

Chapitre 21. « Une source de variations que nous ne pouvons expliquer »

« Le sceptique extrême n'a cure du simple bon sens, il sait que lorsque l'ordre présumé du monde semble en jeu la raison lui accorde un découvert illimité. »

B. Méheust¹.

Nature ayant successivement éreinté l'article de 1988 puis celui écrit en collaboration avec A. Spira, principalement sur des arguments statistiques, on peut supposer que l'article de S. Hirst *et al* était inattaquable sur ce point. On serait évidemment curieux de connaître les commentaires et les remarques des experts qui ont relu cet article. Ce plaisir nous est malheureusement interdit.

Toutefois, bien que les données brutes n'aient pas été communiquées par les auteurs, on peut, à partir des graphiques de l'article, retrouver assez facilement les résultats qui sont à la base de l'analyse statistique. De plus, un rapport statistique de J. Burridge destiné au département de Statistiques auquel ce dernier appartenait alors a été rédigé en mars 1992 et les éclaire singulièrement.^{2, 3}

C'est donc en confrontant le texte de l'article, l'ensemble des mesures déduites des figures et le rapport statistique de J. Burridge que nous allons analyser cet article et montrer en quoi il constitue un cas d'école.

Le protocole expérimental

Trente-six sessions de travail ont été réalisées par Hirst *et al* pour les 3 types de dilutions (Tableau 21.1). Une session était réalisée sur une journée.⁴ Les premières expériences ont commencé début juin 1990. Les échantillons de sang ont été obtenus auprès de 11 donneurs de sang qui ont pu participer à plus d'une session : cinq ont participé une fois et un a participé neuf fois. Les basophiles ont été comptés par un seul expérimentateur « entraîné ».

	Type A Anti-IgE dilué et agité	Type B Anti-IgE dilué (non agité)	Type C Solvant dilué et agité
1 contrôle + 1 anti-IgE dilution 10^2 + 8 hautes dilutions $1/10^{12}$, $1/10^{14}$... $1/10^{26}$ = 30 comptes.	5 sessions	4 sessions	3 sessions
1 contrôle + 1 anti-IgE dilution 10^2 + 8 hautes dilutions $1/10^{30}$, $1/10^{32}$... $1/10^{44}$ = 30 comptes.	5 sessions	4 sessions	3 sessions
1 contrôle + 1 anti-IgE dilution 10^2 + 8 Dilutions $1/10^{46}$, $1/10^{48}$... $1/10^{60}$ = 30 comptes.	5 sessions	4 sessions	3 sessions

Tableau 21.1. Résumé du plan d'expérience de Hirst *et al.* Il y eut 36 sessions. Chaque session correspond à une journée de travail avec la préparation des cellules, la fabrication de la série de dilutions, l'incubation des cellules avec les dilutions puis le comptage des basophiles). Chaque session était consacrée à l'étude d'une série de dilutions au 1/100 pour l'un des types de dilution (antisérum anti-IgE dilué avec agitation, antisérum anti-IgE dilué sans agitation, solvant dilué avec agitation) et pour l'une des 3 étendues de dilutions (3 intervalles ont été définis entre $1/10^{12}$ et $1/10^{40}$)

Chaque session comprenait chacune 30 comptes de basophiles. Le rapport de J. Burrige donne un exemple pour une session « anti-IgE dilué et agité » (Figure 21.1). Les lignes continues entre les tubes sont destinées à montrer comment les tubes étaient « liés » par les dilutions successives. A chaque dilution (d'anti-IgE ou de solvant) correspondent par conséquent 3 comptes de basophiles. Ce sont les moyennes de ces comptes triples qui sont présentées dans les figures de l'article.

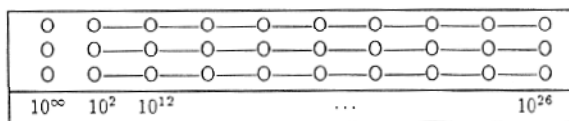


Figure 1: Linking of tubes for Type A session.

Figure 21.1. Cette figure extraite du rapport de J. Burrige correspond à une expérience de « Type A » (anti-IgE dilué et agité) et de « rang 1 » (c'est-à-dire correspondant aux hautes dilutions de $1/10^{12}$ à $1/10^{26}$). Les sessions de type B étaient identiques à part que les tubes n'étaient pas agités entre chaque dilution. Pour les sessions de type C, les seuls tubes qui contenaient des dilutions d'anti-IgE étaient les 3 tubes $1/10^2$; les 27 autres tubes des sessions de type C contenaient le solvant dilué en série dans lui-même avec agitation entre chaque dilution.

Un cas d'école

Avant d'aller plus loin, nous devons d'abord franchir l'obstacle de la figure 1. La première figure que le lecteur a sous les yeux sur la première page après avoir pris connaissance du titre et éventuellement du chapeau de présentation est la suivante.

