



HAL
open science

Des méthodes et des hommes. La production sociale du savoir sur l'efficacité de la peine de mort

Sacha Raoult

► **To cite this version:**

Sacha Raoult. Des méthodes et des hommes. La production sociale du savoir sur l'efficacité de la peine de mort. *Déviance et Société*, 2015, 39 (1), pp.99-121. hal-01196374

HAL Id: hal-01196374

<https://amu.hal.science/hal-01196374v1>

Submitted on 2 Mar 2016

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Des méthodes et des hommes. La production sociale du savoir sur l'efficacité de la peine de mort*

La recherche sur l'effet dissuasif des peines donne lieu à une impressionnante production quantitative, souvent indigeste et discordante. Une exploration des équipes qui travaillent sur la corrélation exécutions-homicides permet d'identifier une « minorité bruyante » de recherches très favorables à la peine de mort. Connaître l'auteur d'un article dans ce domaine est un meilleur prédicteur des corrélations trouvées que le type de données utilisé. On découvre alors un profil type de l'auteur qui a une proportion importante de résultats favorables à la peine de mort: il s'agit d'un économiste américain qui ne se présente pas comme spécialiste du crime et qui enseigne dans une faculté moins prestigieuse que celle où il a soutenu sa thèse, alors que les sociologues et les économistes des facultés hautement classés ont tendance à trouver des résultats plus proches d'un effet « zéro » de la peine de mort sur l'homicide.

Sacha Raoult

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

Introduction

Une des contributions essentielles de la pensée au XX^e siècle aura été la valorisation de *l'importance du regard sur la chose regardée* (Gide, 2012). Les penseurs de l'histoire des sciences ont développé l'idée que le travail scientifique est déterminé par le paradigme spécifique dans lequel il opère (Kuhn, 1983, voire des *conditions sociales de possibilité* épistémique (Foucault, 2014).

Les sociologues ont investi ce champ en s'interrogeant sur les processus de « fabrication » du fait scientifique à partir de positions diverses donnant plus ou moins de place à

* L'auteur tient à remercier pour leurs conseils, relectures et contributions les lecteurs anonymes de *Déviance et Société* qui ont grandement permis de clarifier la présentation de ce travail, ainsi que Philippe Combessie, Bernard Harcourt, Laurent Mucchielli, Andrew Abbott et les membres du Groupe Européen de Recherches sur les Normativités présents lors la présentation initiale de mes résultats à l'inter-labo du 21 juin 2013.

la subjectivité et au constructivisme (Latour, Woolgar, 2006; Barnes, Bloor, Henry, 1996; Abbott, 2001). Dans le même temps, des spécialistes de diverses questions se sont interrogés au cas par cas sur les biais d'auteurs susceptibles de déformer leur propre champ (Vaccaro, Patel, Fisher, 2011; Bresnick, 2008) et ont même proposé des outils statistiques pour repérer les biais de publication (Egger *et al.*, 1997). Ce débat n'a pas été sans conséquence sur l'organisation de la recherche, et participe au développement d'une plus grande transparence dans les rapports entre chercheurs et industries (Thompson, 1993; Bekelman, Li, Gross, 2003).

En revanche, il y a peu de tentatives de mesures systématiques du biais d'auteur en sciences humaines et notamment en sciences criminelles en dehors de la question du conflit d'intérêt financier. Néanmoins, on peut poser l'hypothèse que sur des questions fortement politisées, comme par exemple *l'efficacité de la peine de mort à prévenir des homicides par son effet dissuasif*, les positions antérieures de l'auteur, son parcours, sa formation, ses centres d'intérêts, peuvent prédire en partie le résultat obtenu.

Pour examiner la validité de cette hypothèse, cet article s'appuie sur un recensement récent de toutes les études sur la peine de mort depuis les années 1970 (Yang, Lester, 2008). Ce recensement dénombre une centaine de travaux, dont 85% prétendent répondre à la même question, à savoir « quelle est la corrélation entre exécutions et homicides ? »¹. Nous tentons de mesurer trois choses dans ce recensement. D'abord, on voudra rechercher si les résultats des travaux sur la peine de mort *sont* répartis d'une façon gaussienne, c'est-à-dire s'il y avait autant de résultats publiés au-dessus de la moyenne qu'en dessous et si les résultats « extrêmes » étaient plus rares que les résultats « moyens. » Cette méthode a été choisie car il s'agit du test classique du biais de publication et on en expliquera la signification *infra*. La deuxième étape de ce travail est la mesure du lien entre les résultats d'une publication et les résultats de la prochaine publication d'une même équipe. On y comparera *le pouvoir prédictif de la publication antérieure* avec le pouvoir prédictif du type de données utilisées – (l'explication récurrente des différences de résultats sur la peine de mort). Enfin, cet article établit des *profils de chercheurs* en fonction de leur résultat. On a procédé à une analyse systématique de tous les éléments que l'on pouvait extraire des CV des auteurs. Certains éléments, comme la discipline, le pays, le classement de l'université du dernier poste et les centres d'intérêt apparaissent comme très différents selon que l'on produit des travaux favorables ou défavorables à la peine de mort. Enfin, on proposera des pistes pour approfondir la signification de ces résultats.

Une enquête sur la nature d'un désaccord académique

On peut compter au moins 800 publications internationales qui s'attèlent à mesurer un possible effet dissuasif des peines (Dölling *et al.*, 2009; Yang, Lester, 2008). Certaines

¹ Le lecteur intéressé par un approfondissement des résultats de ce travail trouvera la liste des études analysées ici dans les catégories I, II et III de Yang et Lester (2008), le sous-ensemble des études analysées pour la question du biais d'auteur correspondant aux 64 études émanant de chercheurs ayant au moins produit deux travaux dans cette liste, ce qui rend leur travaux comparables entre eux. Cette bibliographie est accessible sur demande à sraoult@uchicago.edu.

de ces études indiquent que le criminel agirait rationnellement, répondant à la sévérité, à la certitude et/ou à la célérité de la peine comme si celle-ci était un « prix » qu'il était plus ou moins disposé à payer en fonction du gain qu'il espère tirer de l'infraction. Le modèle du délinquant rationnel, suggéré par l'économiste Gary Becker en 1968, serait ainsi confirmé. D'autres travaux, au contraire, sont plutôt compatibles avec la conclusion inverse – la perspective d'une réponse pénale aggravée n'aurait aucun impact sur le comportement des personnes, et l'élément dissuasif de la peine serait *si tant est qu'il existe, d'une importance très éloignée des prédictions du modèle* (Paternoster, 2010).

La peine de mort à elle seule a donné lieu entre 1970 et 2013 à plus d'une centaine de travaux. Certaines études concluent à un effet particulièrement dissuasif de la peine de mort, *même pour les crimes passionnels* (Rubin, 2006) – et leurs auteurs avancent que jusqu'à 18 vies pourraient être sauvées pour chaque personne exécutée. D'autres études, les plus nombreuses, ne trouvent pas de corrélation statistiquement significative entre exécutions et meurtres, ce qui amène leurs auteurs à suggérer que la peine de mort aurait plutôt à un effet « zéro » sur le crime. Un dernier petit groupe d'études aboutissent à un effet dit « brutalisant » de la peine de mort, c'est-à-dire qu'elles révèlent une corrélation positive entre le nombre d'exécutions et le nombre d'homicides, une relation qui correspond à l'argument célèbre de Beccaria selon lequel la peine capitale aurait pour effet de banaliser la mort et d'égarer *une tête déjà disposée au crime* (1870, 104).

Si elle peut surprendre un profane, cette disparité des résultats n'est pas étonnante en soi. Des travaux différents, conduits dans des contextes différents à partir de méthodes et des données différentes peuvent parvenir du seul fait du hasard à des résultats différents. Pour ce qui est de la peine de mort, des auteurs ont proposé plusieurs formes d'explication pour cette différence. Aucune de ces explications ne fait l'unanimité et nous reviendrons sur ce point *infra*. Mais notre impression de départ est que si la disparité des résultats était la seule forme d'hétérogénéité que l'on trouvait dans cette recherche quantitative, il n'y aurait pas de grand mystère à résoudre. On ne regarde pas les mêmes choses, on ne rend pas compte de mêmes choses, certes. Mais ce qui est plus difficile à comprendre, c'est que cet éparpillement des conclusions se retrouve à un autre niveau du monde académique.

Disparité dans les états des savoirs

Les travaux quantitatifs publiés dans les revues à comité de lecture sont souvent introduits par un « état des savoirs » qui résume la littérature précédente dans laquelle celui-ci s'inscrit. Or, si plusieurs études peuvent tout à fait obtenir des résultats différents du fait de choix méthodologiques ou du contexte de leur travail, il n'y a en revanche pas de raison évidente pour que ces états des savoirs diffèrent d'un article à l'autre à une époque donnée. Tous les scientifiques ont théoriquement accès à la même littérature. Si nous prenons la recherche qui se limite à l'établissement de *corrélations entre le nombre d'exécutions et le nombre d'homicides*², on trouvera aisément deux extrêmes dans les « états des savoirs ». Zhiqiang Liu dans son article intitulé *Capital Punishment and the Deterrence Hypothesis* :

² Selon Yang et Lester (2008) ce type d'étude constituait en 2008, 85% des études quantitatives sur l'effet dissuasif de la peine de mort – qu'elles soient appuyées sur des données temporelles, géographiques ou sur des panels de données. 5% des études consistaient à regarder l'évolution du taux d'homicide dans les jours suivants une exécution et 10% à faire le lien entre la publicité donnée à l'exécution et le nombre d'homicides.

Some New Insights and Empirical Evidence publié en 2004 dans *Eastern Economic Journal* résume de la façon suivante cette littérature : *Pour l'essentiel, les économistes ont découvert une corrélation négative et statistiquement significative – c'est-à-dire un effet dissuasif – entre les exécutions et le meurtre* (p. 237). Mais à l'inverse, Zimring, Fagan et Johnson introduisent ainsi leur propre travail publié quelques années plus tard : *les données disponibles [sur la peine de mort] sont faibles, en grande partie parce qu'il y a si peu d'exécutions là où elle est principalement étudiée* (2010, 1) *On ne peut pas détecter un quelconque effet dissuasif aux États-Unis* (2).

