

**L'expérience de l'expérimentation
pour l'étude des inégalités et des discriminations**

Workshop organisé par l'axe « Discriminations et inégalités sociales » du LIEPP

31 mai 2018, Sciences Po, Paris



Compte-rendu du workshop

Ce workshop visait à susciter un échange autour de plusieurs expériences de recherches récentes qui utilisent des méthodes expérimentales ou quasi-expérimentales pour étudier les inégalités et les discriminations. Les intervenant.e.s sont revenu.e.s sur les différents types de méthodes utilisées (expérimentation contrôlée, expérience naturelle, testing...), le contexte institutionnel et académique de leur mobilisation, les difficultés rencontrées et les solutions apportées. Le débat a également porté sur les apports de ces approches méthodologiques dans la compréhension des mécanismes de reproduction et de transformation des inégalités sociales ainsi que leurs implications en termes de politiques publiques.

Intervenant.e.s:

- [Anne Boring](#), Erasmus University Rotterdam / Sciences Po (LIEPP/PRESAGE).

Document de support: [Reducing Discrimination Through Norms or Information: Evidence from a Field Experiment on Student Evaluations of Teaching](#), *TSE Working Paper*, 17-865 (2017), with Arnaud Philippe.

- [Carlo Barone](#), Sciences Po (OSC/LIEPP)

Document de support: [Social origins, relative risk aversion and track choice: A field experiment on the role of information biases](#), *Acta Sociologica*, First Published October 8, 2017, with Giulia Assirelli, Giovanni Abbiati, Gianluca Argentin and Deborah De Luca.

- [Jean-Benoît Eyméoud](#), Banque de France / Sciences Po (LIEPP/Département d'économie) et [Paul Vertier](#), Sciences Po (LIEPP/Département d'économie).

Document de support: [Gender Biases: Evidence from a Natural Experiment in French Local Elections](#) *LIEPP Working Paper* n°78, April, 2018

- [Nicolas Jacquemet](#), Université Paris 1 / PSE.

Documents de support:

Edo A., Jacquemet. N. (2013). [La discrimination à l'embauche, Sur le marché du travail Français](#). *Opuscule du CEPREMAP n°31*, Editions rue d'Ulm

Edo A., Jacquemet. N. (2014). [Discrimination à l'embauche selon l'origine et le genre : défiance indifférenciée ou ciblée sur certains groupes?](#) *Economie et Statistiques*, n°464-465-466, pp.155-172

Edo A., Jacquemet. N., Yannelis C. [Language Skills and Homophilous Hiring Discrimination: Evidence from Gender- and Racially-Differentiated Applications](#). *Review of Economics of the Household*

1- Présentation de la méthode et de son usage

Dans cette première partie sont présentées les démarches expérimentales mises en oeuvre par les chercheurs, tant dans leurs grands principes que dans leurs applications.

Jean-Benoît Eymeoud et Paul Vertier, économistes, inaugurent ce premier temps en présentant la démarche qu'ils ont utilisée dans leur article « *Gender Biases: Evidence from a Natural Experiment in French Local Elections* ». Essentiellement empirique, celle-ci repose sur la méthode des expériences naturelles. Cette méthode consiste à tirer profit de contextes historiques qui s'apparentent à des chocs exogènes pour identifier une relation causale.

Via cette méthode, l'article s'attache à étudier les discriminations en fonction du genre, et les élections départementales de 2015 constituent un cadre d'analyse empirique parfait. En effet, pour la première fois étaient soumis au vote des binômes aux mêmes responsabilités, et formés de manière paritaire. Détail important : l'ordre d'apparition sur le bulletin de vote était déterminé par l'ordre alphabétique des noms des deux colistiers. Dès lors, femmes comme hommes étaient aléatoirement distribués entre première et deuxième place sur les bulletins, permettant la mobilisation de ces élections comme une expérience naturelle, le nom de famille étant une donnée non-manipulable et exogène.