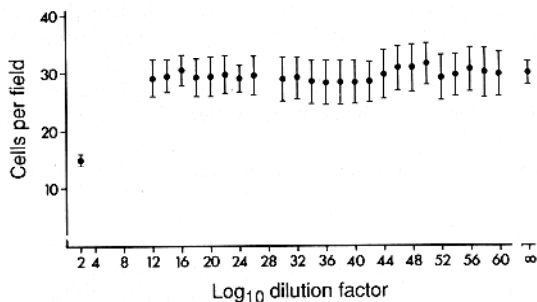


FIG. 1 Mean cell densities (with standard errors) as a function of dilution for all the data. The data for succussed anti-IgE, unsuccussed anti-IgE and succussed buffer have been combined. For the buffer control 10^2 and anti-IgE dilution 10^2 , $n=108$; for the other anti-IgE dilutions, $n=36$.

Figure 21.2. Commentaire : *succussed* = secoué, agité, selon la terminologie homéopathique qui nomme succussion le fait d'agiter la solution entre chaque dilution.

A première vue, cette figure 1 de l'article (Figure 21.2) est cohérente avec le titre puisqu'un examen rapide de cette figure semble indiquer que les hautes dilutions ne perturbent guère les comptes de basophiles. Toutefois, une lecture attentive du texte et de la légende montre qu'en fait les résultats des trois séries de données sont – contrairement à tous les « usages de la profession » – présentés *par leur moyenne commune* ! La légende de la figure indique bien que les « données ont été combinées ».

Mais, cette façon de « présenter » les résultats ne s'applique pas à l'anti-IgE à dose classique (qui est le point à l'extrême gauche du graphique : $1/10^2$) ! En effet ce dernier est présent dans toutes les sessions (qu'elles soient de type A, B ou C). L'impression première s'en trouve donc d'autant renforcée que le nombre de basophiles en présence de l'anti-IgE à dose classique ($1/10^2$), donc actif, est très diminué. Le lecteur habitué à la présentation « normale » des

graphiques scientifiques est abusé car il suppose que l'anti-IgE à dose classique et l'anti-IgE à haute dilution *sont présentés sur un même pied d'égalité, ce qui n'est pas le cas.*

Pour le lecteur peu coutumier des statistiques, illustrons cette idée simplement. Prenons l'exemple d'un médecin scolaire. Au cours de ses visites, ce dernier a eu l'impression que les enfants de l'école Jacques Prévert avaient une croissance ralentie par rapport aux enfants des autres écoles de la ville. Soupçonnant des raisons socio-économiques, il décide de faire une étude épidémiologique simple pour sensibiliser la municipalité et les autorités sanitaires. Pour cela, il recueille dans ses dossiers les tailles de tous les enfants en âge d'entrer au collège dans chacune des écoles. Il obtient ainsi une moyenne pour chacune des écoles. Il constate effectivement que la moyenne des tailles des enfants de l'école Jacques Prévert est plus faible comparée aux moyennes des autres écoles qui ne diffèrent guère entre elles. Ce sont ces moyennes par école que ce médecin communiquera aux services concernés. Dans la même situation, Hirst *et al.*, auraient fait la moyenne de l'ensemble des enfants de toutes les écoles. C'est bien entendu absurde puisque dans ce cas, les résultats de l'école Jacques Prévert risquaient d'être masqués par ceux des autres écoles. C'est pourtant ce qu'on fait Hirst *et al.* dans leur article de *Nature*.

Il est inutile de faire un procès d'intention mais si on avait voulu ne pas faire apparaître un « signal » en le noyant dans le « bruit de fond », on ne s'y serait pas pris autrement.⁵

Quelques nuages bien différents

Reprenons les résultats de l'article qui sont représentés sous forme des moyennes des 3 pourcentages de dégranulation correspondant à chaque dilution (voir figure 21.3).

A chaque dilution d'anti-IgE (ou de solvant) correspondent 3 comptes de basophiles (= un triplet). L'article présente les moyennes de ces triplets. Pour les 30 sessions, il y a donc eu $8 \times 30 = 240$ pourcentages de dégranulation avec les hautes dilutions. S'il n'y a pas de différence entre un puits « contrôle » et un puits « haute dilution », on doit par conséquent s'attendre à ce que les pourcentages de dégranulation fluctuent autour de 0%, puisque selon cette hypothèse les hautes dilutions sont censées n'avoir aucun effet. C'est ce que l'on appelle en statistiques l'hypothèse nulle.