Il n'est pas anodin que la disparité dans les « états des savoirs » reflète en majorité la disparité des résultats. L'article type présente ses conclusions dans le prolongement de la littérature existante et non à contre-courant : une équipe a mesuré ce phénomène pour l'ensemble des études sur l'effet dissuasif des peines. Selon ces auteurs, 70% des papiers présentent en introduction un état des savoirs concordant avec la conclusion de l'article (Dölling *et al.*, 2009, 206). Nous sommes en quelque sorte dans ce que Thomas S. Kuhn appelait la *science normale*, c'est-à-dire la recherche consistant à confirmer le paradigme dominant (Kuhn, 1983), à ceci près qu'il aurait plusieurs normalités concurrentes – une hypothèse que nous approfondirons *infra*.

Disparités dans les témoignages d'experts

Face à une telle cacophonie au sein d'une littérature quantitative indigeste et souvent lourdement mathématisée, on peut s'attendre à ce que le législateur ait besoin de recourir à des « autorités » pour faire la part des choses lorsqu'il se pose une question apparemment aussi basique que l'efficacité de la peine de mort à prévenir des meurtres. Sans surprise, cette expertise reproduit dans un autre lieu les mêmes schémas qu'un parcours de la littérature, souvent avec les même acteurs, puisque ce sont principalement les auteurs des recherches empiriques qui sont amenés à se prononcer comme experts, non pas sur leurs seuls travaux, mais sur l'entièreté du champ de recherche auquel ils participent.

Voici deux séries d'exemples de ces expertises en sens contraire. Au cours de l'année 2011, le Parlement du Connecticut envisagea une loi n° 1035 pour « Abolir la peine de mort et y substituer la réclusion perpétuelle sans possibilité de libération conditionnelle » pour les meurtres les plus graves. Le premier expert, John Donohue, recommande à partir d'un compte rendu de la littérature de déplacer le budget employé pour la peine capitale vers le recrutement de policiers : *À la différence de la peine de mort, il a été démontré qu'investir dans la police était une façon de réduire le coût social des crimes, y compris des homicides* (Connecticut Judiciary committee, 2011). Devant cette même commission, Keri O'Creene s'exprimera en des termes exactement contraires et résumés de façon tout aussi directe par le rapport au soutien de la loi : *Trente années de recherche académique sérieuse démontrent que la peine capitale sauve des vies*. Celui qui est familier avec les codes de ce monde professionnel sera immédiatement frappé du fait que les intervenants n'ont pas le même statut social dans le monde académique. Le premier interlocuteur de la commission est un célèbre professeur de Stanford, mais il nous a été impossible d'avoir des informations sur le second. Cette différence de reconnaissance au sein de l'université n'est pas anecdotique et nous verrons *infra* qu'elle illustre en réalité un phénomène plus général.

Deuxième exemple: le *Senate Judiciary Committee on the Constitution, Civil Rights, and Property Rights* des États-Unis d'Amérique au cours de l'année 2006. Devant cette commission, Paul Rubin, *Professor of Economics and Law* à Emory University fournit un témoignage intitulé « Preuve statistique sur la peine capitale et la dissuasion de l'homicide ». Ce témoignage comporte 3 000 mots, et il est étoffé d'une bibliographie d'une trentaine d'articles. On peut être étonné par la facilité avec laquelle Paul Rubin ordonne la disparité des articles publiés sur la peine de mort: *La quasi-totalité des études modernes, et la totalité des études examinées par des pairs (refereed studies) concluent à un effet dissuasif [statistiquement] significatif de la peine capitale. Une étude seulement interroge ces résultats. Pour un économiste, ce que je vous dis n'est pas surprenant: nous nous attendons à ce que les criminels ainsi que les criminels potentiels répondent à la menace d'une sanction, et la peine de mort est la sanction la plus sévère parmi celles que nous pratiquons.* En somme, écrit-il, *La littérature scientifique est facile à résumer.*

Devant la même commission d'enquête, Jeffrey Fagan, *Professor of Law and Public Health* à la Columbia Law School, parvient à la conclusion exactement inverse. Le témoignage de Fagan est un peu plus long car il s'attarde davantage sur le contexte de chaque étude – au lieu de chercher à trancher le nœud gordien, Fagan le démêle lentement. La synthèse tient néanmoins en un paragraphe: *Les études [favorables à la peine de mort] sont remplies d'erreurs techniques et conceptuelles: données manquantes; variables contrôles non testées; impossibilité de répliquer les résultats à partir des mêmes données, et autres.*

On peut avoir l'impression que les dissensions se résument à un problème de qualité *versus* quantité des travaux évoqués, mais ce n'est pas vraiment selon ce schéma asymétrique que la question est présentée à la commission. Paul Rubin, au sujet de la remise en cause de ses travaux, telle que celle évoquée par Jeffrey Fagan, s'adresse à la commission dans les termes suivants: *Les résultats de cette remise en cause³ n'ont pas été évalués par des chercheurs compétents. La Stanford Law Review, comme toutes les revues de Droit, est éditée par des étudiants qui n'ont pas de compétence particulière en économétrie.* Chacun présente ainsi au politique un argument à la fois quantitatif et qualitatif. Selon chaque expertise, les travaux les plus nombreux sont favorables au point de vue défendu, et les travaux remettant en cause ce point de vue sont de mauvaise qualité.

Deux types d'explications classiques du désaccord

Deux types d'explication ont été offerts pour rendre compte de la disparité des résultats sur l'effet dissuasif des peines et sur la peine de mort en particulier.

Le premier type d'explication regroupe les arguments propres au domaine étudié. L'effet dissuasif de la peine varierait non seulement en fonction des peines et des comportements étudiés, mais aussi, pour la seule peine de mort, en fonction des États étudiés (Shepherd, 2005), du type de données employées (Yang, Lester, 2008), ou encore d'autres paramètres tels que les variables de contrôle utilisées (Dölling *et al.*, 2009). Ce serait parce que le phénomène de l'effet dissuasif des peines est lui-même hétérogène que les travaux publiés

³ Il s'agit de Donohue et Wolfers (2006), une analyse de sensibilité remettant en cause plusieurs travaux favorables à la peine de mort, citée dans le témoignage de Jeffrey Fagan.

rendent compte d'un phénomène hétérogène. La disparité scientifique reflète une disparité dans la réalité, et la mission des scientifiques serait de découvrir les cas dans lesquels la peine fonctionne et ceux dans lesquels elle ne fonctionne pas.

Le second type d'explication n'est pas propre à l'effet dissuasif des peines mais à la recherche empirique en général. On pourrait y regrouper les arguments relatifs au sérieux des études, aux biais et prédispositions des auteurs qui sont évoqués pour expliquer des résultats divergents (Leamer, 1982; McManus, 1985; Donohue, Wolfers, 2006). C'est le regard, et non ce qui est regardé, qui explique la disparité. On comprend alors le « chiffre » comme une *construction sociale* et on s'intéresse à ce qui fait que différentes entreprises scientifiques construisent un chiffre différent à partir d'une même réalité.

Le premier type d'explication a fait l'objet de nombreuses mesures. La communauté scientifique dispose d'un certain nombre d'outils pour quantifier les disparités au sein d'une littérature donnée, et l'un d'eux est la publication de méta-analyses. Trois travaux récents sur l'effet dissuasif des peines cherchent principalement à attribuer la disparité des résultats à de telles *différences de méthodes* (Dölling *et al.*, 2009; Yang, Lester, 2008, Durlauf, Fu, Navarro, 2013), c'est-à-dire en fait à des différences dans les données ou les modèles utilisés. Comme dans de nombreux autres domaines, des méta-analyses produisent des comparaisons entre les travaux selon le type de données qu'ils utilisent ou le type d'infraction auxquels ils s'intéressent. On y apprend alors que les études sur l'effet dissuasif du système pénal ne donnent pas les mêmes tailles d'effet selon que l'on s'intéresse à l'effet dissuasif de la sévérité de la peine ou de sa certitude, mais aussi que les travaux sur la peine de mort sont plus favorables à cette dernière si l'on emploie des séries temporelles plutôt qu'un panel de données. Parfois, ces différences « font sens », alors on saute sur une explication : si la corrélation négative sanction-délit est plus forte dans les études des atteintes aux biens que dans celle des atteintes aux personnes, c'est sans doute car ce sont des délits plus rationnels (Dölling *et al.*, 2009). Les publications sont traitées comme des fenêtres sur la réalité positive particulière dont elles traitent, et le rôle de la méta-analyse serait alors d'accorder ces réalités entre elles, voire de montrer leur cohérence globale avec un modèle.

En concurrence avec ces explications, la recherche d'un biais d'auteur a été lourdement développée en théorie mais n'a pas tellement donné lieu à des grandes tentatives de quantification (v. pour une intégration de la variable « profession de l'auteur » à une méta-analyse, Gerritzen, Kirchgässner, 2013). Pourtant, expliquer la différence de résultats par la méthode employée n'explique pas pourquoi différents auteurs rendent compte de l'intégralité du champ différemment – ce que l'on peut observer en lisant les « états des savoirs » ou en écoutant les expertises. Autrement dit, il se peut bien que certains types de données ou de modèles soient plus à même de fournir des résultats favorables à une thèse que d'autres, mais cela n'explique pas pourquoi tous les auteurs ne comprennent pas l'ensemble de la littérature de la même manière.

