A. Marchand [1]
F. Dress [2]
J. Dayantis [3]

Essai d'approche expérimentale du comportement lors de l'interpolation d'une courbe

Introduction

L'interpolation et l'extrapolation sont des pratiques scientifiques courantes. Elles peuvent s'effectuer par un calcul ou simplement de manière intuitive.

Le calcul d'une courbe interpolée ou extrapolée à partir des points connus suppose une hypothèse (explicite ou implicite) concernant la forme de la courbe ou la nature mathématique de la fonction qu'elle représente.

Lorsque le calcul est remplacé par l'intuition (en général on trace la courbe « à la main » au-delà des points connus ou entre ceux-ci), la méthode mathématique cède la place à une « boîte noire » qui fonctionne dans le cerveau de l'opérateur. Mais, l'existence de cette boîte noire n'exclut pas une hypothèse tout à fait inconsciente sur forme de la courbe (même dans le cas où la courbe est, par exemple, « de toute évidence » une droite... ou un cercle).

Quelle est cette hypothèse inconsciente ? Comment fonctionne la « boîte noire » du cerveau ? Ces questions sont actuellement sans réponses objectives, bien que des suggestions, des propositions aient été avancées. Dayantis (1973) (1), par exemple, introduit la notion de « courbe d'estimation » de l'observateur, entité en général *subjective*, mais que dans certains cas particuliers il admet être une *donnée objective*. Ainsi suppose-t-il que la distribution des valeurs *estimées* (extrapolées ou interpolées, à partir des mêmes points connus, par un grand nombre d'opérateurs différents, ou par un même opérateur au cours d'un grand nombre d'essais) est une donnée aussi objective que celle des valeurs mesurées avec un instrument. Cette hypothèse a été utilisée pour calculer le contenu en information d'une courbe expérimentale [Dayantis (1973) (1) et Marchand (1978) (2)]. Dayantis suppose même que la distribution est normale.

L'étude présentée ici tente de cerner quelque peu le fonctionnement de la « boîte noire » et de mettre en évidence quelques-unes de ses caractéristiques :
- globalement, d'une part,
- de façon différentielle, d'autre part (suivant les caractéristiques du sujet).

Ses résultats jettent des doutes sur les hypothèses d'objectivité et de normalité formulées par Dayantis.

Description du test

Le test utilisé est constitué de dix épreuves. Il a été présenté à 169 personnes volontaires prises dans les Universités de Bordeaux et de Strasbourg, se répartissant ainsi :
- Hommes 120 (71 %)
- Femmes 49 (29 %)
- Droitiers 161 (95,3 %)
- Gauchers 8 (4,7 %)
- Mathématiciens (enseignants ou chercheurs) 16 (9,5 %)
- Scientifiques ou techniciens non mathématiciens 142 (84 %)
- Ni scientifiques, ni techniciens 11 (6,5 %)

Les scientifiques ou techniciens (mathématiciens ou non) avaient un niveau d'études correspondant à l'enseignement supérieur. Les non-scientifiques, non-techniciens avaient au moins le niveau baccalauréat ou équivalent.

Chaque épreuve (voir annexe) se présente sous la forme d'un certain nombre de points que l'on dit appartenir à une même courbe expérimentale. On a laissé volontairement de côté dans cette étude les perturbations causées par de possibles erreurs expérimentales, et les points connus sont supposés parfaitement exacts. Plus précisément ils sont définis à la précision du tracé de la figure.

Les tests sont successivement présentés, dans l'ordre de 1 à 10, à la personne testée, qui y répond dans le même ordre, sans retour en arrière. Il lui est demandé de déterminer, en traçant la courbe « à la main », sans autre instrument qu'un crayon et sans aucun calcul, son point d'intersection avec une droite verticale située soit à droite, soit à gauche, soit au milieu des points connus. Les axes horizontaux et verticaux de la figure sont matérialisés. Les abscisses des points connus sont équidistantes.

Ces points ont en fait (et à l'insu des personnes testées) été pris sur des courbes d'équations connues. On peut donc comparer les réponses à celles données par ces équations.

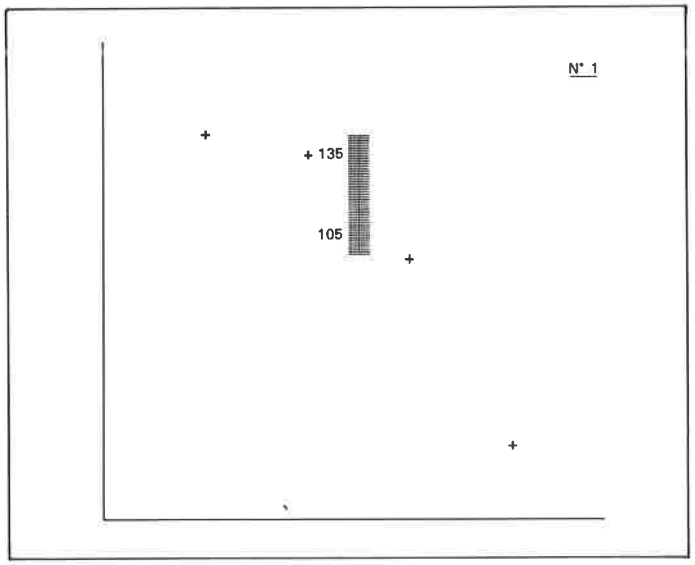
Épreuve 1 : Interpolation à partir de 4 points pris sur une branche de parabole (à droite du sommet) à concavité vers le bas.

