

Eléments de Réponse aux « Fraudistes »

Les « conclusions » proposées par Michel Husson et Thomas Coutrot (document intitulé « L'empreinte indélébile de la fraude » diffusé le 20 août) sont irrecevables, pour trois séries de raisons développées ci-après

- œ la distinction entre lots « conformes » et lots « atypiques » est fondée sur des seuils arbitraires et surtout sur des regroupements extrêmement discutables . Au total, elle n'est pas pertinente¹
- œ il n'y a pas de date ni d'heure fatidique, ni la nuit du 13 au 14, ni le 15 vers 16 heures
- œ et surtout : la constitution des lots n'est pas le fait des seules lois du hasard.

On examinera ensuite le passage « de l'hypothèse fraude à la désignation des coupables » avant de proposer une conclusion.

En annexe, on trouvera quelques observations issues des rapports « Chiche-Rouanet » et « Duménil-Lévy », qui contredisent les thèses de MH-TC sur plusieurs points importants. Une autre annexe, plus spécifique, répond à l'affirmation de MH et TC selon laquelle l'utilisation de « l'écart absolu moyen » serait « préscientifique » et inappropriée ici.

I La distinction entre lots « conformes » et lots « atypiques » est fondée sur des seuils arbitraires et sur des regroupements discutables. Au total, elle n'est pas pertinente.

Il n'y a pas un ensemble de « lots atypiques » qui s'opposeraient à des lots « conformes » mais des degrés d'éloignement de tous les lots par rapport au résultat global. L'examen des nuages de points le montre².

Le fait même de fixer un seuil pour départager deux types de lots prête à discussion, ainsi que le niveau de ce seuil. Toutes les études montrent – à qui veut bien le voir -que « l'atypicité » des lots (ou des lettres) n'est pas une question de nature mais de degré. En d'autres termes, il n'y a pas deux sortes de lots (les « conformes » et les « atypiques ») mais des lots plus ou moins éloignés de la « normale ». En amalgamant les lots « conformes » d'une part, les lots « atypiques » d'autre part, on se prive d'éléments d'analyse précieux et même indispensables. On nie l'hétérogénéité des deux ensembles ainsi constitués. On s'interdit de repérer des informations intéressantes. Cette démarche, l'amalgame inapproprié de données hétérogènes, est une des principales sources d'erreur d'analyse statistique. D'ailleurs, Michel Fenayon et Michel Lasserre apportent des éléments riches d'enseignement, notamment sur l'apparition ou la disparition de lots « atypiques, selon les seuils retenus. ³

Dans le même ordre d'idées, on décrète par endroits que des ensembles sont « proches » ou

¹Voir le document « Analyse Atypicité » diffusé par ailleurs ; on y définit un éloignement « normal » de chaque lot au résultat global en fonction de la taille de ce lot ; on en déduit un éloignement de ce lot par rapport à la « normale » ; on appelle « degré d'atypicité » cet éloignement. Certains des lots désignés par Michel Husson figurent parmi les premiers du classement selon l'atypicité ainsi définie ; d'autres sont plus loin, voire très loin, dans ce classement. De même, un classement établi à partir du rapport Chiche-Rouanet inclut, certes, des lots désignés par Michel Husson, mais pas tous, et il en inclut d'autres parmi les premiers.

²Cf document « Analyse Atypicité » et annexe « Chiche-Rouanet » ci-après.

³A noter que, dans le rapport Chiche-Rouanet, on a apparemment retenu un seuil de 0,75 comme « seuil d'atypicité » ce qui donne une vingtaine de lots « atypiques » et non une dizaine, comme chez MH. Chiche&Rouanet rejoignent ainsi Michel... Fenayon qui a aussi choisi un seuil de 0,75 et qui démontre le caractère très contestable de la sélection opérée par Michel Husson.

homogènes, sans chiffrer le degré de ces proximités (ce que j'ai essayé de faire) ni le risque de rejeter une hypothèse vraie ou de retenir une hypothèse fautive ; ce n'est pas une démarche de science statistique mais une succession de rapprochements plus ou moins pertinents, fondés sur « l'intime conviction » des auteurs.