De plus, il est dangereux de chercher à s'arrêter à une réponse propre à la peine de mort ou à l'effet dissuasif des peines, au risque de manquer derrière tout cela une question plus profonde, plus générale. Après tout, la situation dans laquelle se trouvent les recherches sur la peine capitale n'est pas unique, on la retrouve en fait dans bien d'autres domaines : une littérature quantitative interdisciplinaire relativement indigeste et qui donne lieu à des résultats, des interprétations, des expertises radicalement opposées. Une littérature définie par *ses diversités* (produites par des chercheurs issus de disciplines et de cercles différents,

dans des contextes différents, et qui n'emploient presque jamais deux fois exactement la même méthodologie, ce qui rend les comparaisons difficiles), et *ses régularités*: lieux de publication (revues à comité de lecture), forte mathématisation des recherches, une certaine opacité dans la démarche (par opposition aux simples comparaisons de courbes de criminalité). Cette accumulation de littérature diverse, opaque et discordante, on la trouve dans de nombreux domaines de recherche: l'évaluation du risque de récidive, notamment des délinquants sexuels (où on a récemment tenté de revenir à des données plus sobres, voir Helmus *et al.*, 2012), la mesure de l'effet criminogène de la prison ou de son absence (Jonson, 2010; Lipsey, Cullen, 2007), ou encore de l'influence des « sous-cultures » sur le niveau de criminalité (Pratt, Cullen, 2005) en sont quelques exemples. Certains champs, comme la corrélation entre la libre circulation des armes à feu et les homicides, sont très semblables aux recherches sur l'effet dissuasif de la peine de mort: on y retrouve les mêmes modèles, les mêmes querelles, et parfois les mêmes intervenants (Donohue, Ayres, 2009]. À chaque fois des chapelles, des écoles, des réseaux, des disciplines entières peut-être s'affrontent en produisant des chiffres que l'accumulation de la littérature scientifique rend de plus en plus difficile à résumer. On est donc amené à se demander si c'est à chaque fois des différences particulières de contexte, de domaines propres à la question posée qui expliquent les disparités, ou quelque chose de plus général qui tient à la diversité des perspectives des chercheurs qui réalisent ces travaux.

D'où la nécessité d'employer des méthodes quantitatives établies et pour quantifier le pouvoir explicatif de ce deuxième type de perspective, afin de la mettre sur un pied d'égalité avec les explications plus positivistes, de chercher si, tout autant que la « réalité », l'auteur n'est pas une piste pour comprendre les disparités dans les résultats de la recherche.

La distribution non gaussienne des résultats

Il est attendu que des études conduites sur un même sujet mais dans des contextes différents et employant des méthodes et des données différentes parviennent à des résultats différents. Mais si ces différences méthodologiques ne sont que le fruit du hasard, on pourrait également s'attendre à ce que les résultats soient à peu près également répartis autour de la moyenne qui serait également la médiane. Les champs dont les résultats ne sont pas répartis « normalement » autour de la moyenne souffrent possiblement d'un biais de publication (Egger *et al.*, 1997) – c'est-à-dire que certains résultats sont peut-être disproportionnellement sur-rapportés ou sous-rapportés auprès des revues.

Yang et Lester ont produit une bibliographie de tous les travaux quantitatifs sur la peine de mort (100) publiés entre 1970 et 2008, répartis dans cinq catégories (I. Séries temporelles; II. Données géographiques; III. Panels de données; IV. Publicité de l'exécution; V. Impact d'une seule exécution sur l'homicide les jours suivants). 85 de ces articles (les catégories I, II et III) sont en fait intéressés exactement à la même question, à savoir la corrélation entre le nombre d'exécutions et le nombre d'homicides, et ce n'est ainsi pas la question de recherche qui les différencie mais le type de données (temporelles, spatiales ou mixtes).

La moyenne que Yang et Lester donnent de l'ensemble de ces résultats est de -0,126, ce qu'ils interprètent comme *un résultat au soutien de la thèse d'un effet dissuasif de la peine de mort* (Yang, Lester, 2008, p453). Si les tailles d'effet moyennes de ces articles étaient

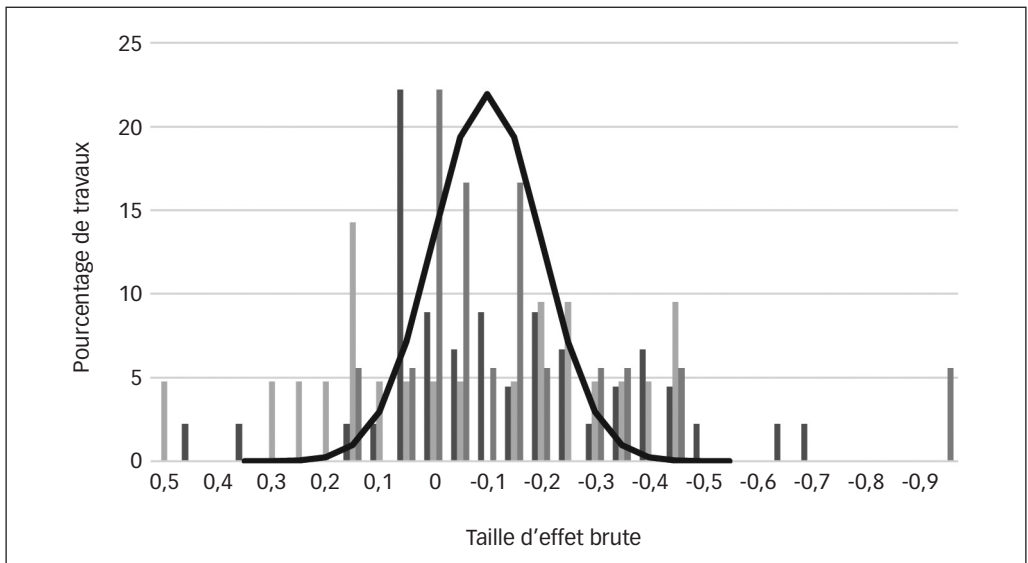
réparties normalement autour de $-0,126$, on devrait s'attendre à ce que les résultats des études soient regroupés sous une courbe de Gauss centrée sur ce chiffre.

Or, le graphique 1 montre que tel n'est pas le cas. Ce graphique exprime à l'intérieur de chaque catégorie la proportion de travaux publiés pour chaque taille d'effet ($\pm 0,025$). Chaque niveau de gris représente une des trois types de données mentionnés antérieurement (le graphique 3 discuté *infra* revient sur la distribution des résultats pour chaque type de données). Sur cette répartition, on peut coucher une courbe de Gauss qui représente une loi normale ayant pour moyenne $-0,125$ (la moyenne calculée par Yang et Lester) et un écart type de $-0,1$. On remarque sans difficulté que cette courbe ne représente pas du tout la façon donc ces résultats sont répartis.

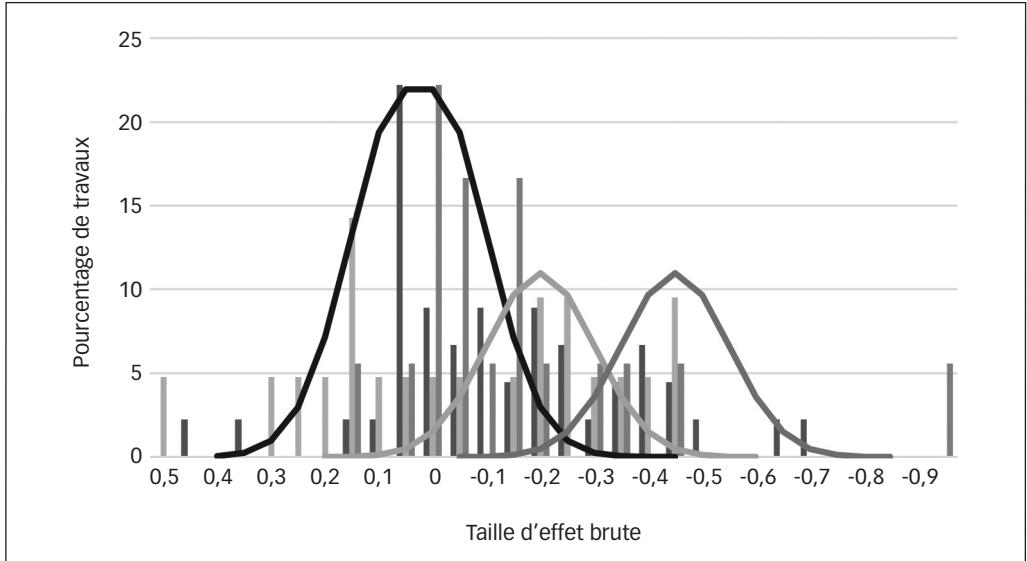
On peut représenter par le graphique 2 la façon dont les résultats sont réellement répartis: une majorité de publications gravitent « normalement » autour de $0,025$ (soit un effet *quasi-nul* de la peine de mort) avec un écart type de $-0,1$, et deux groupes de recherches, plutôt minoritaires, sont réparties à peu près normalement autour de $-0,2$ et sur la gauche de $-0,45$. Ces recherches minoritaires sont extrêmement favorables à la peine de mort, mais semblent correspondre à une autre normalité scientifique. Leur présence et l'extrémisme de leur résultat ont pour effet de tirer la moyenne générale vers le bas et de donner l'impression que l'ensemble du champ tend à démontrer un effet dissuasif de la peine de mort, alors que le résultat le plus commun est nul ou quasi-nul. Ainsi, la moyenne des résultats ne décrit pas une réalité positive qu'une recherche unifiée tente de déceler mais la moyenne de trois normalités scientifiques, et si cette moyenne indique quelque chose sur la production sociale du savoir sur la peine de mort, il n'a pas de grande signification sur la peine de mort comme phénomène.

