



CNRS FORMATION ENTREPRISES

Le CNRS propose un catalogue
de plus de 260 formations technologiques
à destination des entreprises pour l'année 2020

L'organisme de formation continue du CNRS présente un large panel de stages courts (3 jours en moyenne) dans tous les domaines : des biotechnologies à la synthèse chimique en passant par l'intelligence artificielle, le big data et les enjeux sociétaux.

Cet outil de transfert des savoirs et savoir-faire du CNRS permet de **former plus de 1600 professionnels chaque année** issus des secteurs privé (PME et grands groupes) et public.

QUELQUES NOUVEAUTÉS PROPOSÉES PAR CNRS FORMATION ENTREPRISES EN 2020

DONNÉES, CONNAISSANCES, APPRENTISSAGE Apprentissage automatique pour la vision par ordinateur du 15 au 17/06/2020 LYON - 3 jours	CARACTÉRISATION DES MATÉRIAUX - Rhéologie du 01 au 03/04/2020 PARIS - 3 jours - Fluorescence X de type EDX du 28 au 29/05/2020 LE MANS - 2 jours	BIOLOGIE MOLÉCULAIRE ET BIOCHIMIE Préparer des banques NGS à partir d'ADN et d'ARN : les étapes pratiques et méthodologiques appliquées à la technologie Illumina du 09 au 11/03/2020 LYON - 3 jours
BIOINFORMATIQUE Bioinformatique pour la métagénomique du 03 au 05/06/2020 MONTPELLIER - 2,5 jours	CHIMIE, SYNTHÈSE, PROCÉDÉS Mécanochimie pour une chimie de synthèse éco-compatible du 01 au 02/04/2020 MONTPELLIER - 2 jours	BIOLOGIE ANIMALE ET FORMATIONS RÉGLEMENTAIRES Les 3 R en pratique, raffinement, points limites le 03/03/2020 GIF-SUR-YVETTE - 0,5 jour
GÉNIE LOGICIEL ET SYSTÈMES D'INFORMATION Développement logiciel : cycle et prévention des risques juridiques le 06/10/2020 PARIS - 1 jour	CHIMIE ANALYTIQUE Transformation de la biomasse en produits chimiques de spécialité du 11 au 12/06/2020 POITIERS - 2 jours	QUALITÉ ET SÉCURITÉ Maîtrise des risques potentiels liés aux nanomatériaux en laboratoire de recherche le 06/10/2020 NANCY - 1 jour
TERRITOIRE, PATRIMOINE ET ENVIRONNEMENT Initiation aux SIG en écologie : de la collecte au traitement de données géographiques du 18 au 20/05/2020 POITIERS - 2,5 jours	MICROSCOPIE ET IMAGERIE Préparation d'échantillons par fusion alcaline pour analyses ICP du 11 au 12/06/2020 VANDŒUVRE-LES-NANCY - 1,5 jour	SOCIOLOGIE, SCIENCE POLITIQUE ET ÉCONOMIE Jeunesses en exil : quel accueil des mineurs non accompagnés ? du 23 au 25/11/2020 PARIS - 3 jours
SCIENCES DE L'INGÉNIEUR Applications du calcul scientifique dans l'industrie du 17 au 19/03/2020 PARIS - 3 jours	BIOLOGIE CELLULAIRE ET MICROBIOLOGIE Cytométrie en image du 18 au 20/11/2020 LYON - 3 jours	COGNITION ET COMPORTEMENT Les mouvements oculaires : techniques d'oculométrie, recueil et analyse de mouvements oculaires, interprétation, applications du 26 au 27/03/2020 PARIS - 2 jours
PHYSIQUE ET INSTRUMENTATION Interactions plasma / surfaces : utilisation des plasmas froids pour le traitement de surfaces du 24 au 27/03/2020 GRENOBLE - 3,5 jours		

Découvrez l'ensemble des stages de formation sur cnrsformation.cnrs.fr
Ou contactez-nous à cfe.contact@cnrs.fr ou 01 69 82 44 55



L'ESSENTIEL

> Il y a cent ans, l'écrivain Pierre Louÿs a lancé une polémique en affirmant que Molière n'était pas l'auteur des pièces qui lui sont attribuées.