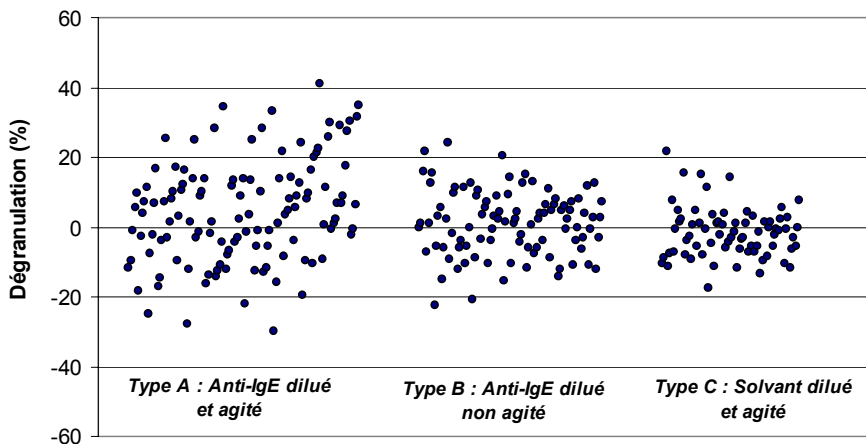


Figure 21.3. La dégranulation correspondant à chacun des points expérimentaux a été mesurée à partir de la figure 2 de l'article de Hirst *et al* (1993). Chaque point est la moyenne de 3 points expérimentaux à haute dilution. On constate que les 3 « nuages » ont des formes très différentes. Ceci suggère que le type de hautes dilutions a une influence sur les comptes de basophiles.

Pourtant, même un œil non exercé constate que les 3 séries d'expériences (dilutions d'anti-IgE avec agitation, dilutions d'anti-IgE sans agitation et dilutions du solvant avec agitation) ne se comportent pas de la même façon (Figure 21.3). Les pourcentages de dégranulation de la solution contrôle diluée et agitée paraissent moins dispersés. En revanche le nuage de points de l'anti-IgE dilué et agité est très dispersé et il semble comporter plus de valeurs positives que négatives. Quant au nuage de points de l'anti-IgE dilué et non agité, il est dans une situation intermédiaire. Le calcul des moyennes et des écarts-types confirme cette impression :

Anti-IgE dilué et agité :	$4,5 \pm 14,7 \%$
Anti-IgE dilué et non agité :	$1,1 \pm 9,3 \%$
Solvant dilué et agité :	$-1,7 \pm 7,1 \%$

Construisons les distributions des résultats

En ce qui concerne la mise en évidence d'un effet à haute dilutions, les séries A et C (anti-IgE dilué et agité et solvant dilué et agité) suffisent (la série B teste la nécessité d'agiter les tubes pour obtenir un effet). Classons chacun des points par classe : dégranulation de 0 à 9% ; 10 à 19% ; 20 à 29%, etc. et comptons combien de points appartiennent à chacune des classes. Nous obtenons alors les distributions de la Figure 21.4.

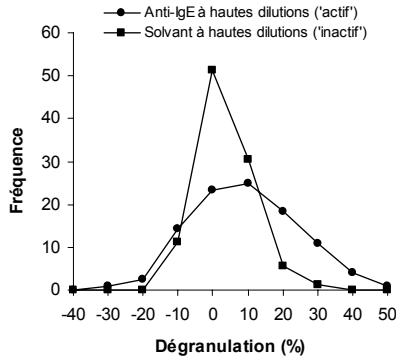


Figure 21.4. Cette figure est construite à partir des résultats de la figure 21.3 pour l'anti-IgE dilué et agité (type A) et pour le solvant dilué et agité (type C). La fréquence des comptes a été calculée pour chaque classe de pourcentage (chaque point sur l'axe des abscisses correspond à la borne supérieure de l'intervalle). On met ici clairement en évidence (quelle qu'en soit la cause) une différence de la distribution des comptes selon que les basophiles ont été en contact avec des hautes dilutions d'anti-IgE (censé être « actif ») ou avec les hautes dilutions de solvant (censé être inactif) préparées dans les mêmes conditions. Pourtant, les auteurs de l'article refuseront d'envisager la possibilité que ces différences soient un argument en faveur d'un effet des hautes dilutions d'anti-IgE.

Nous constatons donc que les deux populations sont différentes. On observe un déplacement de la distribution en faveur des dégranulations plus élevées lorsque les basophiles ont été incubés avec les hautes dilutions d'anti-IgE. Une analyse statistique montre que les deux populations sont effectivement significativement différentes. Dit d'une autre façon, cela ressemble fort à un effet des hautes dilutions d'anti-IgE !

De plus, grâce aux séries d'anti-IgE dilué mais non agité (type B ; non représentées sur la Figure 21.4), les auteurs montrent que les hautes dilutions obtenues avec agitation sont plus actives que les dilutions réalisées sans agitation ! Il est d'ailleurs difficile d'homogénéiser une solution en ne l'agitant absolument pas, ce qui pourrait expliquer la petite activité « dégranulante », bien que statistiquement non significative, qui semble avoir été transmise au cours du processus de « dilution sans agitation ».