Cette division entre une recherche majoritaire qui conclut à un effet nul de l'exécution sur l'homicide et des recherches minoritaires qui parviennent à la conclusion inverse n'est

Graphique 1 : Discordance entre les résultats des recherches de corrélations entre exécutions-homicides et ce qui devrait être attendu d'une répartition gaussienne de leur moyenne



Graphique 2 : La recherche majoritaire et deux recherches minoritaires.



pas dépendante des données utilisées – et ce alors que les données utilisées sont l'explication la plus commune de la disparité des résultats (Yang, Lester, 2008). Le graphique 3 montre que 4 des 5 types de données utilisées ont exactement la même répartition : que l'on s'intéresse à la corrélation exécutions-homicides dans le temps (A), dans l'espace (B), dans un panel de données (C), ou même que l'on cherche le lien entre publicité de l'exécution et homicide (D), ce qui n'est pas vraiment la même problématique, on retrouve à chaque fois le même phénomène : une majorité d'études gravite autour de zéro, alors qu'une minorité conclut à des résultats extrêmes. Le seul sous-groupe dont la répartition des résultats diffère est celui qui est composé d'une poignée de travaux sur l'impact d'une exécution unique sur les homicides dans les semaines suivantes. Il n'y en a que cinq et ils sont produits par seulement 3 équipes différentes ; il est donc difficile de les analyser globalement ou de les comparer aux autres.

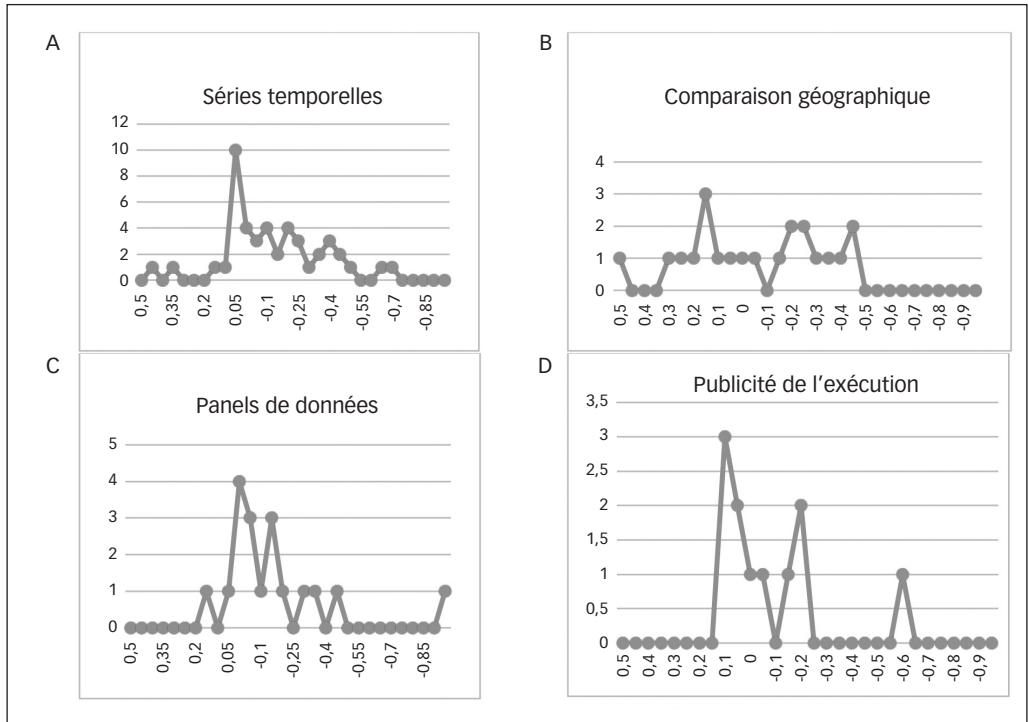
Cette distribution non gaussienne nous amène à nous demander si plutôt que de chercher un biais dans les données, il ne serait pas plus riche d'approfondir le biais d'auteur, et rechercher si ce ne sont pas toujours les mêmes auteurs qui ont tendance à produire les mêmes résultats.

L'identification d'un biais d'auteur

Pour chaque publication, Yang et Lester donnent deux indications :

- *une taille d'effet moyenne* qu'ils ont calculée à partir des données disponibles dans l'article original (2008, 453).
- *les conclusions de l'auteur*, car il se peut qu'il y ait une divergence entre les conclusions de l'auteur et la taille d'effet calculée.

Graphique 3 : Nombre d'études correspondant à chaque taille d'effet dans quatre type de travaux



Par exemple, un auteur peut trouver une taille d'effet favorable (négative) ou défavorable (positive) à la peine de mort, mais insister sur la fragilité de ses résultats. L'exemple le plus représentatif de cette discordance est sans doute cette conclusion d'Edward Leamer suite à une taille d'effet négative : *mon sentiment est que ces données sur l'effet dissuasif de la peine de mort sont trop fragiles pour que l'on y donne foi* (1983, p42). Yang et Lester notent alors *conclusions mixtes* dans leur base de données, à côté de la taille d'effet découverte par Edward Leamer.

Pour notre étude, nous avons repris leur bibliographie en faisant correspondre deux résultats chiffrés à chaque publication :

- la taille d'effet brute qu'ils ont eux-mêmes calculée.
- une taille d'effet corrigée, c'est-à-dire un chiffre qui tient compte des conclusions de l'auteur.

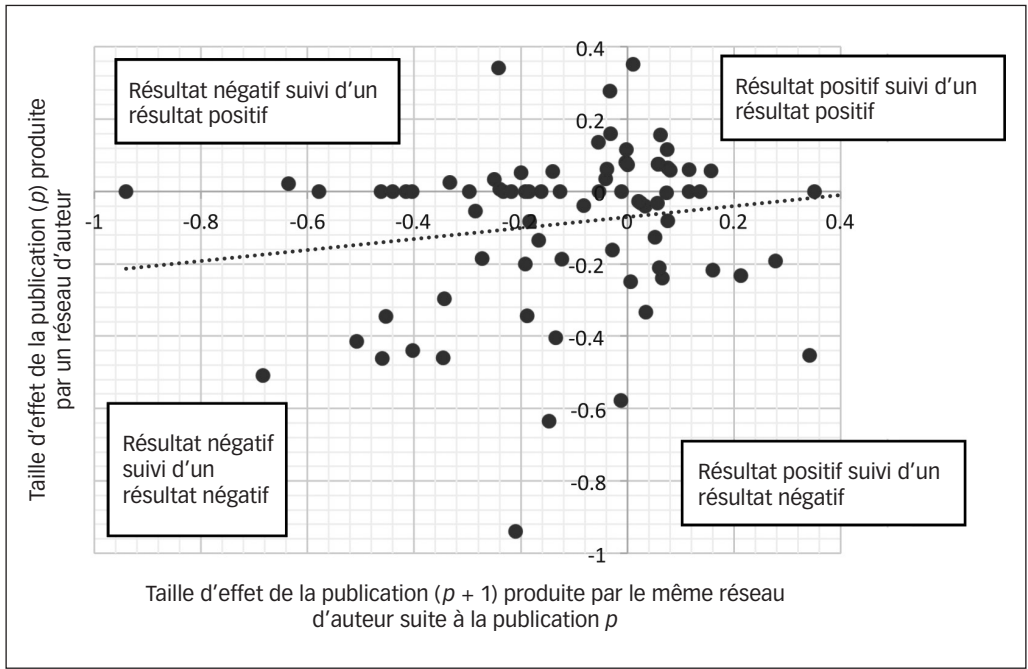
Pour coder une « taille d'effet corrigée », nous avons conservé la taille d'effet négative lorsque Yang et Lester notent *conclut* à un effet dissuasif, et conservé une taille d'effet positive lorsqu'ils notent *conclut* à un effet brutalisant. En revanche, s'il y a une contradiction entre la taille d'effet brute et les conclusions, cela produit une « correction » dans le chiffre : et nous avons ainsi transformé en taille d'effet nul les *conclusions mixtes* et *conclut* à un effet nul.

L'intérêt d'avoir ces deux séries de chiffres pour chaque publication, est qu'on verra que toutes les tendances que nous avons découvertes dans notre étude sont davantage accentuées pour les tailles d'effet corrigées que pour les tailles d'effet brutes. C'est-à-dire que les auteurs ont tendance à donner une conclusion subjective encore plus prévisible que la taille d'effet brute qu'ils découvrent.

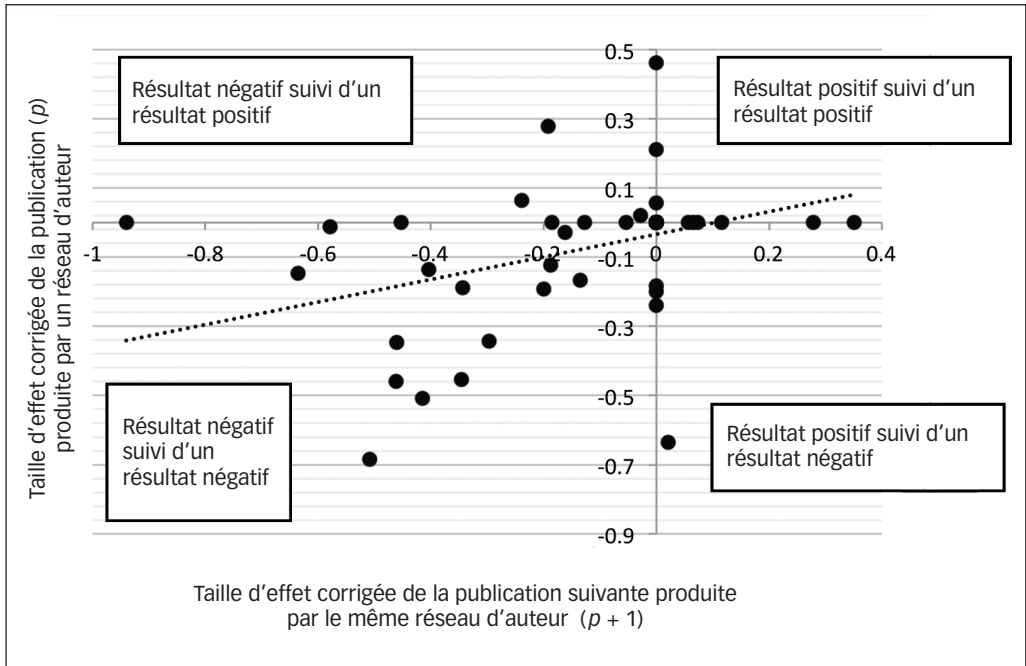
Qui produit les chiffres sur la peine de mort ? Il nous faut distinguer au sein des 85 travaux étudiés une sous-catégorie de publications produites par ceux que l'on pourrait appeler les *professionnels de la corrélation exécutions-homicides*, c'est-à-dire les équipes qui ont consacré plus d'un papier à cette question. On a défini ces équipes de la façon la plus large possible, afin d'avoir un nombre conséquent de données, une équipe comme tout réseau de personnes ayant au moins signé un article ensemble en tant que coauteurs sur notre liste. Selon cette définition, nous pouvons décompter 28 professionnels de la corrélation exécutions-homicides de 1970 à 2008, répartis dans 16 équipes et responsables de 64 articles publiés sur 85, donc d'un part substantielle du champ. En ayant resserré le domaine de cette manière, nous pouvons à présent rechercher la corrélation entre le résultat d'un papier et le résultat précédent de la même équipe.