Basée sur un échantillon de 9 000 binômes testés, l'étude se propose de répondre à la question suivante : les binômes dont la femme est en première position reçoivent-ils moins de votes que les autres ? Une absence de différence de résultat signifierait une absence de discrimination en fonction du genre. Chacune des grandes familles politiques (droite, gauche, extrême-gauche, extrême-droite) a été prise en compte, et l'analyse montre une notable divergence selon cette variable. Là où les différences sont négligeables pour les trois autres, la droite se démarque par le fait que ses binômes ayant pour premier nom celui d'une femme subissent une perte de 1,5 points en moyenne par rapport aux autres, figurant un effet extrêmement robuste. Un deuxième effet en découle, plus décisif par rapport aux résultats des élections : en plus d'avoir un impact brut sur le résultat en soi, une telle perte implique que la probabilité pour ces binômes d'accéder au second tour est diminuée de 4 points de pourcentage.

Une dernière question reste de savoir lequel des deux types de discrimination, à savoir discrimination statistique (basée sur un signal informationnel envoyé) et *taste-based discrimination* (basée sur les goûts), est à l'œuvre. La prise en compte de la PCS et de la position politique antérieure -élu.e ou pas ?- fait disparaître les différences. Cela suggère que c'est plutôt le premier type de discrimination qui est à l'œuvre, et que les discriminations proviennent plutôt de certains *a priori* par rapport à ce que se doit d'être un élu.

Anne Boring a présenté son étude des biais de genre chez les étudiants, notamment garçons, lorsque ceux-ci tendent à mieux noter leurs enseignants hommes au cours des évaluations de fin de semestre. Dans ce travail mené conjointement avec Arnaud Philippe, l'objectif est de traiter un angle mort de la recherche, celui d'évaluer comment réduire une telle discrimination dans ce contexte naturel.

Son terrain a été une cohorte d'étudiants à Sciences Po Paris, répartie sur sept campus considérés ici comme *a priori* peu connectés entre eux. Deux messages ont été testés, chacun sur des campus différents. Le premier, à caractère normatif, enjoignait simplement à ne pas discriminer. Le second accolait à cette même idée des liens vers un graphique montrant qu'il y avait effectivement de la discrimination, et vers un article antérieur d'Anne Boring. Les étudiants de trois campus ont reçu le premier message dans leurs boîtes mail, tandis que ceux de deux autres ont reçu le second. Les étudiants des deux derniers campus n'ont reçu aucun message, faisant office de groupe témoin. Au sein même des campus, la moitié des élèves ont reçu le mail après avoir procédé à leurs évaluations, doublant ainsi le contrôle de l'expérience. Malgré cette double précaution, les effets de

contamination -potentiellement dus à la transmission des informations du mail à des individus faisant partie d'un groupe contrôlé- ont été redoutés, à la fois entre les campus et en leur sein.

Un résultat reste tout de même frappant : l'e-mail purement normatif n'a eu aucun effet, quand il n'a pas paradoxalement augmenté la discrimination -certes de manière faiblement significative-, élément à mettre en lien avec les conclusions de la littérature sur les politiques de diversité en entreprise. Le second mail, quant à lui, a plutôt bien fonctionné, en réduisant le gap.

Nicolas Jacquemet, quant à lui, s'est attaché à non seulement mesurer les discriminations à l'embauche, mais aussi à définir la nature qu'elles revêtent ; à savoir, si elles sont tournées vers certaines catégories de populations, ou si elles sont plutôt mues par une forme d'entre-soi de la part des employeurs. Dans le cadre de la méthode d'évaluation par correspondance -plus connue sous son nom anglais de *testing*-, il cherche à comprendre si l'information dont dispose l'employeur sur les compétences des candidats a une influence sur la discrimination.

De fausses candidatures ont donc été envoyées en réponse à de vraies offres d'emploi, en vue de rendre compte des différences de traitement que les candidats peuvent subir selon telle ou telle caractéristique. Toutefois, un problème méthodologique se pose : comment distinguer l'influence réelle du paramètre testé, et d'autres différences réelles possibles entre candidats, comme la réputation de leur établissement de diplomation ou l'attractivité de leur lieu de résidence ?