Les 3 populations de mesures apparaissent très différentes. Et c'est bien ce qu'indiquent les calculs statistiques de Hirst *et al!* En effet, les auteurs rapportent le tableau suivant qui rassemble les valeurs de p (signification statistique) selon une analyse de variance en fonction du type de traitement (Tableau 2.2).

Table 6: ANOVA Tests* (*p*-values) For Differences Between Dilutions Separately For Each Session

Treatment Type	High Dilution Range		
	$10^{12} - 10^{26}$	$10^{30} - 10^{44}$	$10^{46} - 10^{60}$
A (combined "Fisher" <i>p</i> -value = 0.0027)	0.34	0.028	0.34
	0.066	0.018	0.21
	0.14	0.42	0.17
	0.0043‡	0.42	0.21
	0.70†	0.80	0.40
B (combined <i>p</i> -value = 0.086)	0.56	0.25	0.27
	0.27	0.97	0.27
	0.0073**	0.76	0.16
	0.084	0.48	0.44
C (combined <i>p</i> -value=0.85)	0.92	0.80	0.66
	0.25	0.25	0.21
	0.65	0.91	0.72

* ANOVA tests used the one-way F-test with (8, 18) degrees of freedom, using control and high dilution treatments in each session, ie the 10^2 dilutions were excluded.

† The missing value in this session means that the F-test used (8, 17) degrees of freedom.

‡ See Fig. 6.

**See Fig. 7.

Tableau 21.2. Ce tableau provient du rapport de J. Burridge. Il sera reproduit dans l'article de Hirst *et al* sans changement majeur. Il est sous une forme simplifiée et interprétée dans le tableau 21.3.

Traitement	Valeur de <i>p</i> (significativité statistique calculée par J. Burridge)	Interprétation (pas dans l'article)
Traitement type A (anti-IgE dilué et agité)	0,0027	Très significatif
Traitement type B (anti-IgE dilué non agité)	0,086	Non significatif (tendance)
Traitement type C (solvant dilué et agité)	0,85	Non significatif

Tableau 21.3. Version simplifiée et interprétée du tableau 21.2. Les tests statistiques (analyse de variance) réalisés par J. Burridge indiquent (quelle qu'en soit la cause) que les hautes dilutions d'anti-IgE n'ont pas le même effet – et ceci de façon très significative – sur les basophiles que des dilutions contrôles réalisées dans les mêmes conditions. On note de façon intéressante que l'agitation semble nécessaire pour observer un effet à haute dilution. En dépit de ces résultats, Hirst *et al* concluent que la significativité de ces tests est probablement le résultat d'un « artefact statistique ».

En d'autres termes, les pourcentages de dégranulation ne sont pas nuls pour les hautes dilutions d'anti-IgE. En revanche, les hautes dilutions du solvant ne sont pas significativement différentes de 0. S'il n'existe pas de biais expérimental, cela indique que les cellules n'ont pas le même comportement en présence d'anti-IgE à haute dilution ou en présence d'un contrôle.

Dans le rapport statistique, J. Burridge commente ainsi ces résultats :

« Selon les théories scientifiques classiques, il ne devrait pas, à l'intérieur de chaque session, y avoir de différences entre le traitement contrôle et les huit "traitements" à hautes dilutions. [...] Cette hypothèse peut être testée, séparément pour chaque session, en utilisant un test classique (ANOVA F-test) sur la moyenne des comptes pour chaque tube [...]. Les valeurs de p obtenues sont données dans la Table 6. Ces résultats sont étranges. En effet, si l'hypothèse nulle est correcte, ils devraient être distribués de façon uniforme entre 0 et 1. Ceci ne semble pas être le cas pour les traitements A et B pour lesquels les valeurs de p sont globalement trop faibles. »

J. Burridge va jusqu'à envisager un effet possible des hautes dilutions ! :

« La table 6 et les figures 6 et 7 suggèrent que les moyennes des comptes en triple diffèrent les uns des autres. Les raisons en sont à présent obscures. Une interprétation pourrait être qu'il existe, après tout, des différences entre les traitements – au moins pour certaines sessions et pour certains sujets. »

Laissons le lecteur goûter toute la saveur de ce « après tout ». Ayant écarté rapidement cette hypothèse qui lui paraît manifestement inenvisageable – mais que son analyse statistique est néanmoins censée évaluer ! – J. Burridge poursuit :

« L'hypothèse la plus probable pour expliquer ces effets est qu'il existerait une caractéristique de la procédure expérimentale, non identifiée jusqu'à présent, qui aurait tendance à faire en sorte que les moyennes des comptes en triple diffèrent les uns des autres au hasard ou par chance. Il est possible que la procédure de dilution en série soit responsable de cet effet, bien qu'il soit difficile de comprendre comment. »