> Au début des années 2000, des travaux de linguistique quantitative sont venus appuyer cette thèse, qui a reçu beaucoup de publicité.

> De nouvelles analyses linguistiques, utilisant six méthodes différentes, l'invalident et confirment la paternité de Molière sur ses œuvres.

LES AUTEURS



FLORIAN CAFIERO
ingénieur de recherche
au CNRS et chargé
de cours à l'École
nationale des Chartes
– université Paris-
Sciences & Lettres



JEAN-BAPTISTE CAMPS
maître de conférences
à l'École nationale
des Chartes – université
Paris-Sciences & Lettres

Molière est bien l'auteur de ses œuvres

On a longtemps soupçonné Jean-Baptiste Poquelin de ne pas avoir écrit ses pièces, dont le véritable auteur serait Corneille. Mais l'étude de la langue de Molière et de ses contemporains rend à l'homme de théâtre ce qui lui est dû.

Il y a maintenant cent ans, le poète et romancier Pierre Louÿs lança une polémique: plusieurs articles de lui, notamment le 16 octobre 1919 dans le quotidien français *Le Temps*, soutenaient que l'auteur d'*Amphitryon* n'était pas Molière, mais son illustre contemporain Pierre Corneille. Depuis, le soupçon n'a cessé de peser sur la paternité des pièces signées de Molière. Comment un comédien présumé sans grande éducation littéraire, qui plus est fort occupé entre ses charges de valet de chambre du roi et de directeur de troupe de théâtre, aurait-il pu écrire tant de chefs-d'œuvre?

Au cours des décennies qui ont suivi, plusieurs ouvrages sont parus sur ce problème, défendant l'idée que Molière n'était en fait que l'acteur principal des comédies qu'on lui attribue. Le vrai génie littéraire, Pierre Corneille, serait resté dans l'ombre, profitant des revenus de la pièce sans ternir sa réputation d'auteur sérieux, et sans s'exposer aux controverses que ces comédies souvent polémiques ont pu susciter.

Très discutée, cette thèse a pris une nouvelle dimension lorsque, au début des années 2000, des travaux en linguistique quantitative de Dominique Labbé, de l'IEP de Grenoble, et Cyril Labbé, de l'université Grenoble 1, ont soutenu que les vocabulaires des pièces de Corneille et Molière sont trop proches pour que ces œuvres aient été écrites par deux auteurs différents, et, par conséquent, que Corneille aurait bien écrit les pièces de Molière. La nouvelle a été reprise par la presse mondiale, évoquée ou défendue dans différents reportages et documentaires, et a même donné son sujet à un téléfilm.

Mais la popularité de cette thèse ne doit pas masquer les nombreuses réactions défavorables qu'elle a suscitées dans le monde de la recherche littéraire et linguistique. Pour tenter de trancher la question, nous avons récemment réalisé de nouvelles analyses des textes de Molière, de Corneille et de leurs contemporains. D'après leurs résultats, qui viennent d'être publiés, la théorie niant à Molière la paternité de ses pièces est erronée. >



Une scène de la pièce *Le Malade imaginaire*, de Molière, peinte par Abraham Solomon (1861).