Les CV sont construits de manière cohérente par rapport aux offres d'emploi, avec des informations telles que le nom, le prénom... disposées de manière à évoquer une origine ethnique, et un genre -les deux dimensions ici prises en compte dans la mesure des discriminations. Ces noms sont associés aux CV, des offres d'emploi sont recueillies, et ces candidatures montées de toutes pièces sont envoyées, pour finalement mesurer combien de fois chaque candidat a été convoqué en entretien d'embauche.

La principale difficulté méthodologique à surmonter est qu'il est nécessaire que les différences entre CV ne soient pas les raisons de convocations en entretien d'embauche, et en même temps que les CV ne soient pas les mêmes. Cette contradiction est surmontée en mettant en place une rotation systématique entre CV et candidat. Dès lors, si une différence subsiste systématiquement entre candidats, elle ne peut être attribuée qu'à l'origine ethnique perçue par l'employeur, et à aucune des autres informations fournies.

Mais qu'en est-il de la mise en adéquation des identités que le chercheur souhaite transmettre et évoquer en associant un certain nom et prénom, et celles qui seront effectivement perçues par les employeurs comme véhiculant un certain groupe d'appartenance ? En d'autres termes, comment s'assurer que les identités soient définies de manière claire et homogène par l'ensemble des employeurs ? 50 noms et prénoms ont été présentés à 400 répondants par les chercheurs lors d'une étude préalable, en vue de sélectionner ceux qui étaient le plus unanimement associés à une liste de catégories de population.

Pour finir, **Carlo Barone**, seul sociologue de notre groupe d'intervenants, a commencé par insister sur le fait les méthodes d'expérimentations sont utiles aussi pour les sociologues.

Son usage de l'expérimentation s'inscrit dans une réflexion sur l'impact des inégalités d'accès à l'information sur les inégalités éducatives. En effet, les familles les plus modestes ont tendance à sous-estimer les risques de sortie précoce de leurs enfants du système éducatif (notamment arrêt des études avant l'enseignement supérieur), là où les plus aisées s'y préparent et développent des stratégies palliatives. Dès lors, la question se pose de savoir si, d'un point de vue de politiques publiques, aplanir la quantité et la qualité de l'information donnée aux différentes familles, quel que soit leur niveau socio-économique, peut être un bon moyen de lutter contre ces inégalités.

Carlo Barone pointe les limites, à cet égard, des conclusions qui peuvent être tirées des enquêtes quantitatives classiques, du fait de la difficulté d'isoler l'impact propre du paramètre informationnel, fortement corrélé à de nombreuses autres variables. Face à cette difficulté d'imputation causale, l'expérimentation de terrain (*field experiment*) apparaît alors comme une méthode particulièrement pertinente. Carlo Barone en a souvent fait usage. L'expérimentation sur

laquelle s'appuie sa présentation pour ce workshop a porté sur l'orientation des lycéens en Italie. Deux groupes ont été constitués en donnant un accès aléatoire à l'information à un panel de familles. Toute différence après ce traitement peut donc être attribué à une variable spécifique.

Le résultat de cette expérimentation fut une augmentation de 11 % de poursuites dans l'enseignement supérieur pour le groupe qui a eu le plus d'informations, validant la thèse d'une importance considérable de l'information dans les inégalités de parcours éducatif. Carlo Barone insiste sur le fait que ce type d'enquête revêt une validité externe supérieure à celle des expériences en laboratoire ou à l'analyse statistique classique à partir de bases de données.

2- Difficultés rencontrées et solutions apportées

Cette partie traite des difficultés méthodologiques rencontrées au cours des différents travaux, et de la manière dont les chercheurs y ont fait face. Les problèmes éthiques et pratiques ont notamment été évoqués.

