celui de l'équipe en ayant réalisé une seule. Or, que se passe-t-il si des points de vue ne sont pas du tout représentés ?

2. Contingence de la participation des scientifiques ayant les *a priori* les plus critiques à la recherche empirique. Prenons un contre-exemple. Depuis 1970, une centaine d'études empiriques ont été conduites pour évaluer l'efficacité de la peine de mort à réduire les homicides. Sans surprise, une part substantielle des chercheurs qui produisent de telles recherches considèrent que l'efficacité de la peine de mort est une recherche politiquement utile et scientifiquement faisable. Mais pas tous. On trouve également un certain nombre de chercheurs

aux *a priori* négatifs, qui considèrent par exemple que le problème de la peine de mort n'est pas celui de l'efficacité, ou encore que les données sur les homicides et les exécutions sont trop fragiles pour que le lien entre les deux soit quantifiable¹⁸. Cela n'empêche pas ces chercheurs de mettre ces réserves temporairement « de côté » pour tenter de répondre à la question empirique, comme si les difficultés fondamentales

qu'ils connaissent par ailleurs n'empêchaient pas de faire cette recherche. Or, ces chercheurs qui ont d'emblée un regard critique sur la question ont également tendance à produire une recherche empirique défavorable à la peine de mort. Lorsqu'on fait une synthèse de la recherche relative à la question empirique une part substantielle des résultats publiés tourne autour de zéro (v. Graphique 2-B.).

Graphique 2 - Répartition des résultats empiriques dans deux champs de recherche



Mais les choses sont différentes en matière de récidive. Les chercheurs qui sont *a priori* opposés aux outils actuariels de prédiction de la récidive ne participent pas à la mesure de ce pouvoir de prédiction. La recherche empirique fait apparaître un consensus dans un sous-groupe de chercheur, en quelque sorte un consensus virtuel car l'entièreté de la communauté scientifique n'y participe pas. Les recherches sur le pouvoir de prédiction des outils actuariels sont artificiellement plus homogènes qu'elles ne devraient l'être et plus favorables à ces outils qu'elles ne pourraient l'être (Graphique 2-A.)

La comparaison des deux recherches est très frappante. Les résultats des recherches sur les outils actuariels sont répartis selon une courbe de Gauss (en cloche, c'est-à-dire, au hasard autour d'une moyenne) un peu à la manière des recherches favorables à la peine de mort (2-B, à droite des pointillés). Les deux graphiques sont très similaires à ceci près qu'il manque sur le graphique 2-A le « pic » à gauche de zéro, c'est-à-dire la recherche conduite par des scientifiques ayant des *a priori* critiques et qui conclue souvent en défaveur de l'outil. Il y a sans doute des raisons politiques et techniques à cela. Il est sans doute à la

(18) Par ex., John Donohue ou Edward Leamer.

fois plus facile et plus intéressant pour un chercheur théoriquement critique de la peine de mort de participer également à la recherche empirique sur l'efficacité de cette peine. Mais c'est un élément que l'on doit prendre en compte.

III – Le choix éthique dissimulé dans l'appréciation technique

Tout modèle de prédiction de la récidive se situe entre deux extrêmes : la méthode du risque zéro (« tout le monde est dangereux ») et une méthode légaliste (« le danger n'est pas un critère »). La première méthode repèrera 100 % des « récidivistes » (vrais positifs) mais étiquette également à tort 100 % des « non-récidivistes » (faux positifs). La seconde méthode, puisqu'elle ne se soucie pas de la récidive, ne repèrera aucun « récidiviste » mais aura également l'avantage de ne se tromper sur aucun « non-récidiviste ». Or, dès lors qu'on met au point plusieurs méthodes de prédiction de la récidive à partir de données empiriques, ces dernières (qui ne sont jamais parfaites) se situeront toujours entre ces deux points de référence, et plus souvent qu'on ne le croit, ce qui est gagné dans un domaine (vrai positif) est souvent perdu sur l'autre (faux positifs)¹⁹. Pour pouvoir comparer plusieurs méthodes entre elles, les statisticiens ont imaginé une douzaine de critères qui expriment ce que « fait » un système de diagnostic binaire tel que celui-ci : taux de faux positifs, de vrai négatifs, taux de fausse découverte, coefficient de corrélation de Matthews etc. Cependant, aucun modèle, mis à part une hypothétique prédiction parfaite, ne peut être le meilleur sur tous les critères.

Pour illustrer ce problème, imaginons deux méthodes simplifiées de prédiction de la récidive (la méthode 1 et 2) et com-

parons-les au « risque zéro » et à la méthode légaliste (tableau 1). Aucune méthode n'est meilleure sur tous les critères, il n'y a donc rien de tel qu'une « meilleure méthode » qui peut être décidée avec un regard neutre. Dans ces conditions, il est apparemment impossible de décider « scientifiquement » de la meilleure méthode, car chacune a ses avantages et inconvénients. Ces avantages et ces inconvénients sont d'ailleurs avant tout d'ordre politique, il s'agit de l'éternel dilemme entre davantage de sécurité et davantage liberté – on connaît par exemple le célèbre *ratio de Blackstone* (« dix coupables dehors pour un innocent en prison »). Pour « solutionner » ce problème d'indécidabilité, on a eu recours depuis les années 1980 à un critère dont la vocation a été de remplacer tous les autres : l'aire sous la courbe de ROC.