II Il n'y a pas de date ni d'heure fatidique, ni la nuit du 13 au 14, ni le 15 vers 16H

Il n'y a pas de date fatidique, contrairement à ce que prétendent MH, TC et quelques autres: la distinction la plus claire oppose les votes dépouillés le 17 avec ceux (votes par correspondance) qui l'avaient été auparavant ; **le rapport Chichet-Rouanet indique une homogénéité des lots dépouillés après le 12 (13 inclus) ; il ne trouve donc pas la « rupture » alléguée par MH et TC entre le 13 et le 14 ni au sein des lots dépouillés le 15.** Ce dernier point entre en contradiction avec les assertions avancées ailleurs par J. Tosti⁴.

III La constitution des lots n'est pas le fait des seules lois du hasard

MH et TC considèrent les lots comme des urnes qui auraient été constituées de manière aléatoire ; Chiche-Rouanet insiste à juste titre sur le fait que cette hypothèse est absolument fondamentale. Or, elle est très loin d'être vérifiée.

En effet, divers témoignages, et les travaux de Michel Fenayon sur la lettre G, démontent cette hypothèse, qui ne tient pas la route. En particulier, il est établi que des paquets de bulletins où la première case (Harribey) étaient cochées ont été constitués et affectés à certains lots. Ces manipulations sont peu pertinentes mais elles ne sont pas frauduleuses⁵. Elles modifient évidemment la part des suffrages corrélés avec ceux de JM Harribey dans les lots qu'elles ont affectés. Cet argument, à lui seul, suffit sans doute à ruiner les élucubrations de Michel Husson et de Thomas Coutrot, puisque celles-ci postulent une constitution aléatoire des lots.

Il n'est même pas certain que les votes sont indépendants de la première lettre du nom des électeurs. Pour ne citer qu'un exemple, les noms commençant par la lettre L sont plus fréquents en Bretagne qu'ailleurs (comme le reconnaît d'ailleurs Michel Husson). Jacques Testart a signalé un cas où une étude statistique fondée sur la première lettre du nom s'était avérée fautive. Par ailleurs, les conditions du dépouillement indiquent que des ensembles de bulletins ont très vraisemblablement été affectés à une autre lettre que la leur.

Au total, le mode de constitution des lots est une autre (et même la principale) explication alternative à la fraude. Cela a été dit avec insistance. Pourquoi MH et TC font-ils mine de l'ignorer ?

IV De l'hypothèse fraude à la désignation des coupables

MH et TC proposent de laisser de côté les hypothèses « *qui imaginent une intervention extérieure aux deux listes visant à déstabiliser Attac en délégitimant le scrutin* ».

Cette hypothèse est, il est vrai, audacieuse mais pourquoi l'écarter ? Si on conclut à une fraude – ce qui n'est pas mon cas – il y a plusieurs réponses théoriquement possibles :

œdes farceurs (on a vu mieux...)

œdes « opposants » qui, voyant que Jacques Nikonoff allait dépasser largement les 50% et sans doute arriver en tête, auraient voulu discréditer le scrutin et se réserver en outre une possibilité de

⁴Duménil-Lévy signalent aussi des lots atypiques, à leurs yeux, dès avant le 14.

⁵Tenter d'invalider (en les traitant d'ânes ou de bureaucrates) les témoignages de militants qui, participant au dépouillement, ont cru ainsi bien faire est inadmissible.

le dénoncer en prenant une posture d'accusateurs (c'est « tordu », et je ne retiens pas cette hypothèse, mais on en a vu d'autres, et ce serait parfaitement rationnel)

- œdes « opposés aux opposants » (meilleure manière de désigner les soutiens des « 32 » qui se sont déclarés en désaccord avec la démarche de Susan George) avec deux variantes :
- œces fraudeurs auraient agi sur ordre, ou avec l'aval, de certains « dirigeants »
- œils auraient agi à l'insu de ces dirigeants qui seraient alors seulement « coupables » de ne pas les avoir empêchés d'agir (ce n'est pas la même chose !)
- œune officine des adeptes de la mondialisation libérale : ce n'est pas le plus vraisemblable, .. sauf si on recherche à qui profite le « crime », et qui pourrait se réjouir d'une crise majeure dans ATTAC !