Épreuve 2 : Extrapolation à gauche à partir de 3 points. La concavité est tournée vers le haut. Avec 3 points on ne peut définir sans ambiguïté aucun type de courbe simple. Nous avons utilisé deux valeurs de référence particulières : celles provenant de la parabole d'axe vertical et de l'exponentielle passant par ces 3 points.

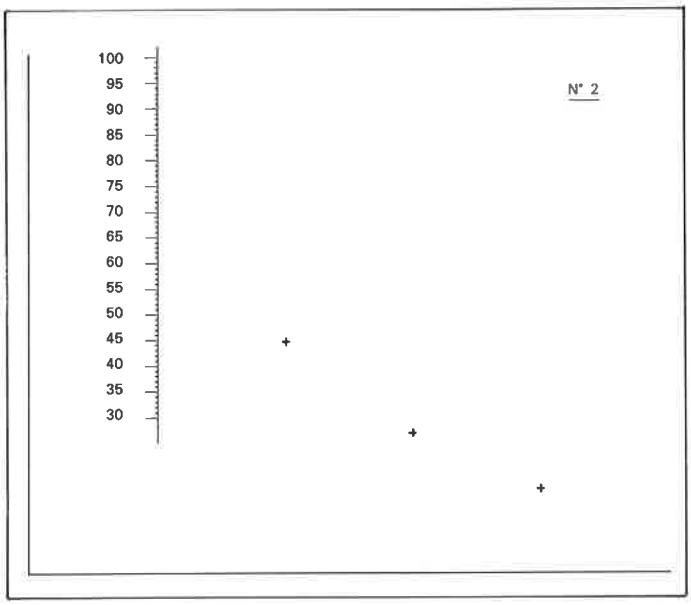
Épreuve 3 : Extrapolation à droite à partir de 3 points. Les figures des épreuves 2 et 3 sont symétriques par rapport à un axe vertical.

Épreuve 4 : Extrapolation à droite à partir de 4 points situés sur une

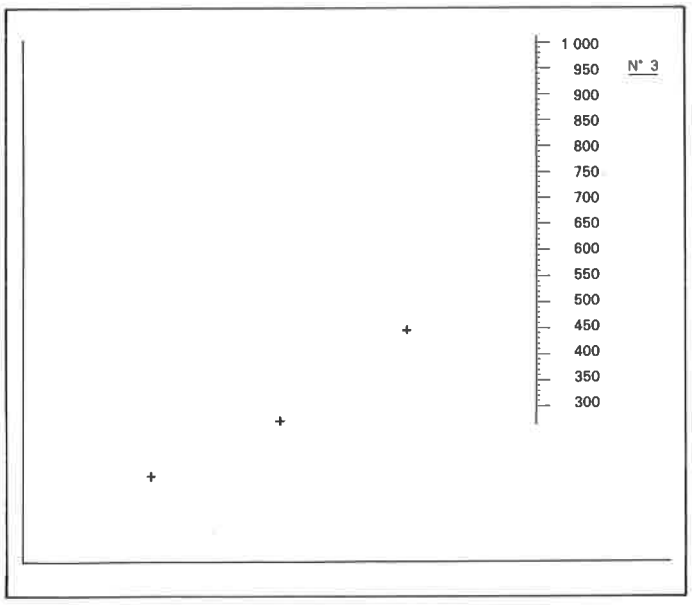
[1] *Centre Paul Pascal (CNRS) et Université de Bordeaux I.*
[2] *UER de mathématiques et d'informatique, Université de Bordeaux I, Laboratoire associé au CNRS n° 226.*
[3] *Centre de recherche sur les macromolécules (CNRS), Strasbourg.*



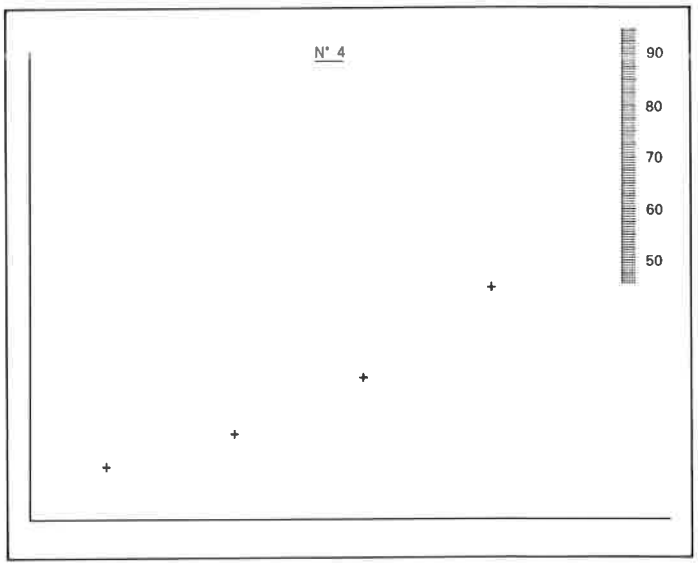
Épreuve 1.