> Les arguments laissant à penser que Molière n'aurait pas écrit ses pièces sont d'abord historiques. Il ne nous est ainsi parvenu aucun manuscrit, pas même une lettre, une note ou un brouillon, qu'on puisse lui attribuer. Il s'agit là d'un cas unique pour un auteur aussi prolifique et célèbre. Le moliériste Georges Forestier, de l'université Paris-Sorbonne, note toutefois que les manuscrits d'auteurs du XVII^e siècle, alors jugés de peu d'importance, étaient très rarement archivés, mais bien plus souvent détruits. Aucun brouillon ou manuscrit de pièces de Corneille n'a ainsi été conservé. De Racine, il ne nous est parvenu qu'un acte d'un projet de pièce finalement abandonné. Et si ses descendants l'ont précieusement gardé, c'est probablement parce que cette pièce n'a jamais été imprimée, et que ce brouillon en était la seule trace restante. Rien de si alarmant, donc, à ne rien retrouver de la main de Molière.

Autre élément ayant prêté au soupçon : Molière aurait commencé à jouer les pièces de Corneille après un voyage à Rouen en 1643, et se serait mis à écrire ses chefs-d'œuvre après un séjour en 1658 dans la même ville. Or Rouen était le fief de Corneille. Ces voyages en terre normande auraient-ils été l'occasion d'ententes secrètes entre les deux hommes ? Les spécialistes de Molière en doutent, et ne sont guère impressionnés par la coïncidence. Il est possible, mais nulle part attesté, que Molière et Corneille se soient rencontrés sur place. Et l'on ne trouve aucune trace d'un quelconque accord conclu entre les deux hommes. Dans sa correspondance, Corneille semble d'ailleurs indifférent au passage de la troupe en 1658. Troisième ville du pays, Rouen était régulièrement visitée par les troupes de théâtre, qui n'y allaient pas nécessairement voir Corneille.

Dernier argument majeur : l'éclosion tardive du talent de Molière. Celui-ci écrivit tous ses chefs-d'œuvre passé l'âge de 37 ans, ce qui n'a



Molière, de son vrai nom Jean-Baptiste Poquelin (1622-1673), gravure de James Poppelwhite publiée dans *The Gallery of Portraits: with Memoirs*, d'Arthur Thomas Malkin (Londres, 1833).

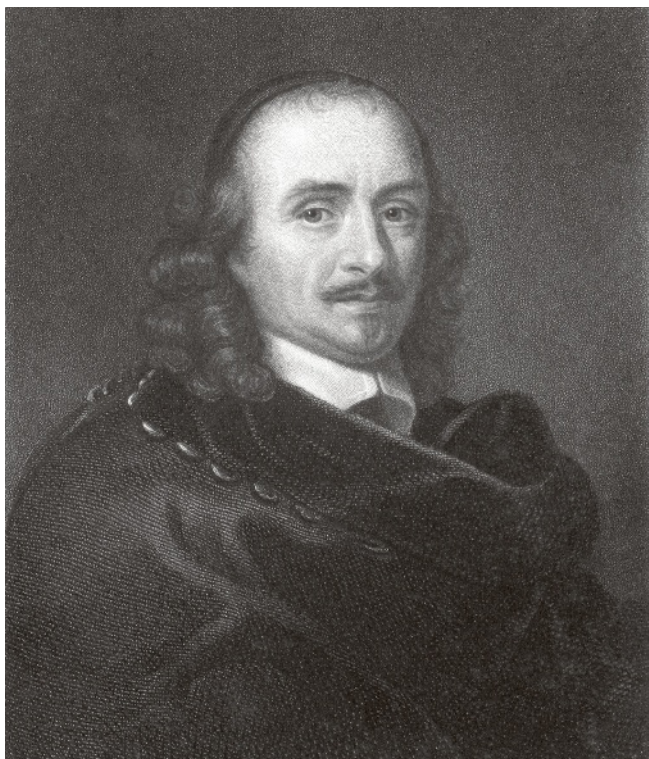
pas manqué d'étonner. Selon Hippolyte Wouters, avocat belge et auteur de théâtre, ce serait le seul cas « où un auteur médiocre jusqu'à 40 ans devient non seulement profond, mais surtout une des plus belles plumes de son temps ». Si cette situation est en effet assez rare dans l'histoire littéraire, elle n'est cependant pas inédite – Marcel Proust, James Conrad ou Umberto Eco ont débuté des carrières littéraires tout aussi tard, et ont été couronnés d'un succès certain.