Si l'on résume la démarche de J. Burridge, notre raison nous défendant d'envisager la possibilité d'une effet à hautes dilutions, il existe probablement un autre effet dû à la procédure expérimentale, mais « non identifié » celui-ci ! Une fois de plus, l'esprit de Descartes a paradoxalement encore frappé outre-

Manche : « Et les démonstrations de tout ceci sont si certaines qu'encore que l'expérience nous semblerait faire voir le contraire, nous serions néanmoins obligés d'ajouter plus de foi à notre raison qu'à nos sens ». ⁶

Les critiques (non publiées) du statisticien J. Burridge vis-à-vis de l'article dont il est signataire

Manifestement, les compétences du statisticien J. Burridge ont été mises à profit *après* l'obtention des résultats. Ce qui apparaît certain c'est qu'il n'a pas participé à la mise au point du protocole expérimental. Il se plaint à plusieurs reprises dans son rapport à propos de défauts dans la conception du protocole qui rendent l'analyse délicate. Ses principales critiques sont les suivantes :

1) Absence de randomisation entre les sessions :

« [...] il existe certaines caractéristiques qui gênent quelque peu l'analyse et d'autres qui rendent parfois l'interprétation problématique. Par exemple, il ne semble pas que l'on ait pris la peine soit de randomiser l'ordre dans lequel les sessions étaient comptées dans le temps ou d'équilibrer les rangs des dilutions au sein de chaque type de traitement par rapport aux donneurs de sang volontaires. Ainsi, les sessions de type A ont été faites d'abord, suivies par les sessions de type B puis les sessions de type C. De la même façon, au sein de chaque type de sessions, les sessions de rang 1 ont été faites d'abord, puis les sessions de rang 2 suivies par les sessions de rang 3. La répartition des volontaires dans les différentes sessions a été réalisée inévitablement de façon hasardeuse. L'absence de randomisation de l'ordre des sessions est une caractéristique malvenue de l'expérimentation et signifie que certains effets du "traitement" pourraient être attribuables à des tendances au cours du temps (et pourraient être dues en particulier à un effet de l'apprentissage acquis par les expérimentateurs tout au long de l'étude). »

2) Les « liens » entre les 3 séries de dilutions ont été « brisés » :

« [...] pour la plupart des sessions les tubes étaient "liés" par la procédure de dilution en série. Ceci signifie que, grâce à cette liaison, les résultats des différentes dilutions successives pourraient être corrélés au sein des séries de neuf dilutions de telle sorte qu'idéalement ces séries pourraient être analysées comme des entités uniques avec leurs propres moyennes et covariances. Toutefois ce lien n'a pas été enregistré au cours de la procédure de randomisation qui faisait suite et par conséquent ne peut pas être prise correctement en compte pour l'analyse statistique. »

Or, concernant ce deuxième point, l'un des résultats mis en exergue dans l'article est qu'il n'existe pas de périodicité dans les dégranulations en fonction des dilutions. J. Burridge reconnaît pourtant implicitement que les auteurs de l'article ne se sont pas donnés les moyens d'analyser ce point précis. En effet, lorsque les tubes ont été codés, il n'a pas été noté à laquelle des trois dilutions en série ces tubes appartenaient (voir les « liens » entre les dilutions de chacune des trois séries de dilution sur la Figure 21.1). Il est tout à fait possible que les auteurs de l'article ne pensaient pas qu'ils seraient amenés à analyser la périodicité car ils n'avaient pas envisagé qu'un effet global significatif puisse être observé. Il est probable que les différences d'effet des différents traitements aient surpris les auteurs de l'article et qu'ils aient cherché ensuite à les « balayer sous le tapis » en s'attachant à montrer que les « vagues » de dégranulation décrites dans l'article de *Nature* de 1988 n'étaient pas retrouvées.

Comment les auteurs de l'article ont-ils surmonté cette difficulté pour « démontrer » néanmoins – en dépit de leurs résultats – que les hautes dilutions d'anti-IgE étaient sans effet ?

Un « artefact statistique »...

Bien entendu, ces critiques de J. Burridge à propos des problèmes de méthodologie n'apparaissent pas dans l'article. En effet, si les auteurs reprennent bien dans des termes voisins la description ci-dessus concernant les comportements différents des 3 traitements (correspondant ci-dessus à « Selon les théories scientifiques classiques ... les valeurs de p sont globalement trop faibles »), à aucun moment ils n'envisagent la possibilité que l'effet observé soit lié aux hautes dilutions d'anti-IgE. Pour expliquer ces différences qu'ils n'attendaient pas entre les traitements, les auteurs introduisent une notion nouvelle et curieuse, celle d'« artefact statistique » :

« Bien qu'il soit possible que les effets observés soient un artefact statistique, ils est possible que ces derniers puissent s'expliquer par un élément non identifié de la procédure expérimentale. C'est un point intéressant de nos résultats mais il ne doit pas servir, bien entendu, à soutenir les résultats de Davenas *et al.* et il permet de souligner une fois de plus la complexité de l'analyse de variance dans un essai de ce type. ».