Notre première façon de mesurer cette relation est de comparer le résultat d'une publication p produite par une équipe et la publication suivante $p + 1$. Si les tailles d'effet brutes de la publication p et $p + 1$ d'une même équipe étaient entièrement indépendantes, les points ayant pour coordonnées $(p, p + 1)$ seraient répartis hasardeusement sur le graphique 4. Or, cela n'est pas le cas et on voit une légère tendance pour les publications ayant un certain résultat à être suivi d'un résultat relativement proche (c'est la tendance représentée par la

Graphique 4 : Tendance d'un réseau d'auteur à confirmer ses résultats obtenus dans la publication p lors de la publication subséquente $p + 1$ (tailles d'effet brutes)



Graphique 5 : Tendance d'un réseau d'auteur à confirmer les résultats obtenus dans la publication p lors de la publication subséquente $p + 1$ (tailles d'effet corrigées)



droite). Dans le quart inférieur gauche du graphique, qui correspond aux résultats négatifs (favorables à la peine de mort) suivis d'un résultat autre résultat négatif, on voit se dessiner une droite, qui indiquent des résultats se suivant et relativement similaires (les résultats les plus extrêmes sont suivis de résultats extrêmes, les plus modérés suivis de résultats modérés). Le graphique 5 applique ce même regard aux tailles d'effets corrigés, et la tendance est alors aggravée, ce qui signifie que les auteurs ont une encore plus grande tendance à produire deux fois d'affilée les mêmes conclusions qu'à donner deux fois consécutives le même chiffre.

Une seconde comparaison que nous pouvons faire est la moyenne des résultats des publications selon qu'on les catégorise par données utilisées ou par réseau d'auteur. Cette comparaison est faite au graphique 6. On y découvre que connaître le premier résultat obtenu dans la carrière d'une équipe d'auteurs est un meilleur prédicteur du résultat des études subséquentes que de savoir si l'étude a employé des données temporelles, géographiques ou un panel de données, alors même que ce critère est régulièrement présenté comme étant le plus grand « biais » susceptible d'affecter une étude sur la peine de mort.

Nous avons à présent deux éléments. D'abord, la recherche sur la corrélation entre le nombre d'homicides et le nombre d'exécutions est toujours, quelles que soient les données utilisées, répartie de la même manière : une majorité d'études ne trouve pas de corrélation, et une minorité d'études trouve des corrélations anormalement fortes. Le second élément que nous avons découvert est la tendance des chercheurs à confirmer leurs résultats antérieurs. Ainsi, nous pouvons répartir les équipes en fonction de leur taux de résul-

Graphique 6 : Moyenne des tailles d'effet brutes des publications selon les données utilisées (colonnes 1 à 3) ou moyennes des tailles d'effet brutes des articles subséquents d'une équipe selon le premier résultat publié par cette équipe (colonnes 4 à 8)

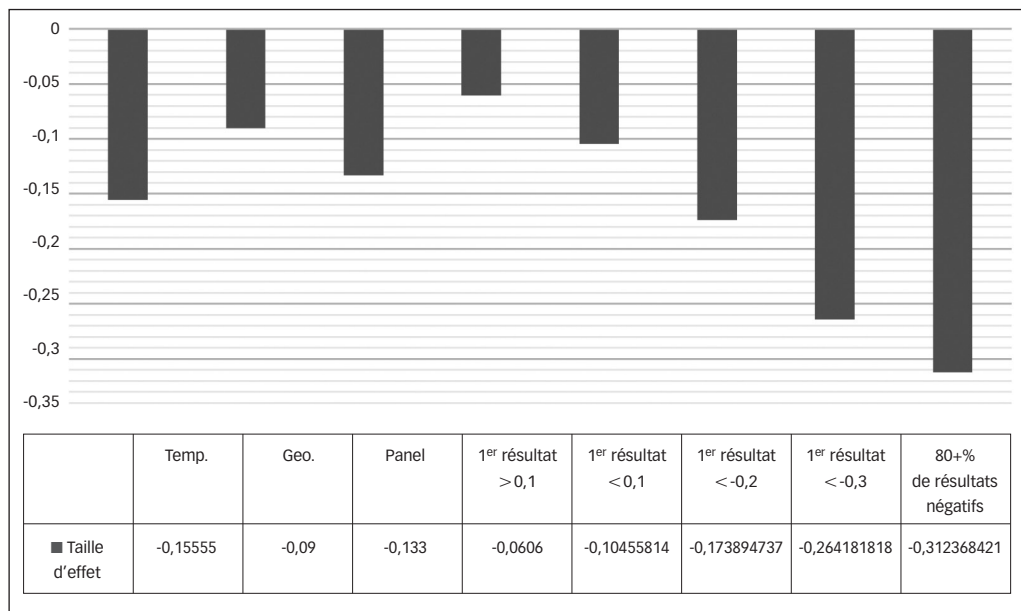


Tableau 1 : Réseaux d'auteurs ayant publié en 2008 au moins deux études quantitatives sur la corrélation exécutions-homicides

Équipe	Nombre d'études	Tailles d'effet corrigées négatives	Probabilité hypergéométrique
Bechdolt	2	0%	39%
Bowers et Pierce	2	0%	39%
Decker et Kohfeld	5	0%	8,6%
Fox et Radelet	2	0%	39%
Mcaleer et Veall	2	0%	39%
Passel e Taylor	2	0%	39%
Peterson et Bailey	20	10%	0,2%
Yang et Lester	4	25%	37%
Avio	2	50%	47,6%
Leamer	2	50%	47,6%
Zimmerman	2	50%	47,6%
Dezhbakhsh, Rubin et Shepherd	5	80%	5,6%*
Cloninger, Blumm et Marchesini	4	100%	1,7%
Ehrlich et Liu	4	100%	1,7%
Layson	3	100%	4,9%
Yunker	3	100%	4,9%

* 5,6% est la probabilité qu'une équipe ayant fait 5 observations ait trouvé exactement 4 résultats négatifs sur 5. La probabilité que cette équipe ait trouvé au moins 4 résultats négatifs est de 6,1%.

tats négatifs (tableau 1). Alors que le taux global de résultats négatifs est de 37%, quatre équipes ont 100% de résultats négatifs, et une cinquième en a 80%. La différence entre cette répartition et une répartition égale est assez étrange pour mériter une explication ($\chi^2 = 0,028$).

Sur ce premier tableau, on a également calculé la probabilité hypergéométrique pour qu'une équipe ait obtenu exactement le ratio de résultats négatif/positif qu'il a obtenu, en partant du principe évidemment fictif qu'il doit y avoir au final 24 résultats négatifs sur 64. On remarque un seuil à partir de 80% de résultats négatifs. Ainsi, si tous les chercheurs observaient la même chose de la même façon, la probabilité pour une même équipe d'accumuler autant de résultats favorables à la peine de mort serait très basse, respectivement 6,1%, 4,9% et 1,7%. Autrement dit, il est probable que ce soit la façon dont ces auteurs regardent (car les analyses de sensibilités montre que l'on peut, à partir des mêmes données, voir autre chose, Yang, 1998; Donohue, Wolfers, 2006; Kirchgässner, 2011) qui explique pourquoi ils ont autant de résultat négatifs.

Le profil des chercheurs en fonction de leurs résultats

Deux groupes de chercheurs. On pourrait répartir les chercheurs en deux groupes A et B selon où ils se situent par rapport à ce seuil de 80% de résultats négatifs. C'est à partir de ce seuil que la probabilité hypergéométrique d'avoir obtenu ces observations par hasard en observant de la même façon que les autres chercheurs chute drastiquement et définitivement en dessous de 6%. Avant ce seuil, seules deux équipes ont une probabilité inférieure à 9%, la plupart se situant autour des 30-40%.

Ce seuil est également intéressant car la seule équipe dans ce groupe A qui a eu un résultat positif est également par ailleurs très favorable à la peine de mort. En effet, l'interprétation des conclusions par Yang et Lester que Joanna Shepherd publie seule en 2005 et qui est le seul résultat non négatif de ce groupe est, pour l'objectif que nous nous donnons de repérer la recherche la plus favorable à la peine de mort, incomplète. Madame Shepherd calcule la corrélation exécutions-homicides dans plusieurs États américains. Si l'on fait la moyenne de tous les résultats, on trouve une légère taille d'effet positive, mais l'auteur, quant à elle, interprète ces chiffres de la façon suivante: *Mes résultats ont des implications politiques importantes: pour atteindre un niveau dissuasif, les États doivent exécuter un nombre minimal de personnes* (1). Alors que Yang et Lester notent l'auteur conclut à un *effet brutalisant*, nous remarquons que ces résultats sont interprétés par l'auteur comme confirmant également – certes avec des nuances – la thèse de l'effet dissuasif de la peine de mort. En complément, prenons l'article de vulgarisation que Joanna Shepherd publie dans *Christian Science Monitor*. Elle nous y indique: *Mon explication intuitive [de ces résultats positifs], c'est qu'on peut voir un effet dissuasif seulement quand un État exécute suffisamment de gens; dans la plupart des États, quand le nombre d'exécutions dépasse un certain stade, l'effet dissuasif dépasse l'effet brutalisant*. Il y a donc clairement un groupe de chercheurs qui obtient systématiquement des résultats favorables à la peine capitale, la seule demi-exception étant celle-ci.