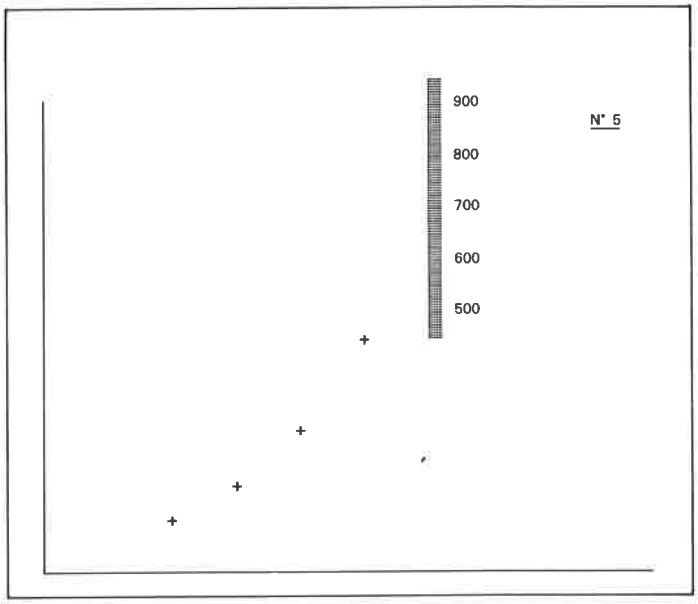
Épreuve 2.



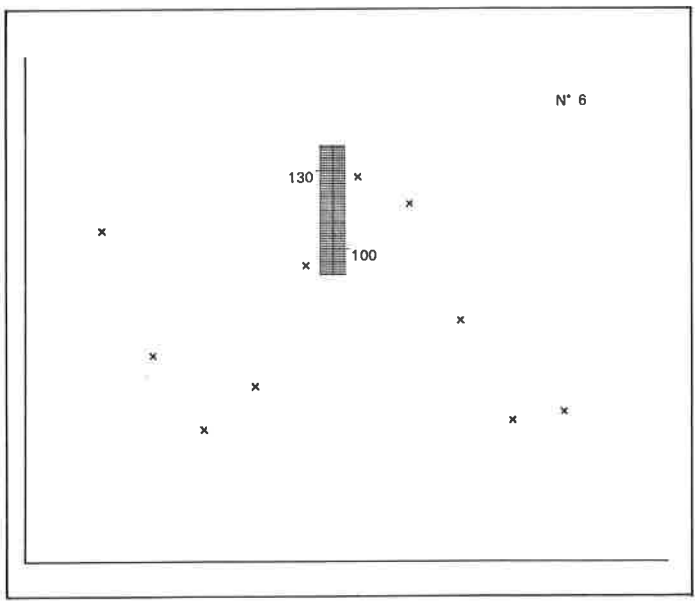
Épreuve 3.



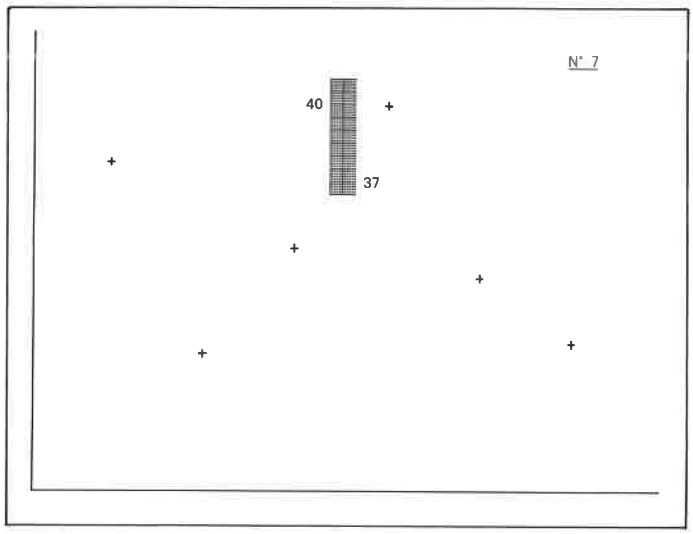
Épreuve 4.



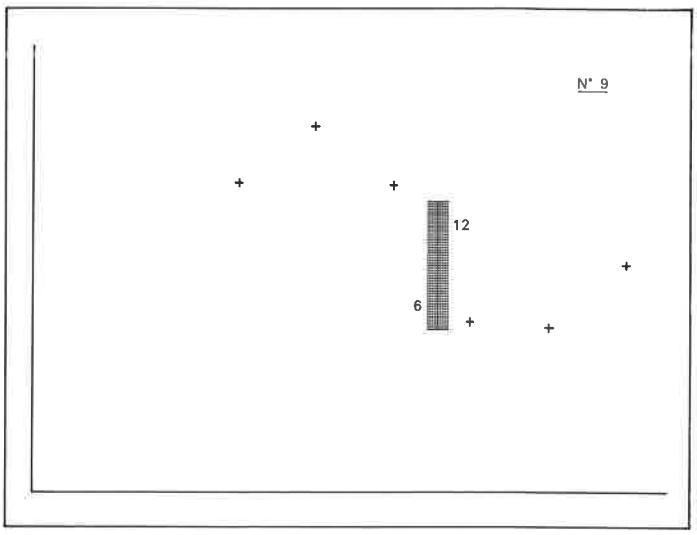
Épreuve 5.