La notion tout à fait originale d'« artefact statistique » joue ici le rôle de *joker*. Des résultats inattendus ayant été obtenus, un « artefact statistique » est alors invoqué pour les réfuter. De plus, on ne saurait tout contrôler se plaignent les auteurs ; il est possible que notre protocole expérimental nous joue des tours. Et puis, semblent-ils dire, toutes ces analyses sont si compliquées...

Mais tout ceci n'empêche pas les auteurs de conclure l'article avec front :

« Nous ne sommes pas parvenus à mettre en évidence un quelconque effet reproductible par les très hautes dilutions d'anti-IgE, agitées ou non agitées, sur la dégranulation des basophiles humains ».

On aboutit donc à une conclusion en désaccord avec les résultats obtenus. *A minima*, si les auteurs avaient conscience que leur protocole expérimental comportait des faiblesses et n'était donc pas capable de répondre à la question posée, il leur était loisible de le notifier (en évoquant les critiques du rapport de J. Burridge sur lesquelles l'article est muet). Mais leurs résultats auraient-ils alors mérité d'être publiés dans les pages de *Nature* ? Au final, c'est la stratégie « tout sauf les résultats de Benveniste » qui a prévalu.

Les arguments de la réponse de J. Benveniste et A. Spira

C'est sur une unique colonne que J. Benveniste et A. Spira peuvent s'exprimer dans *Nature* et, comme nous l'avons déjà mentionné, huit longs mois après la publication de l'article⁷. Ils notent quinze différences entre la méthodologie de Hirst *et al* et celle de l'article de *Nature* de 1988 mais se contentent, étant donné la place limitée, de pointer les plus importantes. Ainsi Hirst *et al* ont inclus dans leur analyse toutes les expériences réalisées, y compris celles pour lesquelles la dégranulation avec l'anti-IgE à dose classique était faible. Or, explique une nouvelle fois J. Benveniste, si on n'observe pas de dégranulation avec l'anti-IgE à doses classiques, on a peu de chance d'observer un effet avec les hautes dilutions d'anti-IgE.

Une autre critique exprimée par J. Benveniste et A. Spira est d'avoir évalué dans des expériences séparées les différents traitements que l'on souhaitait comparer. C'est le reproche le plus important car c'est l'essence même de l'expérimentation en biologie de ne faire varier qu'un facteur à la fois. La procédure correcte aurait consisté à comparer dans la même session les hautes dilutions d'anti-IgE et les hautes dilutions du solvant. Enfin, une étape de centrifugation a été ajoutée après l'incubation des cellules avec les hautes dilutions. Cette procédure est susceptible d'augmenter la dispersion des comptes.

Dans une deuxième partie, J. Benveniste et A. Spira questionnent la méthode statistique employée et pointent les « tactiques » utilisées pour masquer les différences statistiquement significatives. Pour terminer, ils soulignent que les auteurs ont refusé de communiquer les données brutes des expériences.

L'exercice de la critique est toutefois délicat. On ne peut en effet reprocher aux auteurs à la fois de ne pas avoir respecté le protocole expérimental original et affirmer dans le même temps que les résultats prouvent néanmoins l'existence d'un effet à hautes dilutions.

Dans la presse, J. Benveniste s'exprime de façon plus directe en affirmant que Hirst *et al* ont « commis plusieurs fautes méthodologiques et déontologiques ».⁸ Pour la fameuse « Figure 1 » en particulier, il considère qu'il s'agit d' « une manipulation des résultats sans précédent dans l'histoire des sciences (mélange de résultats actifs et contrôles) ».

Puis, il poursuit :

« Aussi peu déontologique [...] est le fait que je n'ai pas été approché pour la mise au point des nombreux détails nécessaires à la bonne pratique d'expériences aussi complexes et que je n'ai appris l'existence de cet article que par la presse. Il est extrêmement surprenant de voir une revue comme Nature, qui se pose en archétype de l'excellence et de l'intégrité scientifique, se livrer à une telle manipulation. La question est : quels en sont les motifs réels ? »

Plus technique, A. Spira déclare :

« Au total [...] je ne pense pas que ces résultats soient contradictoires avec les nôtres et je pense qu'il serait nécessaire que nous puissions échanger nos données brutes de façon à comparer les résultats des deux séries d'expériences en utilisant la même stratégie d'analyse statistique ».