Finalement, même si la démarche statistique suivie était convaincante – ce qui n'est pas le cas – on ne pourrait établir une fraude, et a fortiori l'imputer à tel ou telle, que si des indices « graves et concordants y conduisaient. L'examen physique de certains lots serait un des ces indices : il est frappant de voir que les arguments des « fraudistes » sont martelés sans être assortis d'une demande de vérification sérieuse sur ce point.

IV Conclusion : déceptions et occultation

On peut comprendre la déception de ceux qui avaient cru avoir gagné au vu des bulletins dépouillés les premiers jours. Mais cette déception n'autorise pas à considérer les résultats à mi-parcours comme « conformes » ni à qualifier « d'atypiques » les lots qui s'en écartent ⁶!

Au passage, Michel Husson qualifie de « représentatif » le résultat d'un sondage sur un échantillon de 1000. Une telle affirmation est bien surprenante. Il est vrai qu'il la nuance et, en fait la contredit ! D'une part, c'est l'échantillon – et non le résultat - qui pourrait être représentatif : les résultats seraient plus ou moins « significatifs », ce qui n'est pas la même chose. D'autre part, pour un score qui avoisinerait 50%, un « échantillon » de 3000 présenterait un « intervalle de confiance à 95% d'environ 2% ; cela signifie qu'il y a une « chance » sur vingt que le pourcentage vrai soit inférieur à 49 ou supérieur à 51. Les exemples abondent, ces dernières années, de résultats serrés où les résultats annoncés sur la base de sondages effectués sur de vrais votes - avec des échantillons beaucoup plus gros que 3000 - ont finalement été inversés... Dans un scrutin serré, le seul sondage significatif est le dépouillement de tous les bulletins. Michel ne peut l'ignorer.

Remarquons enfin que la polémique sur la fraude occulte un enseignement fondamental du scrutin : Jacques Nikonoff a recueilli environ 60% des suffrages des adhérents. On est loin du désaveu espéré par ceux qui ont mené contre lui une campagne très active et fortement concentrée sur sa personne.

Remarques sur l'étude de Gérard Duménil et Dominique Lévy

L'étude GD-DL les conduit à dire « nous voyons mal comment des experts pourraient éviter de conclure à l'existence de manipulations ». Admettons un instant cette affirmation. Mais remarquons que des « Manipulations » ne permettent pas nécessairement de conclure à une fraude, et encore moins de désigner nommément des fraudeurs ! **Des fraudes consistent à transgresser la sincérité du scrutin en diminuant le nombre des suffrages de certains candidats et/ou en ajoutant à d'autres. Des « manipulations » (éventuellement peu judicieuses, mais non frauduleuses) peuvent consister en des déplacements de bulletins qui seront comptabilisés ailleurs que dans leur urne (ou « leur lot »). C'est d'ailleurs ce qui s'est passé, selon plusieurs témoignages, et que certains ne veulent décidément pas voir.**

Cette remarque invalide les conclusions globales de GD-DL quand, de « manipulation » ils glissent vers « fraude ».

Notons aussi que l'utilisation des rangs et du « rang moyen » n'est qu'un succédané de l'utilisation des %, puisque les rangs se déduisent sans ambiguïté des % alors que l'inverse n'est pas vrai. Le rapport Chiche-Rouanet se conclut d'ailleurs en indiquant qu'il est peu utile de travailler sur les rangs quand on a travaillé sur les %. Autre inconvénient d'une analyse qui se fonderait exclusivement sur les rangs : ils amplifient l'effet de faibles variations des suffrages (au milieu du « classement ») ; bien entendu, cela relativise mais n'invalide pas l'analyse de GD-DL, d'autant qu'ils utilisent d'autres indicateurs.

Remarquons quand même la conclusion de cette partie (page 5) : [la « rupture du 13 », qui aurait été massive et orientée en faveur de la « liste de Jacques Nikonoff » ...] « n'a pas affecté des candidat(e)s particuliers, par exemple les candidat(e)s proches du 24ème rang, dont l'élection était en jeu... » Cette conclusion de GD-DL est strictement contradictoire avec ce qui constitue LA pièce jugée décisive par Michel Husson dans son argumentation !