Paul Vertier et Jean-Benoit Eyméoud évoquent comme principal problème le fait qu'il faille prouver le caractère aléatoire de cette expérience. En effet, comment s'assurer qu'il n'y ait pas eu de manipulation de l'ordre d'apparition, avec par exemple des appareils de parti ayant anticipé des effets du genre sur le résultat, et qui auraient donc veillé à ne pas mettre de femmes en première position ? Pour ce faire, des données exhaustives renseignant le plus de caractéristiques possibles des candidats sont essentielles, pour voir si l'ordre d'apparition est conditionné par celles-ci.

Certaines de ces données, collectées par le ministère de l'Intérieur, sont disponibles sur datagouv.fr : Catégorie socio-professionnelle, âge et expérience politique. Il est dès lors possible de voir si l'ordre d'apparition des élus sur le bulletin dépendait des caractéristiques croisées des élu.e.s mentionné.e.s, et de mettre en évidence que les partis de droite notamment mettent ensemble des élu.e.s expérimenté.e.s, et ensemble des élu.e.s non-expérimenté.e.s. Toutefois, rien ne semble indiquer que le taux d'expérience respectif des colistiers soit déterminant dans l'ordre d'apparition sur le bulletin.

De la même manière, il apparaît que l'ordre d'apparition des premières lettres est également réparti chez les hommes et les femmes. Dès lors, quelles sont les variables à inclure, et quelles sont les données à considérer ? Dépenses de campagne, origine ethnique, sont les éléments testés en ce moment, donnant *a priori* les mêmes résultats que précédemment.

Une autre piste explorée est de regarder du côté des biais cognitifs possiblement présents chez les électeurs, du fait d'une attention limitée ou d'un manque de compréhension des règles du scrutin. Si l'électeur comprend qu'il n'y a pas de différence entre deux noms sur un bulletin, alors il ne devrait y avoir aucune différence de vote selon le nom situé en première position. À l'inverse, s'il perçoit une différence, alors ce critère aura un impact sur son vote. La question de savoir si la compréhension de l'électorat est homogène ou non est donc primordiale, et peut être résolue en regardant les différentiels de communication faites autour du scrutin entre les cantons.

Regarder le format du bulletin de vote peut aussi fournir une piste d'explication : une disposition horizontale ou verticale peut-elle influencer le nombre de votes ? De plus, dans quelle mesure peut-on étendre l'étude aux autres informations disponibles sur le bulletin de vote, telles que la photographie ou l'âge des candidats ? Telles sont quelques-unes de réflexions en cours en vue du prolongement de cette recherche.

Le travail d'**Anne Boring**, en traitant des étudiants sans que ceux-ci en aient été informés, lui a surtout posé des problèmes éthiques. Ces réticences ont été contrebalancées d'une part par le soutien de l'administration de Sciences Po, qui lui a permis l'accès aux données, et d'autre part par la validation du protocole expérimental par un certificat CITI délivré par le MIT ainsi que par un IRB de Jpal, cette dernière institution arguant que l'expérience ne causerait de dommages à qui que ce soit. Ce dernier point reste vrai dans la mesure où la discrimination ne serait pas augmentée par l'expérimentation, ce qui a été le cas finalement. L'administration en a été avisée en amont, pour

veiller à ce qu'elle ne pénalise pas les enseignantes concernées. En outre, l'approbation de Jpal a permis d'empêcher de faire du data-mining après l'opération, tant le protocole devait être clair et contrôlé.

Un autre problème, plus méthodologique celui-ci, est dû à la différence naturelle existant entre groupes de contrôle et groupes testés, le test ayant été commencé après qu'un certain nombre d'étudiants -constituant le groupe de contrôle- aient déjà faits leurs évaluations pédagogiques. Dans la même veine que cette première considération, aucun contrôle n'est possible par le temps, l'avant étant le groupe de contrôle, et les professeurs n'étant évalués qu'une seule fois.