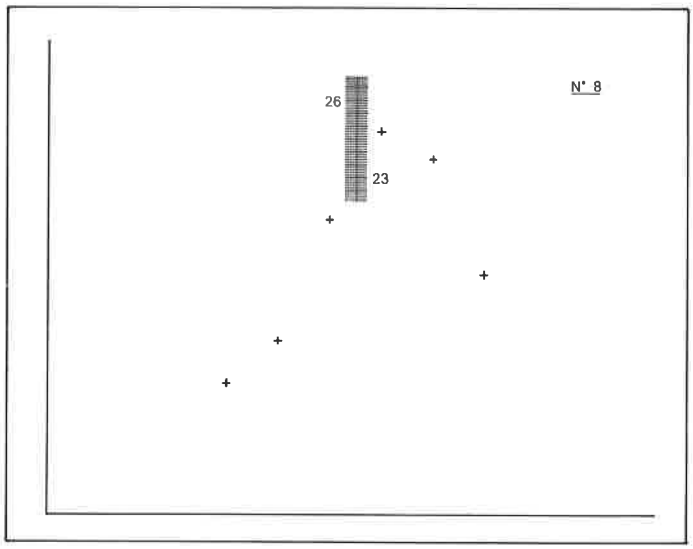
Épreuve 6.



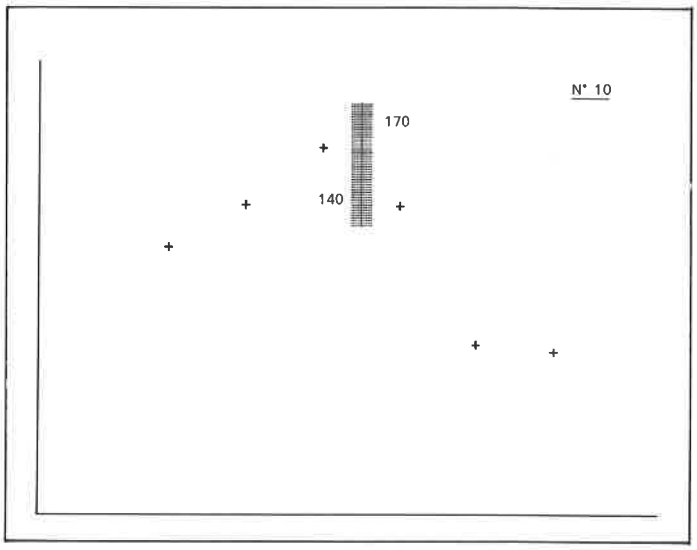
Épreuve 7.



Épreuve 9.



Épreuve 8.



Épreuve 10.

exponentielle E identique à celle de l'épreuve 3 : les 3 points de droite de l'épreuve 4 sont identiques aux points de l'épreuve 3.

Épreuve 5 : Identique à l'épreuve 4, mais les abscisses y sont contractées dans le rapport 1/2.

Épreuve 6 : Interpolation (au voisinage d'un maximum) à partir de 10 points choisis sur une sinusoïde de manière telle que l'allure sinusoïdale de la courbe apparaisse.

Épreuve 7 : Interpolation au même point de la même sinusoïde que l'épreuve 6. Mais 6 points seulement de cette courbe sont donnés. L'allure sinusoïdale reste cependant apparente.

Épreuve 8 : Interpolation à partir de 6 points pris sur la même sinusoïde que les épreuves 6 et 7; mais l'allure sinusoïdale de la courbe n'est plus apparente : on a simplement supprimé les 4 points extrêmes de l'épreuve 6.

Épreuve 9 : Interpolation à partir de 6 points pris sur une « courbe de dispersion » ($y = a - bxe^{-x^2}$). La zone à interpoler est située au voisinage d'un point d'inflexion et de pente maximale.

Épreuve 10 : Interpolation à partir de 6 points pris sur la même courbe que l'épreuve 9 : la zone à interpoler est située près d'un sommet.