- il est franchement gênant (pages 8 et 9) de voir qu'on écarte de l'analyse des paquets décelés comme « atypiques » mais défavorables à Susan George, et dépouillés lors de la première phase. Ne faisons pas de procès d'intention, mais on peut y voir une conséquence d'un a priori implicite. Heureusement, d'autres ont analysé ces lots.

Remarques sur l'étude de Jean Chiche et Henry Rouanet

Comment les « fraudistes » peuvent-ils se réclamer de cette étude qui, sur des points majeurs, contredit totalement leurs assertions ? Ils semblent n'en avoir extrait que ce qui les arrange, en oubliant de mentionner les précautions, indispensables et décisives, prises par les auteurs !

En effet, selon les conclusions de cette étude, les « urnes » (c'est-à-dire les « lots ») auraient été réparties en fonction de critères « politiques » permettant de les typer. **Ces conclusions s'imposent en effet si on admet l'hypothèse très explicite**, mentionnée en introduction, et résumée en conclusion (page 16) : « **si la procédure de répartition des enveloppes en urnes a été parfaitement respectée** ». Précaution claire, très utile que d'autres n'ont pas prise et n'ont pas vue dans ce rapport !

A ce « détail » majeur près si l'on admet la « neutralité statistique » d'un dépouillement par lettre (ce qui n'est pas acquis) , et à l'exception de quelques détails mineurs⁷, l'étude est intéressante par la méthodologie employée et par certains de ses résultats.

Il est notamment indiqué :

- que les urnes dépouillées les 10, 11 et 12 sont très homogènes entre elles
- et que les urnes dépouillées les 13, 14 et 15 le sont entre elles.

De cette assertion, il se déduit deux remarques, totalement contradictoires avec les affirmations des « fraudistes » :

- s'il y a une « rupture », elle intervient entre le 12 et le 13 et non entre le 13 et le 14 (alors que les « fraudistes » ont prétendu prouver une fraude « massive » entre le 13 et le 14,

- dire que les lots dépouillés le 15 sont homogènes avec ceux qui l'ont été le 13 et le 14, c'est contredire l'affirmation qu'on rencontre ici et là, selon laquelle une partie importante des bulletins dépouillés le 15 seraient proches de ceux de la première période (du 10 au 12) ; en effet, si c'était le cas, le « bloc » constitué par les lots du 13 au 15 ne serait pas homogène) ;

Par ailleurs, Chiche Rouanet mettent en évidence les corrélations entre chaque « urne » (=lot) et le cumul global ; ils utilisent le coefficient de corrélation de Pearson et signalent les lots pour lesquels ce coefficient est inférieur à 0,75 ⁸. Même si, comme ils le disent, ces calculs ne permettent pas encore de conclure, on peut alors classer les lots. Les lots « les plus éloignés de la normale » sont alors, dans l'ordre :

C6 0,180
Q1 0,318
T1 0,404
U1 0,457
NO1 0,485
B4 (0,563)
P4 0,650
W1 0,658
B3 (0,665)
R2 0,687

⁷A) Une erreur factuelle, mais minime, aisément décelable page 16 : les écarts sont LE PLUS SOUVENT mais pas TOUJOURS signés de la même façon dans le tableau 8

B) une contradiction apparente entre « l'absence d'effet chronologique » et le fait que les lots qui favorisent le plus les « anti-géorgiens » seraient tous situés le 13, le 14 ou le 15

⁸Comme Michel Fenayon, et à la différence de Michel Husson qui a « choisi » 0,65. Chiche-Rouanet mettent alors l'accent sur une vingtaine de lots.

B5 0,691

L4 0,697

T3 et I1 0,698

P2 0,701

J1 0,719

C1 0,738

G4 0,741

D4 0,749

O1, 0,783 (et 7 autres lots PM et V.PRO qui ne sont pas ventilés par lettre).

Ce classement des lots selon « l'atypicité » est très éloignée de celle de Michel Husson.