C'est l'attaque de Pearl Harbor qui a conduit les statisticiens de Washington à inventer la méthode de la courbe de ROC (*receiving operating characteristics*, caractéristiques opératoires du récepteur). Cette courbe est une représentation visuelle du ratio vrais positifs/faux positifs d'un outil de diagnostic binaire pour chaque palier de sensibilité (v. graphique 3). C'est-à-dire qu'au lieu de concevoir un outil avec deux étiquettes (dans notre cas, dangereux et non-dangereux) on conçoit un outil attribuant un « score » (par ex. de 0 à 20) et on mesure à chaque score le taux de succès et d'échec de l'outil.

Pour les radars, cette visualisation permet de savoir lesquels ont le plus de « bruit » à quel palier de sensibilité. Pour la récidive, chaque palier horizontal de la courbe représente le ratio entre les « récidivistes » que l'on a repéré (apparemment) à raison et les « non-récidivistes » que l'on a classés comme dan-

(19) Le meilleur ex. de ce modèle gagnant-perdant est la comparaison entre méthodes actuarielles (avec moins de faux positifs) et méthodes cliniques (avec davantage de vrais positifs), v. Litwack T. R., *Actuarial versus clinical assessments of dangerousness, Psychology, Public Policy, and Law* 7, n° 2, 2001, p. 409.

Méthode	Etiquette	« Récidivistes »	« Non-récidivistes »	Total
Méthode 1	Dangereux	10	30	40
	Non dangereux	5	55	60
Méthode 2	Dangereux	5	10	15
	Non dangereux	10	75	85
Zéro Risque	Dangereux	15	85	100
	Non dangereux	0	0	0
Légaliste	Dangereux	0	0	0
	Non dangereux	15	85	100

Tableau 1. Quatre méthodes hypothétiques de prédiction pour évaluer une situation avec 15 % de « récidive »

géreux (apparemment) à tort. La courbe de ROC est méthode très intéressante car elle permet de visualiser immédiatement ce que fait la méthode d'évaluation selon comment on veut l'utiliser, mais une telle représentation visuelle ne permet pas en elle-même de classer les diverses méthodes. Il y a bien des courbes dont on « voit » qu'elles sont « strictement meilleures » que d'autres (c'est-à-dire qu'elles ont un meilleur ratio à tous les seuils), mais souvent ce n'est pas le cas.

Pour parvenir à un classement linéaire (de la meilleure méthode à la pire méthode), il faut donc réduire les contributions de la courbe de ROC à « un » chiffre. Tant qu'on a deux chiffres différents, il se peut toujours qu'une méthode soit meilleure sur un des chiffres et moins bonne sur un autre, et les techniciens ne pourront pas dire au politique quelle est la « meilleure ». L'aire sous la courbe (AUC, *area under the curve*) est l'indice qui a été choisi pour remplir cette mission. Les méthodes actuelles

sont presque exclusivement classées selon ce chiffre, et on juge par exemple que la méthode X est meilleur que la méthode Y car elle a une AUC de 0,71 au lieu de 0,69. C'est le cœur de la recherche empirique actuelle.

Pour comprendre ce que signifie cet indice, imaginons que la « dangerosité » des individus soit évaluée au hasard. Dans ce cas, quel que soit le degré de sensibilité qu'on retient en utilisant cet outil, il devrait y avoir à peu près, pour autant qu'on ait un grand nombre d'individus, la même proportion de vrais positifs (y) et de faux positifs (x) : peu quand l'outil est peu sensible, moyennement quand il est moyennement sensible, beaucoup quand il est très sensible. La relation entre les deux variables sur un graphique cartésien est linéaire : c'est une ligne droite $f(x)=y$ qui coupe le graphique en deux. Si l'aire totale du graphique est de 1 alors l'aire sous cette droite est donc de 0,5. Voilà notre score d'aire sous la courbe.

Étape suivante : si la dangerosité est évaluée selon de véritables données empiriques, il devrait y avoir un moment où le taux de vrais positifs dépassera le taux de faux positifs au fur et à mesure que l'on fait varier la sensibilité de l'outil. La courbe se détachera alors de la ligne droite et l'on peut à nouveau calculer l'aire sous cette courbe, qui dépassera donc 0,5. Une AUC de 1 signifie une prédiction parfaite, et bien évidemment, un outil capable de faire une telle prédiction n'existe pas. Ainsi la signification de ces deux valeurs (0,5 et 1) ne fait pas débat. La question est que signifie véritablement une AUC entre 0,5 et 1 ? Par exemple, que signifie le fait qu'une AUC soit 0,01 points supérieure à une autre, le genre de différences qui amènent les spécialistes à privilégier une méthode plutôt qu'une autre ?