Résultats : caractéristiques globales

Le tableau 1 rassemble les résultats obtenus par l'ensemble des sujets testés sur chacune des 10 épreuves : allure symétrique ou non de la courbe de distribution des réponses, caractéristiques dissymétriques éventuelles, valeur moyenne et écart-type de la distribution. On y a également indiqué pour chaque épreuve les valeurs calculées à partir de l'équation ou de l'une des équations simples compatibles avec les points connus de la courbe. Enfin, des tests de comparaison de moyennes (tests de Student) ont été effectués, entre la moyenne des réponses et la (ou les) valeur(s) calculée(s) dans le tableau 1, entre les moyennes de sous-populations dans la suite de l'étude. Le nombre de degrés de liberté est très grand (supérieur à 100), et la variable t de Student possède donc une distribution normale. Une extrême prudence nous a conduits à utiliser le vocabulaire suivant :

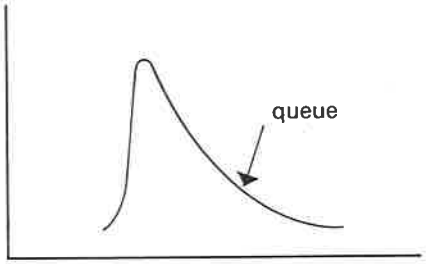
- très significatif : si $t > 3{,}89$ (probabilité d'erreur $< 0{,}001$)
- significatif : si t entre 3,29 et 3,89 (probabilité d'erreur entre 0,000 1 et 0,001)
- faiblement significatif : si t entre 2,58 et 3,29 (probabilité d'erreur entre 0,001 et 0,01)
- non significatif : si $t < 2{,}58$ (probabilité d'erreur $> 0{,}01$).

On notera que, dans la plupart des épreuves, un certain nombre de sujets ont donné une réponse « à l'infini » ou hors des limites de la figure, alors que cette possibilité n'était pas offerte explicitement.

Tableau 1

Numéro de l'épreuve	(Int)	(Ext)	(Ext)	(Ext)	(Ext)	(Int)	(Int)	(Int)	(Int)	(Int)
	1	2	3	4	5	6	7	8	9	10
Dissymétrie de la distribution des réponses	+	+ +	+ +	+ +	+ +	+	0	0	+	+
Caractéristiques dissymétriques	Queue vers convexité de la courbe (haut)	Réponses ∞ 7 % queue vers concavité	Réponses ∞ 5 % queue vers concavité	Réponses ∞ 2 % queue vers concavité	Réponses ∞ 20 % queue vers concavité	Queue vers convexité (haut)	Réponses ∞ 4 %	Réponses ∞ 2 %	Queue vers le bas (voisinage de l'inflexion)	Réponses ∞ 5 % queue vers convexité
Valeur théorique calculée à partir de l'équation	120,0 (P)	(P) 69,3 (E) 73,9	(P) 693 (E) 739	73,9 (E)	739 (E)	115,4 (S)	38,54 (S)	23,54 (S)	10,02 (D)	159,0 (D)
Moyenne des valeurs des réponses	121,9	76	763	74,4	779,2	118,2	38,7	24,0	8,3	164,3
	3,9	8,1	73,1	6,5	8,1	3,9	1,1	0,6	0,9	6,9
Écart-type de la distribution Test de Student sur l'écart entre moyenne et valeur théorique	t = 6,33 TS	(P) t = 10,4 TS (E) t = 3,25 FS	(P) t = 12,1 TS (E) t = 4,15 TS	t = 0,99 NS	t = 5,5 TS	t = 9,3 TS	t = 1,85 NS	t = 9,8 TS	t = 24,8 TS	t = 9,7 TS

courbe de distribution dissymétrique



NS : non significatif ($t < 2{,}58$)
FS : faiblement significatif (t entre 2,58 et 3,29)
S : significatif (t entre 3,29 et 3,89)
TS : très significatif ($t > 3{,}89$)
(P) : équation d'une parabole
(S) : équation d'une sinusoïde
(E) : équation d'une exponentielle
(D) : équation $y = a - b \times e^{-x^2}$
(Int) : interpolation
(Ext) : extrapolation

1. La première caractéristique générale mise en évidence est la dissymétrie de la courbe de distribution des réponses. Dans 8 cas sur 10 cette dissymétrie est évidente, alors qu'elle est douteuse dans les 2 autres cas. La courbe de distribution présente donc, le plus souvent, un côté « abrupt » et une pente plus douce (une « queue ») de l'autre côté.

On peut chercher à discerner une corrélation entre le côté où se trouve cette « queue » et la concavité ou la convexité de la courbe à inter/extrapoler. On note que la queue est du côté de la concavité dans tous les cas d'extrapolation et du côté convexe pour les interpolations (dans l'épreuve 9, l'interpolation est faite au voisinage d'un point d'inflexion, de sorte que la concavité/convexité est négligeable).

La queue est aussi toujours vers le haut de la figure, à une exception près (épreuve 9).

On observe, enfin, que la queue se trouve toujours (par rapport à la valeur moyenne des réponses) *du côté opposé* à la valeur théorique calculée à partir de l'équation de la courbe (queue vers les valeurs supérieures à la moyenne pour les épreuves 1, 2, 3, 4, 5, 6, 10 et vers les valeurs inférieures à la moyenne pour l'épreuve 9).

Nous reviendrons sur ces corrélations, mais *la dissymétrie* de la courbe expérimentale de distribution des réponses *infirme* déjà l'hypothèse *a priori* de Dayantis (1) concernant le *caractère gaussien* de cette courbe.

2. La seconde caractéristique générale évidente est que l'opérateur (ou l'observateur) imagine toujours une courbure plus importante que celle correspondant aux équations simples qui sont compatibles avec les points connus.