Différence entre les deux groupes. Une des hypothèses récurrentes sur la disparité des résultats sur la peine de mort est que les études favorables à la peine de mort sont produites par des économistes alors que les études défavorables sont produites par des sociologues. Des économistes et des sociologues se réclament de cette opposition (Goertzel, Goertzel, 2008; Rubin, 2006; Gerritzen, Kirchgässner, 2013). Elle n'est pourtant que *partiellement correcte*. En fait les groupes A et B n'opposent pas deux groupes homogènes (« l'économie » et la « sociologie »), mais correspondent à

- (A) un *groupe homogène* formé d'économistes américains d'universités mal classées.
- (B) un *groupe hétérogène* formé des sociologues, des économistes européens et des économistes américains des universités hautement classées.

Pour parvenir à ces profils, nous avons extrait des CV de ces 28 auteurs tous les éléments disponibles. Certains, qui auraient évidemment été particulièrement intéressants pour l'étude du biais d'auteur (la source du financement des études, divers indicateurs de prises de position politiques ou idéologique) étaient très peu renseignés et ne permettaient pas de faire ressortir quelque chose de fiable avec un si petit échantillon. Mais des différences très fortes sont malgré tout apparues sur les points qui étaient presque systématiquement renseignés.

Dans le groupe A, c'est-à-dire au-delà de ce seuil des 80% de résultats favorables à la peine de mort, seule Joanna Shepherd est une femme, seule à avoir publié un résultat non négatif et la seule à avoir un poste de Droit (en fait, Joanna Shepherd est à la fois une juriste et une économiste, car ce poste succède à deux postes d'économie et à une thèse en *Law and Economics*, sa spécialité sur la page de la faculté est *Law and Economics* et elle publie régulièrement dans des revues d'économie avec des économistes). Elle est donc à plusieurs titres une exception au sein de ce groupe – même si à part ces points-là elle a un parcours relativement similaire aux autres membres du groupe. Hormis elle, le groupe A (10 chercheurs) contient 100% d'hommes économistes (poste *Economics* ou *Finance*) aux **États-Unis**. À l'opposé, le groupe B (18 chercheurs) contient 50% d'économistes parmi ceux où cette position est renseignée, et 17% de ses membres sont en poste en dehors des États-Unis (Canada et Europe).

D'autres éléments du CV diffèrent, et on a une impression d'ensemble très opposée pour les deux groupes.

D'abord, la **matière la plus enseignée** dans le groupe A est la micro-économie, c'est-à-dire la modélisation de l'agent rationnel, alors que la matière la plus enseignée par les membres du groupe B est *Crime and Punishment*, que l'on peut traduire par la criminologie ou étude du système pénal.

Ensuite, les « **spécialités** » sont très différentes dans chaque groupe. En effet, les CV débutent très souvent par un paragraphe ou une courte liste présentant la ou les spécialités de l'auteur (soit en tête du CV, soit sur la page web de l'auteur à côté du lien permettant de le télécharger). À côté de leurs 2 à 5 publications sur la peine de mort, les membres du groupe A se présentent alors dans ce paragraphe comme spécialistes de « l'économie régionale », du « développement chinois », « des pharmaceutiques », voire « de l'utilisation des réseaux de neurones dans les décisions de prêt », alors qu'on lit dans le groupe B des personnes qui se présentent comme intéressées par, outre la peine de mort, « les armes à feu », « les gangs », « les homicides multiples », « la violence », la « délinquance juvénile ».

Dans le groupe A, les travaux sur la peine de mort sont presque exclusivement le prolongement de recherches sur le consommateur rationnel. Dans le groupe B ces recherches sont soit le prolongement de recherches sur le crime et l'institution pénale, soit des exercices méthodologiques montrant la fragilité de certains travaux économétriques.

Les **publications** diffèrent également d'un groupe à l'autre. Mis à part la peine capitale, quand d'autres institutions juridiques ou plus spécifiquement pénales sont étudiées par les membres du groupe A, c'est dans les mêmes termes d'efficacité et de pouvoir dissuasif – c'est notamment le cas des armes à feu. La rationalité et la réponse individuelle au changement de norme sont la seule chose qui intéresse dans la peine de mort les membres de groupe A, qui ne publient par exemple jamais de travaux sur les inégalités et les injustices dans le fonctionnement de la justice pénale, alors que c'est un thème majeur des membres du groupe B. Cette absence d'intérêt pour les inégalités est d'ailleurs étonnante : la peine de mort ayant en pratique une application très inégale, on pourrait s'attendre à ce qu'un chercheur intéressé par la question de son effet dissuasif s'interroge, au passage, sur ce que cette inégale application peut avoir comme impact sur la criminalité.

L'élément qui est le plus systématiquement renseigné et que l'on découvre avec surprise être un très bon prédicteur des résultats d'un travail sur la peine de mort est **le classement de l'université du dernier poste de l'auteur**. En effet, les universités sont classées par plusieurs systèmes qui donnent en général des résultats convergents – nous prenons ici le classement du *Times Higher Education* 2013-2014⁴. Harvard, Yale, le MIT, l'University of Chicago, de Stanford sont en haut. La Western Illinois University, par exemple, n'est pas classée et l'Emory University est en milieu de tableau. Ce classement se fonde sur une série de critères choisis par le *Times* : rigueur du niveau d'admission des étudiants, budget par étudiant, production scientifique du personnel. Sans avoir à discuter de la teneur de ces critères pour distinguer les « bonnes » des « mauvaises » universités, il n'est pas controversé que ce classement reflète à peu près, aux États-Unis, les universités socialement perçues comme prestigieuses ou non, à la fois dans le monde académique et dans le monde étudiant. Or, il est *a priori* très surprenant que ce classement soit fortement corrélé aux résultats sur la peine de mort produits par les chercheurs qui y sont en poste.

Cette corrélation s'accroît drastiquement si l'on s'intéresse aux équipes exclusivement composées d'économistes (*Economics, Finance, Law and Economics*), avec un coefficient de corrélation R^2 très élevé à 0,55. Ces résultats sont illustrés aux graphiques 6 et 7.

Le tableau 2 permet de se faire une idée plus précise de ces sous-groupes, en comparant notamment le prédicteur important qu'est le classement de l'université du dernier poste avec un autre critère, le classement de l'université où chaque auteur a soutenu sa thèse, qui, lui, n'est pas un prédicteur. En effet, Il n'y a pas tellement de différence entre le niveau des « universités de départ » des chercheurs : 8 des 10 économistes du groupe A ont fait leur thèse dans le top 200, comme 6 des 8 économistes du Groupe B et 12 des 18 chercheurs du groupe B en général. En revanche, il y a une importante différence au niveau

4 Pour le classement d'un individu nous avons pris le classement 2013-2014 du *Times Higher Education* de l'université du dernier poste qu'il a occupé, pour le classement d'une équipe nous avons pris l'université la mieux classée parmi les derniers postes occupés par les membres de l'équipe. Il s'agit d'un classement du top 400. Pour les universités non classées, elles sont codées comme 401^e. Pour les universités entre 200 et 400, le classement *Times of Higher Education* donne une position à 25 places près (par exemple 301-350), et nous avons donc pris la position moyenne de la fourchette (dans ce cas 325,5).

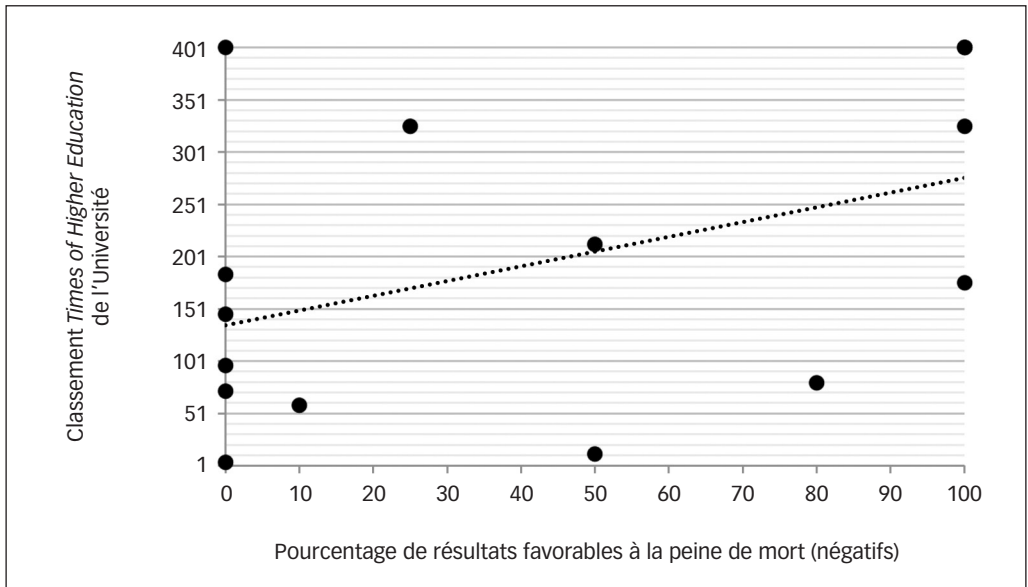
des « universités d'arrivée ». La présence de trois économistes du groupe B dans le top 50 (en fait, dans le top 15) alors qu'aucun économiste du groupe A ne s'y trouve est statistiquement significative (*test de Student*). Ce résultat est néanmoins relativement fragile du fait de la petitesse de l'échantillon⁵.

On peut quantifier l'ascendance ou la descendance d'un chercheur entre l'université d'arrivée et l'université de départ par un « indice d'ascendance », sous la forme d'un coefficient de type $\frac{\text{type rang de l'université d'arrivée}}{\text{rang de l'université de départ}}$.