Echanger des données, comparer des résultats, ne serait-ce pas faire de la recherche ? Mais était-ce le souci des auteurs et de ceux qui ont promu la publication de cet article ?⁹

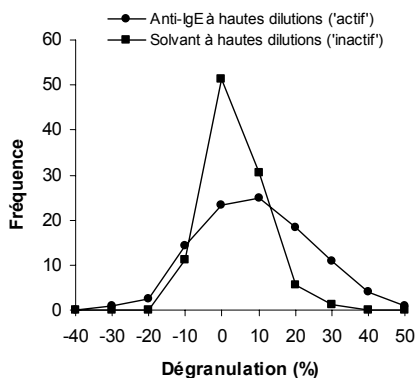
Comparaison avec les résultats de l'article des Comptes Rendus de l'Académie des Sciences

La comparaison de ces résultats avec les résultats de l'étude faite en collaboration avec l'équipe d'A. Spira et publiée dans les *Comptes Rendus de l'Académie des Sciences* est par ailleurs fort étonnante. En effet, la présentation des résultats sous forme de distribution des dégranulations conduit à des profils similaires (Figure 21.5).

Il est extrêmement curieux de constater que des résultats, au fond très semblables, sont interprétés avec des conclusions opposées. Dans les deux cas,

les hautes dilutions ont un comportement différent de celui des dilutions contrôles et les courbes « sinusoïdales » ne sont pas présentes. On s'en souvient, J. Benveniste avait dit : « Nous retrouvons les résultats publiés dans *Nature*. C'est la même fille, toujours aussi belle. Il ne lui manque qu'un peu de maquillage. »¹⁰ Mais pour Hirst *et al* cette absence de « maquillage » est un argument décisif pour assurer : « Nous ne pouvons mettre en évidence aucune variation périodique ou polynomiale de la dégranulation en fonction de la dilution de l'anti-IgE » et que par conséquent en aucun cas on ne pouvait affirmer que les résultats de J. Benveniste avaient été confirmés.¹¹

Benveniste et al (1991)



Hirst et al (1993)

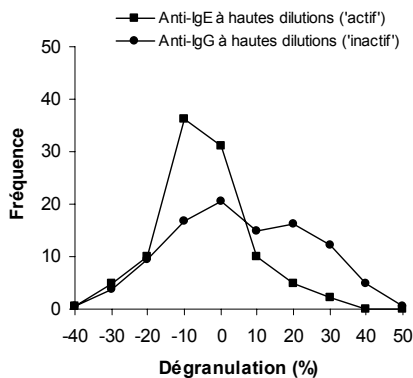


Figure 21.5. Comparaison des résultats de l'article de J. Benveniste et A. Spira¹² de 1991 dans les *Comptes Rendus de l'Académie des Sciences* et celui de Hirst *et al* de 1993 dans *Nature*. Des résultats extrêmement voisins sont obtenus dans les deux cas mais des conclusions diamétralement opposées quant à un effet des hautes dilutions d'anti-IgE en sont tirées.

(Chaque point sur l'axe des abscisses correspond à la borne supérieure de l'intervalle).

J. Maddox et l' « esprit poppérien »

Pour en finir avec les rapports de J. Benveniste et de la revue *Nature*, nous concluons sur une réflexion, sous forme de prédiction, que J. Maddox avait inséré dans le long commentaire final de quatre pages qu'il écrit (non cosigné par les autres enquêteurs) dans *Nature* du 27 octobre 1988.

En effet, en septembre de l'année 1988, J. Maddox écrit aux équipes israéliennes, italiennes et canadiennes pour leur demander si elles désirent commenter les événements des derniers mois.¹³ L'équipe de Toronto lui répond en particulier qu'elle a entrepris de poursuivre ses travaux sur les hautes dilutions dans un « esprit poppérien ». J. Maddox ajoute dans son texte final du

27 octobre qu'il serait « heureux de publier dans la rubrique Correspondance Scientifique la conclusion générale de l'un ou l'autre de ces groupes lorsqu'ils seraient prêts. »¹⁴ Puis, à propos de J. Benveniste et de son équipe, il forme le vœu que ces derniers « compteront dorénavant les basophiles en répétant les comptes, en suivant la procédure standard pour contrôler les erreurs d'échantillonnage et qu'ils élimineront le biais inévitable lié à l'observateur en faisant en routine des expériences à l'aveugle ».

Enfin, il conclut :

« Je m'attends à ce que ces résultats ne soient guère différents de ceux obtenus dans les trois expériences à l'aveugle (chacune avec deux expérimentateurs) à Clamart les 9 et 10 juillet (*sic*)¹⁵ ; ce serait extrêmement intéressant s'il en était autrement, mais je ne doute pas que le Dr Benveniste préférera publier dans un autre journal ».

La première prédiction de J. Maddox ne s'est donc pas accomplie. En effet, au cours de la série d'expériences faite en collaboration avec A. Spira et son équipe un effet des hautes dilutions a bien été retrouvé dans des conditions expérimentales correctes et beaucoup plus rigoureuses que pour les expériences citées par J. Maddox.