Ceci apparaît si l'on compare *les valeurs moyennes* des réponses avec les valeurs calculées à partir des équations des paraboles, exponentielles, sinusoïdes, etc., passant par les points donnés. On constate, par exemple, que dans l'épreuve 1 l'ordonnée calculée pour la parabole est 120,0 alors que la moyenne des réponses est 121,9 : les observateurs ont donc tendance à situer le point *interpolé* du côté *convexe* de la parabole, c'est-à-dire qu'ils imaginent une courbe *plus courbée* que la parabole. Inversement, dans l'épreuve 5, l'ordonnée calculée pour l'exponentielle est 739 alors que la moyenne des réponses vaut 779,2 : les observateurs placent donc le point *extrapolé* du côté *concave* de l'exponentielle. Là encore ils imaginent une courbe *plus courbée* que l'exponentielle. On constate ainsi que les réponses moyennes sont toujours, dans l'interpolation, décalées du côté convexe de la courbe calculée (épreuves 1, 6, 7, 8, 10) et que dans l'extrapolation elles sont toujours décalées du côté concave (épreuves 2 à 5). Les écarts correspondent donc toujours à une exagération de la courbure (cf. figures 1*a* 1*b*) et les tests de Student montrent qu'à deux exceptions près (épreuves 4 et 7) ils sont significatifs.

Si la « boîte noire » du cerveau paraît préférer l'exponentielle à la parabole (n° 2 et 3) c'est sans doute parce que, dans la région considérée, la courbure de l'exponentielle est supérieure à celle de la parabole.

Il est très probable que la dissymétrie observée dans les courbes de distribution des réponses provient du même phénomène. Nous avons en effet noté que, d'une part, la « queue » de la courbe de distribution est toujours du côté concave dans l'extrapolation et du côté convexe dans l'interpolation, et que, d'autre part, elle est toujours (par rapport à la valeur moyenne) du côté opposé à la valeur théorique. Ces deux observations *traduisent en fait la même constatation :* la queue et la valeur moyenne sont toutes deux du même côté de la valeur théorique, du côté concave dans l'extrapolation et du côté convexe dans l'interpolation.

Le pourcentage (parfois élevé) de réponses « à l'infini » correspond sans doute au même phénomène.

Cette amplification de la courbure semble intervenir d'autant plus que la courbure est déjà importante : c'est ce que suggèrent les résultats comparés des épreuves 4 et 5, où les courbes à extrapoler ne diffèrent que par l'échelle des abscisses. L'épreuve 5 correspond à des abscisses contractées, donc à une courbure plus importante, que les « opérateurs » ont eu tendance à exagérer fortement, indiquant parfois des valeurs infinies (20 %) et obtenant pour l'ensemble des autres valeurs une moyenne très supérieure à la valeur théorique calculée pour une exponentielle (test de Student très significatif) alors que les abscisses dilatées de l'épreuve 4 donnent à la courbe une allure relativement rectiligne, d'où une exagération moindre de la courbure (test de Student non significatif). La même observation est sans doute valable pour les épreuves 7 et 8 qui représentent la même sinusoïde, mais où les points connus de 7 sont plus écartés que ceux de 8, et donc où la courbure de 8 est plus apparente que celle de 7.

3. L'extrapolation « à droite » et l'extrapolation « à gauche » conduisent au même résultat. C'est au moins ce que semblent bien montrer les épreuves 2 et 3. Il faut tenir compte de ce que l'unité utilisée pour les ordonnées de 3 est 10 fois plus petite que celle de 2. Exprimés avec la même unité, les valeurs moyennes et les écarts-types sont 760 et 81 pour l'épreuve 2, 763 et 73,1 pour l'épreuve 3. Le test de Student appliqué à l'écart entre les deux valeurs moyennes donne $t = 0,35$: cet écart n'est donc pas du tout significatif.

Il aurait été intéressant de confirmer cette première impression par une comparaison des performances des sujets droitiers et gauchers. Le nombre trop réduit de réponses de gauchers ne l'a pas permis.

4. La dernière constatation générale que l'on peut faire est que l'influence de l'aspect global de la figure l'emporte sur celle de l'aspect local. Il y avait en effet deux couples d'épreuves conçus en vue de cette question.

• *a*) Épreuves 3 et 4 : 3 résulte simplement de la suppression du point *le plus à gauche* de 4. Les résultats montrent que les indications fournies par ce 4e point permettent une extrapolation *à droite* bien plus précise de l'exponentielle, et que l'exagération de la courbure, très marquée dans l'épreuve 3, est négligeable (voir test de Student) dans 4.

• *b*) Épreuves 6 et 8 : 8 résulte de la suppression des deux points extrêmes à droite et des deux points extrêmes à gauche de 6, de sorte que l'allure sinusoïdale de la courbe, évidente dans 6, n'apparaît pas dans 8. L'interpolation dans les épreuves 6 et 8 est demandée au même endroit. Exprimés avec la même unité, on obtient les résultats présentés dans le tableau 2.