Ils en extraient deux listes au vu de l'analyse en composantes principales :

- les lots qui expliquent le mieux le « quart de plan Sud-est » (disons les « géorgiens »)

T1 W1 D4 F1 V3 (qui, disent-ils, ne représenteraient pas une série chronologique plutôt qu'une autre)

- les lots qui expliquent le mieux le quart Nord-est (les « non-géorgiens ») :

B4 T3 R2 B3 L4 V1 G4 J1 G3 P2 ajoutant que « ce ne sont que des urnes dépouillées du **13** au 15 juin »

Les autres lots (demi-plan « ouest ») correspondent aux non élus : Chiche-Rouanet ne les ont pas analysés plus précisément , mais certains de ces lots pourraient l'être.

« Défense et illustration » de l'écart absolu médian (EAM)

Selon Michel Husson et Thomas Coutrot : « *Le calcul de la « distance » revient à travailler sur l'écart entre le pourcentage obtenu par chaque candidat dans un lot donné et son pourcentage d'ensemble. Comme ces écarts sont par définition positifs pour certains candidats et négatifs pour d'autres, et que leur somme s'annule, il n'existe que deux solutions⁹ : soit prendre la valeur absolue des écarts (c'est la solution d'Alain Gély), soit prendre la somme des carrés des écarts, ce que font des indicateurs comme le coefficient de corrélation ou le Khi-deux. La théorie statistique s'est développée à partir du second choix, parce qu'il permet d'établir des lois statistiques précises et tabulées, ce que la valeur absolue ne permet pas. La solution d'Alain Gély est donc pré-scientifique et, en tout état de cause, ne devrait pas conduire à des classements différents ».*

Observons d'abord que si « elle ne doit pas conduire à des résultats très différents », pourquoi la récuser ? Parce que des travaux menés sur cette base ne conduisent pas aux mêmes conclusions que MH et TC ?

Et venons-en au caractère prétendument « préscientifique » de l'EAM.

L'argument de Michel et Thomas n'est pas « scientifique » mais scientifique ou techniciste. L'élévation au carré a, il est vrai, un avantage : elle permet la dérivabilité et autorise, effectivement, des développements mathématiques dont je n'ai nullement l'intention de nier l'intérêt.

Mais j'observe que MH et TC invoquent des lois statistiques « précises et tabulées ». Ce n'est pas un argument suffisant pour discréditer l'EAM. D'autant que, sauf omission de ma part, ces lois n'ont vraiment été utilisées que par Christophe Osswald (s'attirant une réponse de M. Husson qui indique que « le Khi-deux ne convient pas ») et implicitement par Duménil-Lévy et Chiche-Rouanet. Où sont, dans « l'analyse » de MH-TC les « intervalles de confiance », les « risques de première et de seconde espèce » et autres outils scientifiques que n'autorise pas l'utilisation de l'EAM ?

En fait, l'écart absolu moyen a un gros avantage : en s'épargnant - justement ! - l'élévation au carré des écarts, on évite d'introduire des variables peu compréhensibles pour le « profane » ; le non-statisticien peut ainsi mieux suivre l'ensemble des raisonnements. J'ajoute que l'élévation au carré exagère le poids des valeurs extrêmes de la variable étudiée. Enfin, l'EAM a une vertu peu connue : c'est par rapport à la médiane qu'il est minimum. Or, la médiane est, à bien des égards, un « indicateur de tendance centrale » préférable à la moyenne... même si ses propriétés mathématiques en limitent l'utilisation.

Les arguments employés en faveur de l'EAM sont transposables à la somme (ou la moyenne) des valeurs absolues des écarts entre deux distributions statistiques.

⁹il existe d'autres possibilités que le travail sur des « % en colonnes » : ne serait-ce qu'en travaillant sur les « pourcentages en ligne », sur les rangs de classement des candidats ou encore sur les nombres de suffrages et non sur les pourcentages qui se déduisent de ces nombres. On peut aussi travailler par lettre, et non seulement par lot. Par ailleurs Chiche-Rouanet ont effectué une analyse en composantes principales, d'ailleurs fort intéressante : si j'ai bonne mémoire, l'indicateur de distance au sein d'un nuage de points est encore un autre indicateur, même s'il présente quelques analogies avec l'écart-type.