Plus ce taux coefficient est bas, mieux l'université d'arrivée est classée par rapport à l'université de thèse. Un coefficient de 1 signifie que le chercheur enseigne dans une université aussi bien classée que celle de son directeur de thèse, un coefficient de 3 qu'il est en poste dans une université 3 fois moins bien classée, et un coefficient de 0,5 qu'il est dans une université deux fois mieux classée. Pour les membres du groupe A, ce coefficient atteint 10,5 alors que pour les seuls économistes du groupe B il chute de moitié à 5,6.

Ainsi, ce ne sont pas « les économistes » qui sont favorables à la peine de mort alors que les sociologues y sont défavorables, mais plutôt un certain type d'économiste. Les travaux les plus prudents sur la peine de mort sont souvent le fait des économistes les plus prestigieux. John B. Taylor par exemple, de l'université de Stanford, est un nobélisable, dont les théories sur les taux d'intérêts sont utilisées par de nombreuses banques centrales dans le monde entier. Edward Leamer, l'inventeur du test de sensibilité (*extreme bound*

Graphique 7 : Corrélation entre le classement de l'université d'une équipe et le pourcentage de résultats négatifs ($R^2 = 0,18$)



⁵ Si l'on s'intéresse aux postes actuels au lieu du dernier poste occupé, et qu'ainsi l'on ne compte pas ceux qui sont donc sortis du monde universitaire, les économistes du groupe B perdent un poste dans le top 50 et le résultat perd de sa mesurabilité.

Graphique 8 : Corrélation entre le classement de l'université d'une équipe d'économistes et le pourcentage de résultats négatifs ($R^2 = 0,55$)

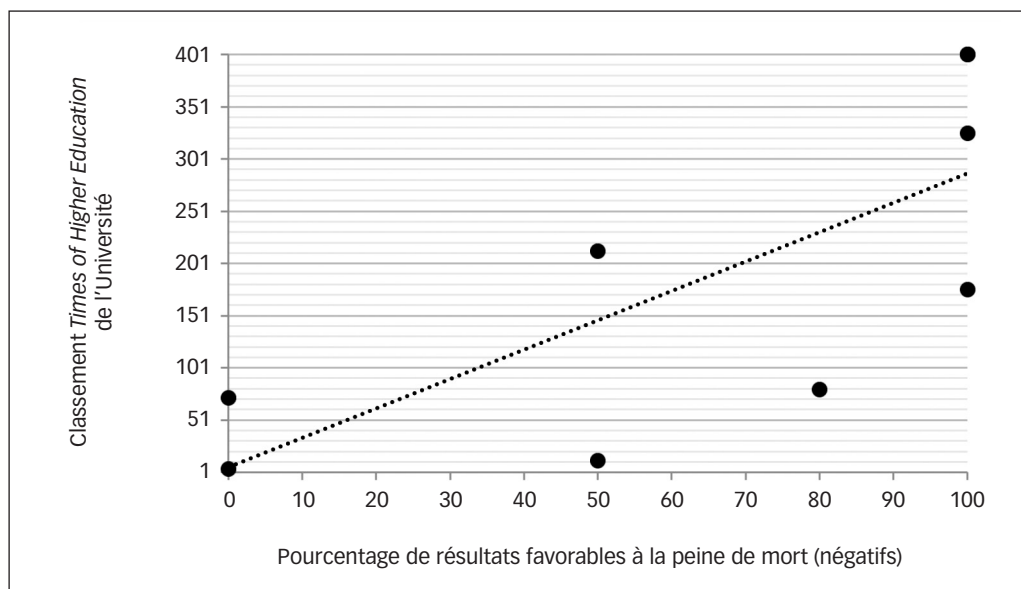


Tableau 2 : Comparaison entre les niveaux d'université de départ et d'arrivée du groupe A et B

	Tous	Groupe A	Économistes du groupe B	Autres membres du Groupe B
Total	28	10	8	10
Thèse dans le top 200	20	8	6	6
Thèse dans le top 100	20	8	6	6
Thèse dans le top 50	13	4	5	4
Poste dans le top 200	16	5	5	6
Poste dans le top 100	10	3	5	2
Poste dans le top 50	3	0*	3*	0*
Indice d'ascendance	9,50	10,42	5,67	11,64

* Signifie $p < 0.05$

analysis) et professeur à UCLA, est un des grands noms de l'épistémologie en économétrie. C'est lui qui a, quelques années après la parution de la première étude d'Isaac Ehrlich, mis les chercheurs en garde sur les possibilités de *data mining* qu'offrait la spécification *ad hoc* des régressions en matière de peine de mort, dans un article cité plus de 1800 fois selon *Google Scholar* intitulé avec humour *Let's take the Con out of Econometrics* (1982). Peter Passel, avant de rejoindre le *think-tank* du *Milken Institute*, était professeur à Columbia et membre du staff économique du *New York Times*. Autrement dit, chacun à sa manière, ces trois économistes américains ont obtenu des reconnaissances importantes et ont enseigné dans les facultés les plus hautement regardées des États-Unis.

À l'opposé, le parcours typique d'un économiste du groupe A est descendant : il s'agit d'un chercheur dont le dernier poste est généralement dans un lieu bien moins prestigieux que celui où il a lui-même étudié. Le parcours de ces chercheurs n'est pourtant pas différent en début de carrière : le point de départ de plusieurs de ces auteurs est à peu près aussi souvent une grande université. Mais, alors que la majorité des économistes du Groupe B qui ont soutenu leur thèse dans le top 50 ont par la suite enseigné dans une université du même niveau, Issac Erlich passe de *Columbia* (13) à l'*University of Buffalo* (178), Stephen Layson de l'*University of Chicago* (9) à l'*University of North Carolina* (301-350), Roberto Marchesini de l'*University of Texas at Austin* (27) à l'*University of Houston* (non classée) et James Yunker de la *Northwestern University* (22) à la *Western Illinois University* (non classée). Les résultats sur la peine de mort sont corrélés à l'université d'arrivée, et non à celle de départ. Ce résultat est d'autant plus frappant que même à l'intérieur du groupe A, on retrouve cette division. En effet, les économistes les mieux classés au sein de ce groupe (ceux qui ont abouti dans *Emory University*, une université méthodiste classée 80^e), ont des résultats favorables à la peine de mort, mais moins extrêmes. C'est ce qui donne au graphique 6 un aspect tellement propre : les résultats les plus proches de zéro émanent de meilleures universités, les plus proches de -1 des moins bien classées et ceux qui sont entre les deux émanent d'universités en milieu de tableau, d'où le très fort coefficient R^2 , qui nous indique que malgré la petitesse de l'échantillon, cette piste est intéressante. On retrouve donc au sein même des économistes favorables à la peine de mort une même subdivision entre ceux qui y sont systématiquement favorables et ceux qui sont un peu plus nuancés, subdivision qui se fait selon le même critère (rang de l'université) – ce qui correspond à la forme de fractale identifiée par Andrew Abbott dans son histoire des sciences sociales (2001).

Les travaux systématiquement favorables à la peine de mort ont donc pour caractéristique notable d'être produits par les universitaires ayant la position la plus descendante entre leur université de doctorat et leur poste le plus récent. Ce fait est étrange puisque les professeurs qui les ont accompagnés sur leurs pistes de recherches étaient, eux, dans des universités hautement classées, mais est-ce un hasard si certains ont pris de la distance sur ces résultats ? Ainsi, Gary Becker, prix Nobel, éminent professeur d'économie à Chicago, celui qui fut à l'initiative de ce mouvement sur la rationalité criminelle et le mentor de certains des économistes du Groupe A – n'a jamais personnellement produit de telle étude, et il est bien plus prudent que ses élèves sur leurs conclusions. Alors que les membres du Groupe A présentent l'effet dissuasif de la peine de mort comme « essentiellement » ou « unanimement » établi par la littérature empirique qu'ils ont produite, Gary Becker écrivait sur son blog il y a quelques années : *il ne fait guère de doute que ces travaux empiriques quantitatifs donnent des résultats mixtes*, et ajoutait que s'il était toujours personnellement convaincu de l'effet dissuasif de la peine de mort, c'était également pour des raisons philosophiques qui allaient au-delà de ce qui peut être démontré par ces études. Dans une conférence tenue peu de temps avant son décès, Gary Becker indiquait à nouveau que *les résultats de ces travaux [sur la peine de mort] sont mixtes*, et qu'il préférerait que ce sujet ne monopolise pas trop la discussion de ses théories (Becker, Ewald, Harcourt, 2013).

Il faut évidemment être prudent vis-à-vis de l'interprétation de ces données. Tout d'abord, nous parlons ici d'un petit échantillon de chercheurs (28 au total) sur un petit domaine (65 publications). Néanmoins, cette découverte ouvre toute une série de pistes de recherches, que l'on pourrait sans doute approfondir en répliquant cette méthodologie dans des champs

plus larges. Par exemple, on peut se demander en quoi le parcours de vie des chercheurs et leur appartenance à des sous-communautés scientifiques spécifiques influencent les résultats qu'ils trouvent. On peut également se demander si certains résultats ne peuvent être obtenus que par des travaux de moindre qualité qui, s'ils parviennent à être publiés dans des revues respectées – nous reviendrons sur ce point –, ne passent pas au crible d'un comité de recrutement dans une grande université, peut-être parce que cette recherche a justement la réputation de n'être pas sérieuse depuis que les travaux d'Isaac Ehrlich ont été disséqués dans les années 1970. On peut également rechercher dans quelle mesure il peut y avoir un biais de recrutement des universités américaines, ou encore une corrélation entre certains résultats économétriques et certaines opinions politiques. Il existe en bref une discussion à avoir sur les liens entre les marqueurs du « prestige » académique – qui ont fait par ailleurs l'objet de plusieurs théorisations marquantes (Bourdieu, 1984; Latour, Woogar, 2006) – et les résultats empiriques sur un domaine donné. Une autre piste est de faire le lien entre les résultats obtenus et la diversité des cultures académiques.