Il est clair que, d'une part, l'exagération de la courbure est plus marquée dans l'épreuve 8 (de 115,4 à 120,0) que dans l'épreuve 6 (de 115,4 à 118,2), que, d'autre part, les deux moyennes (118,2 et 120) sont faiblement significativement différentes (test de Student $t = 3,24$), et enfin que les valeurs des écarts-types montrent une dispersion des résultats nettement plus importante dans l'épreuve 8.

• *c*) On peut donc conclure que les informations données par l'aspect global de la figure orientent de manière prépondérante la

Tableau 2

	Valeur moyenne	Écart type	Valeur calculée (sinusoïde)
Épreuve 6 (sinusoïde apparente)	118,2	3,9	115,4
Épreuve 8 (sinusoïde non apparente)	120,0	6,0	115,4

réponse du sujet testé. Moins ces informations sont précises, plus la réponse sera hésitante et même éventuellement fausse, quels que soient par ailleurs les renseignements fournis par l'aspect local de la figure au voisinage de la zone à inter — ou extrapoler. La « boîte noire » du cerveau fonctionne donc surtout, semble-t-il, à partir d'un « *modèle global* » (probablement non explicité) de la courbe présentée, et non pas tellement à partir des données locales concernant le « champ opératoire ».

La comparaison des résultats des épreuves 9 et 10 suggère d'ailleurs un processus qui pourrait être celui du fonctionnement intuitif : *le positionnement préalable* (et implicite) du sommet éventuel de la courbe le plus proche de la zone opératoire. Dans ces deux épreuves la courbe présentée est la même, ainsi que le nombre de points connus. Mais dans 10 la position du sommet proche de la zone d'interpolation est assez bien suggérée, alors que dans 9 la position de ce sommet (qui est un minimum) est très douteuse (les sujets testés *ne savent pas* que les courbes 9 et 10 sont identiques). L'expérience montre que (exprimé avec la même unité) l'écart absolu entre la moyenne des réponses et la valeur théorique est nettement plus important dans l'épreuve 9 ($8{,}3 \times 5 = 41{,}5$ à $10{,}02 \times 5 = 50{,}1$) que dans l'épreuve 10 (159 à 164,3) : en règle générale les sujets semblent, dans l'épreuve 9, avoir situé le minimum nettement plus haut qu'il n'est sur la courbe théorique imaginée par les auteurs du test, ce qui correspond d'ailleurs bien à une exagération de la courbure de cette partie de la figure.

Résultats : caractéristiques différentielles

Nous espérions initialement disposer de suffisamment de réponses pour que certaines catégories minoritaires soient représentées de manière significative. Le nombre de gauchers (8) et le nombre de mathématiciens (16) ou de non-scientifiques (11) se sont trouvés malheureusement trop faibles pour que des conclusions sures puissent être obtenues.

Tableau 3

N° de l'épreuve	1	2	3	9
Valeur calculée	(P) 120,0	(P) 69,3 (E) 73,9	(P) 693 (E) 739	(D) 10,02
Moyenne générale	121,9	76	763	8,34
Moyenne hommes	122,1	75,5	759,5	8,32
Moyennes femmes	121,5	77,3	772	8,39
Écart-type hommes	4,4	7,4	69,7	0,8
Écart-type femmes	2,3	9,5	80,3	1,1
t de Student entre les moyennes hommes et femmes	0,90 NS	1,27 NS	0,98 NS	0,46 NS

Tableau 4

	Épreuve 2	*Épreuve 3*
Valeur calculée	(P) 69,3 (E) 73,9	(P) 693 (E) 739
Moyenne générale Moyenne mathématiciens	76 81,8	763 794,9
Écart-type général Écart-type mathématiciens	8,1 7,7	73,1 96,4
Test de Student sur les moyennes (*t*)	2,74	1,61

Les seules catégories vraiment utilisables sont les hommes (120) et les femmes (49). Or aucune différence significative n'apparaît entre les deux groupes de réponses, ainsi que le tableau 3 le montre sur quelques exemples.

Il semble, en revanche, qu'une indication puisse être tirée des réponses des mathématiciens, malgré leur faible nombre. Le tableau 4 permet de comparer le groupe des mathématiciens à l'ensemble, pour les épreuves 2 et 3.

On constate dans ces deux cas que, avec un écart-type voisin de celui de l'ensemble des réponses, la moyenne des mathématiciens se situe encore plus loin des valeurs calculées que la moyenne générale. En d'autres termes, les mathématiciens semblent exagérer la courbure plus encore que l'ensemble de ceux qui ont répondu au questionnaire. Le test de Student ne permet cependant de considérer comme significatif (faiblement) que le premier écart et ces deux résultats nous autorisent seulement à poser la question.

Conclusion

Deux idées souvent admises chez les scientifiques : objectivité de certaines situations d'interpolation et d'extrapolation, caractère gaussien de la courbe de distribution des réponses, ont été mises en doute par cette étude de caractère artisanal.

Quelques éléments d'une approche du fonctionnement de la « boîte noire » impliquée ont pu être avancés : attrait de la courbure locale, et en même temps fabrication inconsciente d'un modèle global.