Imaginons deux méthodes. Une première avec une AUC de 0,65 et une seconde avec une AUC de 0,66 (donc prétendument « plus précise »). Qu'est-ce qui fait, concrètement, que la seconde est « meilleure » et doit être recommandée aux fabricants du Droit ? Représentons-nous le moment où l'AUC d'un modèle s'amplifie de 0,01 points pour passer de 0,65 à 0,66. Chaque mouvement de la courbe de ROC qui lui permet de s'éloigner de la ligne droite (et d'accroître ainsi l'aire sous la courbe) ne peut être que de deux natures (v. graphique 3). Il s'agit :

- soit d'un pas sur la gauche du graphique (x diminue, c'est-à-dire qu'il y a moins de faux positifs à un seuil donné) ;
- soit d'un pas vers le haut du graphique (y augmente, c'est-à-dire qu'il y a plus de vrais positifs à un seuil donné).

Ces deux mouvements sont les deux façons pour une AUC de passer de 0,65 à 0,66. S'il y a deux façons différentes d'augmenter cette aire de la même quantité, cela veut dire qu'il va bien falloir à un moment donné qu'un « poids » soit attribué aux vrais et faux positifs, il va falloir placer une équivalence entre un certain nombre de vrais positifs et un certain nombre de faux positifs. Concrètement, cela veut dire que n récidivistes apparem-

Graphique 3 - Exemple de courbe de ROC avec les deux mouvements qu'elle peut effectuer pour agrandir l'AUC.



ment repérés doivent valoir m non-récidivistes étiquetés « dangereux », apparemment à tort. Pour simplifier, il faut que ce 0,01 point d'AUC qu'on vient de gagner provienne soit d'une certaine quantité de faux positifs (si on l'a gagné en se déplaçant vers la gauche), soit d'une certaine quantité de vrais positifs (si on l'a gagné en se déplaçant vers le haut).

Or, par un hasard du sort, dans ces courbes mises en place pour les radars, l'équivalence se fait en termes de « pourcentage de signaux ». Ainsi pour augmenter l'AUC de 0,01 point, 1 % de vrais positifs à un certain seuil sont équivalents à 1 % de faux positifs. C'est comme cela que la méthode a défini l'équilibre, et c'est ce même ratio qu'on utilise, sans s'interroger sur sa signification, en sciences criminelles.

Ainsi, pour augmenter l'aire sous la courbe de ROC d'une quantité donnée, 1 % de vrais positifs est équivalent à 1 % de faux positifs à un même seuil. A chaque instant où l'aire sous la courbe s'amplifie, c'est par ces termes là qu'elle peut se mouvoir. En termes humains, cela veut dire que :

(a) le ratio *récidivistes repérés / non-récidivistes identifiés comme dangereux à tort* qui permet de juger de l'amélioration d'une méthode est entièrement

déterminé par le taux de « récidive » observé dans un contexte donné, (b) plus la « récidive » est rare dans un contexte donné, plus un « récidiviste » identifié à raison pèse par rapport à un non-récidiviste identifié à tort, (c) à titre d'illustration, dans un contexte tel que la délinquance sexuelle, si la « récidive » observée tourne autour de 15 % [ce qui est la moyenne des travaux internationaux récents sur la dangerosité ⁽²⁰⁾], un « récidiviste » identifié à raison vaut 5,67 « non-récidivistes » labélisés à tort ⁽²¹⁾.

Pour résumer, dans un contexte où il y a 15 % de « récidive », on peut *améliorer* un modèle d'un même nombre de points soit en identifiant à raison un récidiviste de plus à un seuil donné, soit en « innocentant » 5,67 non-récidivistes jusque-là identifiés à tort à ce même seuil.

Reprenons notre hypothétique méthode 1 (v. tableau 1). Cette méthode a une AUC de 0,66. Imaginons deux versions « améliorées » [méthode 1-A et méthode 1-B] de cette méthode. La méthode 1-A permettrait de repérer un récidiviste de plus à un seuil de sensibilité donné et la méthode 1-B en innocenterait 5,67 au même seuil. Les deux nouvelles méthodes auront la même AUC, elles seront jugées « équivalentes » dans la littérature actuelle. Mais évidemment, derrière la complexité du jargon et des modèles, c'est bien une question éthique qui est tranchée, la question du marchandage d'Abraham avant la destruction de Sodome et Gomorrhe, la question du ratio de Blackstone. Comment s'est-il trouvé que cette question soit aujourd'hui tranchée par « l'aire sous la courbe des caractéristiques opératoires du récepteur » ?

(20) L. R. Helmus et al., préc.

(21) 5,67 équivaut à 1 puisque qu'1 % d'une masse de 85 non-récidivistes équivaut à 1 % d'une masse de 15 récidivistes.