Ce qui est certain, c'est que les grandes tendances que nous avons mises en lumière s'aggravent au fur et à mesure que l'on donne davantage de contrôle aux scientifiques sur leurs résultats, c'est-à-dire que l'on s'intéresse aux conclusions plutôt qu'aux données brutes. Si les tailles d'effet des travaux sur la peine de mort sont généralement corrélées aux tailles d'effets des travaux précédents du même chercheur et au niveau de son université d'arrivée, cette corrélation s'accroît pour les tailles d'effet corrigées et s'accroît encore pour les seules études comprenant des variables contrôlées.

Conclusion

Que faire de ces résultats? Une première question est de savoir quels sont les outils dont le monde scientifique dispose afin d'éviter que des « biais d'auteurs » ne polluent les champs d'analyse. Il y a tout d'abord le niveau des revues – mais il ne permet pas de solutionner ce problème ici. Si les auteurs des résultats les plus favorables à la peine de mort enseignent dans les facultés les moins prestigieuses, cela ne signifie pas pour autant que ces travaux soient publiés dans des *revues grises*. Il s'agit d'*American Economic Review*, *Journal of Criminal Justice* ou encore d'*American Law and Economics Review*. Néanmoins, il est vrai que l'absence de publicité des bases de données rend le travail critique *a posteriori* difficile à effectuer sans une collaboration des auteurs – ce qui est gênant dans le cadre de travaux fortement mathématisés et opaques.

Pour ce qui est des méta-analyses, il y a depuis plus d'une décennie un certain nombre d'initiatives pour permettre que ces analyses systématiques d'un champ ne soient pas polluées par les biais de publication. Un auteur proposa une série de tests pour mettre en lumière les résultats statistiquement anormaux dès la fin des années 1990 et ce afin de détecter la tricherie ou la rétention de données dans les recherches financées par l'industrie pharmaceutique (Egger *et al.*, [1997]). On peut regretter que ce test ne soit pas suffisamment employé, car il est utile en tout domaine. La *Campbell corporation* propose une méthodologie poussée d'exclusion des études fragiles ou suspectes. Jean-Claude Combessie montre, au sujet des inégalités de destin, que l'on peut à minima présenter tout nouveau résultat en fonction de chaque méthode d'analyse disponible afin de ne pas masquer pour le lecteur

la pluralité des choix méthodologiques qui peuvent en aval donner des résultats différents (Combessie [2011]). Suite à notre travail, on pourrait rajouter un test « résultat par réseau d'auteur » dans chaque méta-analyse afin de mettre en lumière des groupes de recherches qui produisent systématiquement les mêmes résultats – que ce biais soit lié à des financements, à une méthodologie, à une idéologie ou bien qu'il soit réductible à une question de sous-domaine de recherche. On pourrait de cette façon prendre acte du fait que la « vérité scientifique » ne saurait être une « moyenne » des résultats obtenus dans diverses communautés.

Enfin, un dernier élément est porteur d'interrogations. De Sellin (1959) à la comparaison de Hong-Kong et Singapour (Zimring, Fagan, Johnson, 2010), l'évolution des taux d'homicide dans les juridictions démographiquement proches mais ayant des politiques sur la peine de mort drastiquement différentes a toujours suggéré un effet zéro des exécutions sur les homicides. Autrement dit, quand deux zones géographiques ont un taux d'homicide qui évolue en parallèle et que l'une d'elles change drastiquement sa politique sur la peine de mort, cela n'empêche pas les taux d'homicide de continuer à évoluer en parallèle par la suite. On pourrait se demander si c'est par coïncidence que ce résultat brut et pour lequel il n'est pas besoin de mathématiques sophistiquées correspond également à la recherche majoritaire, hétérogène ainsi qu'à celle produite dans les universités les plus prestigieuses alors que la recherche quantitative qui parvient à renverser ces données brutes est minoritaire, produite par un petit groupe très homogène et marginalisé dans l'université.

La prochaine étape de notre recherche est sans doute d'appliquer ces mêmes méthodes de méta-analyses désanonymisées et d'identification de biais d'auteurs et profils de chercheurs à d'autres champs afin de voir s'il se dessine des régularités ou si les recherches sur la peine de mort sont plutôt l'exception que la règle. Un champ plus large aura l'avantage d'offrir des résultats plus robustes et moins sensibles au hasard et à l'erreur.

Sacha Raoult

Bibliographie

- ABBOTT A., 2001, *Chaos of disciplines*, Chicago, University of Chicago Press.
- BARNES B., BLOOR D., HENRY J. (Eds), 1996, *Scientific knowledge: A sociological analysis*, Chicago, University of Chicago Press.
- BECCARIA C., 1870, *Des Délits et des peines*, Paris, Guillaumin et Cie.
- BECKER G S., EWALD F., HARCOURT B.E., 2013, Becker and Foucault on Crime and Punishment, *University of Chicago Coase-Sandor Institute for Law & Economics Research Paper*, 654.
- BEKELMAN J.E., LI Y., GROSS C.P., 2003, Scope and impact of financial conflicts of interest in biomedical research: a systematic review, *Jama*, 289, 4 454-465.
- BOURDIEU P., 1984, *Homo academicus*, Paris, Éditions de Minuit.
- BRESNICK S.D., 2008, Does Extensive Financial Support Create Author Bias? The Juvéderm Injectable Gel Study, *Aesthetic Surgery Journal*, 28, 5, 596-597.
- COMBESSIE J.-C., 2011, Analyse critique d'une histoire des traitements statistiques des inégalités de destin, *Actes de la recherche en sciences sociales*, 3, 4-31.
- DÖLLING D., ENTORF H., HERMANN D., RUPP T., 2009, Is deterrence effective? Results of a meta-analysis of punishment, *European Journal on Criminal Policy and Research*, 15, 1-2, 201-224.

- DONOHUE J.J., WOLFERS J.J., 2006, Uses and abuses of empirical evidence in the death penalty debate, *Stanford Law Review*, 58, 1, (791-846).
- DONOHUE J.J., AYRES I., 2009, More guns, less crime fails again: the latest evidence from 1977 –2006, *Economic Journal Watch*, 6, 2, 218-238.
- DURLAUF S.N., FU C., NAVARRO, S., 2013, Capital punishment and deterrence: understanding disparate results, *Journal of Quantitative Criminology*, 29, 1, 103-121.
- EGGER M., DAVEY SMITH G., SCHNEIDER M., MINDER C., 1997, Bias in meta-analysis detected by a simple, graphical test, *Biomedical Journal*, 315, 7109, 629-634.
- EHRlich I., 1975, The Deterrent Effect of Capital Punishment: A Question of Life and Death, *The American Economic Review*, 65, 3, 397-417.
- FOUCAULT M., 2014, *Les mots et les choses. Une archéologie des sciences humaines*, Paris, Gallimard.
- GERRITZEN B., KIRCHGÄSSNER G., 2013, Facts or Ideology: What Determines the Results of Econometric Estimates of the Deterrence Effect of Death Penalty? A Meta-Analysis, document de travail disponible sur en prépublication sur Social Science Research Network SSRN.
- GIDE A., 2012, *Les Nourritures terrestres/Les Nouvelles nourritures*, Paris, Gallimard.
- GOERTZEL T., GOERTZEL B., 2008, Capital punishment and homicide rates: sociological realities and econometric distortions, *Critical Sociology*, 34, 2, 239-254.
- HELMUS L., HANSON R.K., THORNTON D., BABCHISHIN K.M., HARRIS A. Jr., 2012, 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, 9, 1148-1171.
- JONSON C.L., 2010, *The impact of imprisonment on reoffending: A meta-analysis*, Diss., Cincinnati, University of Cincinnati.
- KIRCHGÄSSNER G., 2011, Econometric estimates of deterrence of the death penalty: Facts or ideology?, *Kyklos*, 64, 3, 448-478.
- KUHN T.S., 1983, *La structure des révolutions scientifiques*, Paris, Flammarion.
- LATOUB B., WOOLGAR S., 2006, *La vie de laboratoire. La production des faits scientifiques*, Paris, La découverte.
- LEAMER E.E., 1983, Let's take the con out of econometrics, *The American Economic Review*, 73, 1, 31-43.
- LIPSEY M.W., CULLEN FT., 2007, The effectiveness of correctional rehabilitation: A review of systematic reviews, *Annual Review of Law and Social Science*, 3, 297-320.
- LIU Z., 2004, Capital punishment and the deterrence hypothesis: some new insights and empirical evidence, *Eastern Economic Journal*, 30, 2, 237-258.
- MCMANUS W.S., 1985, Estimates of the Deterrent Effect of Capital Punishment: The Importance of the Researcher's Prior Beliefs, *The Journal of Political Economy*, 93, 2, 417-425.
- PATERNOSTER R., 2010, How much do we really know about criminal deterrence, *Journal of Criminal Law & Criminology*, 100, 765.
- PRATT T.C., CULLEN ET., 2005, Assessing macro-level predictors and theories of crime: A meta-analysis, *Crime and Justice*, 373-450.
- RUBIN P., 2006, *Statistical Evidence on Capital Punishment and the Deterrence of Homicide*, Written testimony for the Senate Judiciary Committee on the Constitution, Civil Rights, and Property Rights February 1.
- SELLIN T., 1959, *The death penalty*, Philadelphia, American Law Institute.
- THOMPSON D.F., 1993, Understanding financial conflicts of interest, *New England Journal of Medicine*, 329, 573-573.
- VACCARO A R., PATEL A.A., FISHER C., 2011), Author conflict and bias in research: quantifying the downgrade in methodology, *Spine*, 36, 14, E895-E896.
- YANG, B., LESTER D., 2008, The deterrent effect of executions: A meta-analysis thirty years after Ehrlich, *Journal of Criminal Justice*, 36, 5, 453-460.
- ZIMRING F.E., FAGAN J., JOHNSON D.T., 2010, Executions, deterrence, and homicide: a tale of two cities, *Journal of Empirical Legal Studies*, 7, 1, 1-29.