Il est bien évident qu'il ne s'agissait là que d'un travail d'exploration. Des recherches ultérieures plus approfondies sont indispensables pour confirmer et préciser ces premières indications, et pour examiner l'effet des caractéristiques du sujet.

Bibliographie

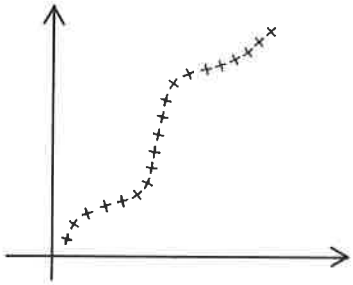
(1) J. Dayantis : « Essai de quantification de l'information obtenue lors de mesures à un appareil de mesure », *J. Chim. Phys.*, 1973, 70, 865.
(2) A. Marchand : « Calcul de l'information fournie par une courbe expérimentale : la règle d'affinité et la conservation de l'information », *J. de Phys.*, 1978, 39, 117.

Annexe page suivante

Annexe
Le test tel qu'il était proposé

Quelques explications sur ce test

Une courbe peut être définie par un certain nombre de points (voir par exemple la figure ci-dessous). Plus ces points sont nombreux, mieux la courbe est définie. Il faut naturellement tenir compte de la précision avec laquelle les points sont fournis : pour un nombre de points donné, la courbe est d'autant mieux définie que la précision est plus élevée.



A partir des points connus, on peut déterminer par extrapolation ou par interpolation, des points inconnus avec une plus ou moins bonne précision.

Si l'opération est faite par un calcul à partir d'hypothèses sur la « forme » * de la courbe, et *si on se fixe une précision a priori*, la courbe est complètement définie.

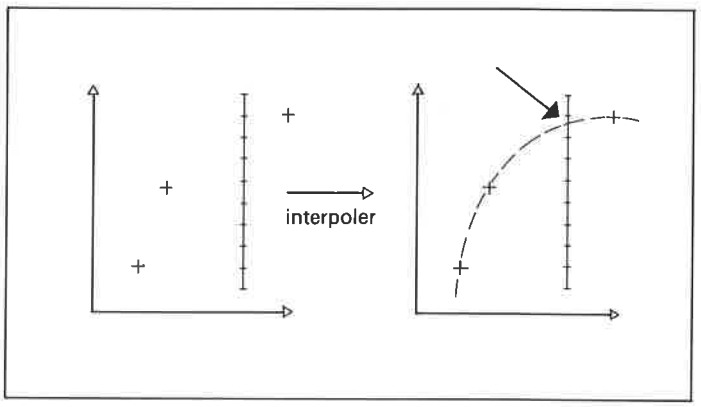
Si l'opération est réalisée sans calcul par un expérimentateur se fiant à son intuition, le résultat sera influencé par le nombre de points connus et des hypothèses sur la « forme » de la courbe, formulés de façon tout à fait inconsciente. Le résultat risque également d'être influencé par la manière dont la courbe est disposée (échelle, position des points, symétries éventuelles, etc.) et les épreuves présentées ici sont destinées à étudier ces influences.

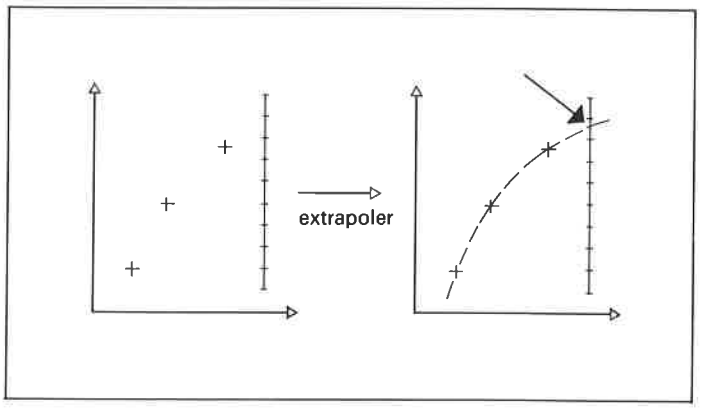
Chaque figure représente un certain nombre de points d'une courbe. La courbe est différente pour chaque figure.
La courbe passe exactement par tous les points indiqués, à la précision du dessin. Sur chaque figure est aussi présenté un segment de droite gradué : on vous demande d'indiquer, *sans vous servir d'aucun instrument autre qu'un crayon*, à quel endroit la courbe rencontre l'échelle graduée (vous pouvez, si vous le désirez, tracer cette courbe au crayon).

* *Pour un mathématicien, cela signifie « forme polynômiale, forme exponentielle, forme logistique... ».*

Nom :
Prénom :
Adresse :
Profession :

Êtes-vous gaucher ou droitier ?





N.B. 1) Vous devez travailler sans retourner la feuille de bas en haut, même si cela vous paraît incommode.
2) Si vous avez l'impression que la courbe extrapolée ne rencontre pas l'échelle graduée, indiquer le par une flèche.

Il devient absolument nécessaire que vous rédigiez vos manuscrits en respectant la symbolique et la terminologie édictées par l'IUPAC (publiées dans le « Manuel des symboles et de terminologie des grandeurs et des unités physico-chimiques »).