L'homogénéité des campus peut elle aussi être mise en doute, d'autant plus que des effets de contamination peuvent avoir eu lieu. Un groupe féministe sur l'un des campus a notamment relayé l'information de la transmission de ce mail à quelques étudiants, faussant potentiellement et dans une certaine mesure l'expérience.

Enfin, un point sur lequel moins d'attention a été accordée reste celui de savoir si les individus ont effectivement été traités. À ce sujet, une sorte d'accusé de réception adjoint au mail n'aurait peut-être pas été superflu pour affiner un peu plus encore les résultats et contrôler la viabilité de l'expérimentation.

Nicolas Jacquemet présente l'expérience qu'il a menée comme très peu invasive ; elle a donc posé peu de problèmes éthiques. Les complications sont surtout apparues lors de sa mise en place, autour de l'importance scientifique qu'il y avait de définir la nature de la discrimination parmi les deux types évoqués dans la première partie. En vue de tester la manière dont sont discriminés des groupes n'appartenant explicitement pas au groupe majoritaire, deux candidatures étrangères ont à chaque fois été testées : une subissant effectivement des discriminations -et rattachées au groupe des prénoms maghrébins-, et une perçue comme non-française tout en n'étant associée à aucune origine géographique ou ethnique particulière dans le cadre de l'enquête préliminaire elle aussi citée plus haut -seuls 10 % des répondants y avaient associé une origine, et les réponses étaient très hétérogènes.

La seconde difficulté rencontrée fut due aux limites posées en terme de terrain. Celui-ci devait en effet être suffisamment riche en offres d'emploi -alors même que la conjoncture économique en 2009 n'étant pas florissante- et donc indépendant des fluctuations économiques. Le domaine de la comptabilité fut choisi pour ces raisons. De même, trois niveaux hiérarchiques ont été utilisés, pour pallier les différences de genre -les postes de secrétaires comptables étant majoritairement féminins, là où les experts étaient mixtes. Il a donc fallu construire autant de batteries de CV que de types de postes à pourvoir auxquels ils ont été envoyés.

Le principal problème rencontré par **Carlo Barone** fut d'ordre éthique. Deux paramètres ont dû être pris en compte dans l'évaluation du risque éthique lié à l'expérimentation. Le premier renvoie au fait que le groupe de contrôle ne reçoive pas le bénéfice possible de l'intervention. Cet effet a été estimé peu significatif dans le contexte italien où les lycées disposent d'une large autonomie et exposent d'ores et déjà leurs élèves à une grande diversité d'interventions en matière d'orientation. De plus, on peut offrir aux établissements la possibilité d'étendre *a posteriori* le traitement aux groupes de contrôle. Le second risque, concernant les effets potentiellement négatifs du traitement sur certains élèves, a été estimé plus sérieux, conduisant à un infléchissement du dispositif expérimental. En effet, l'expérimentation exposait les enfants de milieux défavorisés au risque de suivre des voies d'enseignement les menant potentiellement à l'échec. Pour y pallier, n'ont été retenus que des étudiants avec de hauts niveaux de capacité, afin de contrôler ces risques au maximum. Concernant l'enjeu éthique entourant l'expérimentation, Carlo Barone ajoute qu'il serait utile d'étendre le questionnement éthique à la pratique consistant, pour les pouvoirs publics, à investir des montants considérables dans des interventions qui ne font l'objet d'aucune évaluation.

3-Portée pour l'action publique

Dans cette partie finale, les conclusions auxquelles peut mener l'usage d'une telle méthode ont été évoquées, tout comme les difficultés d'interprétation. Leur réception par les acteurs publics, et la manière dont ces travaux ont alimenté la réflexion sur la lutte contre les inégalités et discriminations ont finalement été abordées par les intervenants.