La deuxième prédiction ne s'est pas non plus réalisée. En effet, comme on l'a vu plus haut, le manuscrit rapportant ces expériences a bien été proposé à *Nature*. On a pu constater également comment J. Maddox s'est employé à disqualifier une nouvelle fois ces résultats. On a vu comment au contraire l'article de Hirst *et al* a passé la barrière des experts apparemment sans trop de difficultés en dépit des questions légitimes qu'il aurait dû soulever si les mêmes critères avaient été appliqués que pour l'article de J. Benveniste et A. Spira.

L'esprit poppérien consiste précisément à mettre en doute et à éprouver ses propres convictions à la lumière de l'expérimentation. Il semble bien que pour J. Maddox l'esprit poppérien s'applique à tous les scientifiques, à l'exclusion toutefois du Directeur de *Nature*.

Notes de fin de chapitre

¹ B. Méheust. Somnambulisme et médiumnité, tome II, *Les Empêcheurs de Penser en Rond / Synthélabo*. 1998.

² J. Burridge. A repeat of the « Benveniste » statistical analysis. Department of Statistical Science. University College London. Research Report N° 100, March 1992.

³ Ce rapport m'a été aimablement communiqué en 2001 sur simple demande auprès du secrétariat du Département de Statistiques du *Royal College* de Londres.

⁴ Si les expérimentateurs ont découpé ainsi les expériences c'est probablement qu'ils estimaient que réaliser une expérience comportant 30 puits était suffisant pour une journée de travail. Il est dommage qu'ils n'aient pas bénéficié des conseils et de l'aide de W. Stewart qui avait une conception nettement plus stakhanoviste du comptage des basophiles...

⁵ Une explication possible de cette « approche » statistique pourrait résider dans le fait que si certaines dilutions censées être actives faisaient varier le nombre de basophiles, la variance des comptes en serait augmentée par rapport aux dilutions « inactives ». Il reste que cette démarche est pour le moins inhabituelle.

⁶ R. Descartes. *Principes de la philosophie* (1644).

⁷ J. Benveniste, B. Ducot et A. Spira. Memory of water revisited. *Nature*, 4 août 1994, p. 322.

⁸ F. Nouchi. Une équipe de chercheurs anglais n'a pu reproduire les travaux du docteur Benveniste sur la « mémoire de l'eau ». *Le Monde*, 11 décembre 1993.

⁹ Il semble néanmoins que le manuscrit ait nécessité quelques améliorations puisqu'il a été reçu par la rédaction le 16 avril 1993 et accepté le 22 octobre de la même année.

¹⁰ M. de Pracontal. *Les mystères de la mémoire de l'eau*, p. 200.

¹¹ Cette incohérence entre les résultats et les conclusions des auteurs a été remarquée néanmoins par certains scientifiques. En particulier les analyses indépendantes d'Italo Vecchi ainsi que celles de Jean-Pierre Pharabod diffusées sur Internet peu après la publication de l'article de Hirst *et al* conduisent aux mêmes conclusions, à savoir que l'hypothèse nulle (c'est-à-dire qu'il n'existe pas de différence entre les contrôles et les tubes « actifs ») doit être rejetée et que par conséquent les résultats contredisent le titre même de cet article.

¹² Pour les résultats de J. Benveniste et A. Spira, les pourcentages de dégranulation au sein de chaque expérience sont calculés par rapport à la moyenne des comptes des puits anti-IgG à haute dilution.

¹³ L'équipe israélienne répondit en deux vagues. Tout d'abord B. Robinzon expliqua qu'il lui était difficile de commenter une enquête à laquelle il n'avait pas assisté (lettre du 18 septembre 1988 à J. Maddox). Puis tout en commentant plusieurs points techniques, il réaffirma sa conviction qu'« un phénomène biologique avait bien été observé et pas un artefact, même s'il n'y avait pas d'explication pour le moment ». Il conclut : « Je n'ai pas fait de commentaires jusqu'à présent [...] car je ne pense pas que c'était à nous de prouver que le phénomène que nous avons observé était réel ou était un artefact, en

particulier après votre rapport qui nous a discrédités ». Il demanda toutefois à J. Maddox de ne pas publier sa lettre. J. Amara et M. Oberbaum répondirent longuement plus tardivement en reprenant de même plusieurs points techniques et se plaignirent eux aussi du traitement par la revue britannique des travaux sur les hautes dilutions qui selon eux n'était « pas convenable pour un journal du calibre de *Nature* à propos d'un travail scientifique » (lettre de J. Amara et M. Oberbaum du 11 décembre 1988).

L'équipe italienne répondit début octobre à la lettre de J. Maddox et reprocha le traitement « déloyal » de sa contribution.

¹⁴ J. Maddox. Waves caused by extreme dilution. *Nature*, 27 octobre 1998, p. 760.

¹⁵ Les enquêteurs étaient repartis le vendredi 8 juillet 1988; J. Maddox veut parler du jeudi 7 et du vendredi 8.