Jean-Benoit Eymeoud et Paul Vertier insistent sur les difficultés d'interprétation des résultats de cette expérience naturelle. En effet, les résultats observés sont sous-tendus par deux mécanismes difficiles à dissocier : une possible incompréhension des règles électorales (des électeurs pensent que l'ordre d'apparition sur le bulletin a un effet sur le résultat du scrutin), et une discrimination à l'encontre des femmes. Les chercheurs insistent sur les risques importants d'instrumentalisation politique des résultats de leur enquête, raison pour laquelle ils sont très prudents dans la présentation de leurs travaux.

En terme d'action publique, cette recherche montre qu'une discrimination peut être résorbable, à condition que les acteurs publics prennent à bras le corps le problème de l'accès à l'information, et remplissent leur devoir de faire comprendre les nouvelles règles aux électeurs. De plus, une homogénéisation des mises en forme des bulletins serait souhaitable.

Bien qu'ils n'aient pas encore été publiés et n'aient suscité aucune discussion avec les pouvoirs publics, les travaux **d'Anne Boring** débouchent sur une conclusion très importante : quand des campagnes antidiscriminatoires sont menées – comme cela a été le cas dans le métro par exemple – à partir d'un message strictement normatif (« Ne discriminez pas »), cela peut rester sans effet, voire risquer d'avoir l'effet inverse à celui visé. Il est donc primordial de tester ce type de messages avant de dépenser des millions d'euros pour leur diffusion.

Son enquête a toutefois aussi mis en évidence la difficulté d'isoler les mécanismes – par exemple, les étudiants ont pu être influencés par des lectures préalables. Il s'agit là d'une difficulté propre aux expériences de terrain.

Trois résultats sont énumérés par **Nicolas Jacquemet**.

Tout d'abord, conclusion qui n'est pas spécifique à cette étude, il apparaît qu'un candidat maghrébin a 40 % de chances en moins de recevoir une convocation à un entretien, chez les hommes comme chez les femmes. Toutefois, le fait que cette discrimination semble se généraliser à tout candidat ayant une consonance non-française en change l'interprétation : cette discrimination semble relever plus d'une préférence pour l'entre-soi que de l'exclusion d'un groupe spécifique.

Ensuite, une différence selon le genre est mise au jour : parmi les candidatures dont les noms sont à consonance française, on peut voir du favoritisme à l'égard des noms à consonance féminine, notamment quand l'employeur a lui-même un nom à consonance féminine. Ainsi, il est possible d'émettre l'hypothèse que le retard des femmes sur le marché du travail serait dû au déroulement des carrières, et pas tant que ça à l'insertion.

Enfin, cette enquête conduit à s'interroger sur le rôle de la discrimination informationnelle dans l'explication des résultats observés. Autrement dit, un plus grand nombre d'informations données pourrait faire diminuer les discriminations. L'information sur laquelle s'est concentrée l'étude fut la maîtrise de la langue, en précisant par exemple des activités extra-professionnelles suggérant une bonne maîtrise du français sur certains CV. De là, une différence selon le genre s'observe : là où les candidats aux noms à consonance maghrébine et féminine sont moins discriminés lorsqu'un tel élément figure sur le CV, aucun changement n'est visible dans le cas des candidatures avec nom à consonance masculine. Deux types de conclusions en découlent : celle d'une discrimination par les goûts envers les hommes, ou bien d'une différence de discrimination entre hommes et femmes, basée sur des stéréotypes de genre autour du comportement ou de la soumission ou non à l'autorité notamment.

Concernant l'action publique, des recommandations ont été formulées dans un [Opuscule du CEPREMAP](#), en vue d'apporter des éléments de réflexion autour de ce sujet, sur lequel le poids de l'idéologie reste conséquent et nuit à la discussion.

En guise de conclusion et d'impact possible sur l'action publique, **Carlo Barone** évoque quant à lui un système de référencement qui donne aux chefs d'établissement des outils pour répondre aux questionnements des parents d'élèves. Il insiste toutefois sur la portée limitée que peut avoir une expérimentation isolée. Il importe de multiplier les démarches expérimentales, puis de conduire des méta-analyses. Celles-ci permettent en effet de répondre de manière plus efficace à la question de la validité externe.