RETOUR AUX TEXTES

Faute d'archives pouvant prouver l'imposture supposée, la controverse historique semblait mener à une impasse. Pour apporter de nouveaux éléments aux débats, des chercheurs ont décidé de s'intéresser aux textes eux-mêmes. Si les textes signés de la main de Molière étaient stylistiquement proches de ceux de Pierre Corneille, ce pourrait être le signe que ce dernier en est l'auteur.

Le constat de départ de Dominique et Cyril Labbé est simple : deux textes écrits par un même auteur ont plus de vocabulaire en commun que deux textes écrits par des auteurs différents. En prenant comme base la fréquence d'apparition de chaque mot dans un texte, ils ont calculé une « distance intertextuelle », qui mesure la différence entre deux textes dans le choix du vocabulaire. Si la distance intertextuelle séparant deux pièces est

L'éclosion tardive du talent d'un auteur est une situation rare, mais pas inédite



Pierre Corneille (1606-1684), gravure de Thomas Woolnoth publiée dans *The Gallery of Portraits : with Memoirs*, d'Arthur Thomas Malkin (Londres, 1833). On a supposé, à tort semble-t-il, qu'il était l'auteur des pièces de Molière.

inférieure à un certain seuil, cela signifie selon eux qu'elles ont été écrites par le même auteur.

D'après les calculs de Dominique et Cyril Labbé, les œuvres attribuées à Corneille et de nombreuses pièces de Molière sont trop proches pour avoir été écrites par deux personnes différentes. Cela porterait à croire que Corneille a écrit la plupart des chefs-d'œuvre signés de Molière (l'hypothèse selon laquelle Molière aurait écrit les pièces de Corneille étant invraisemblable et impossible pour des raisons chronologiques).

Depuis ces travaux fondateurs, les techniques permettant de découvrir l'auteur d'un texte ont évolué. Leur usage, tant par des philologues cherchant à identifier la plume derrière un poème jamais signé que par les services de renseignement pour démasquer l'auteur d'une lettre anonyme, s'est également étendu.

ANALYSER LES HABITUDES D'ÉCRITURE ET LES TICS DE LANGAGE

Ces méthodes reposent sur l'analyse statistique des habitudes d'écriture et des tics de langage. En écrivant un texte, nous laissons involontairement des empreintes qui trahissent notre identité : chaque individu utilise avec une fréquence particulière des mots, des expressions ou des structures grammaticales. En les relevant systématiquement, on peut se faire une excellente idée de qui a écrit un texte.

Depuis quelques années, la démarche la plus couramment adoptée pour révéler l'auteur d'un texte repose sur l'intelligence artificielle. On réunit d'abord une sélection de textes dont on connaît les auteurs avec certitude. On entraîne alors l'ordinateur à reconnaître les auteurs de chacun de ces textes avec la meilleure précision possible. Puis on lui fournit des textes anonymes, pour lesquels la machine prédit l'auteur le plus vraisemblable.

Mais pour construire le profil type des textes d'un auteur, encore faut-il être d'accord sur les pièces que cet auteur a réellement rédigées. Or, dans notre débat, tout le monde n'est pas d'accord sur qui a écrit quoi. La théorie des « comédiens poètes », défendue notamment par Dominique Labbé, voudrait que, comme dans les spectacles d'humour d'aujourd'hui, dont on connaît l'interprète mais rarement le ou les auteurs, on aurait au XVII^e siècle mis en avant le comédien principal et entièrement omis de citer le dramaturge. Quelque 90% des comédies de l'époque auraient alors été signées par un prête-nom. Si l'on veut prendre au sérieux cette possibilité, il faut donc adopter une autre approche.

Faute d'accorder une totale confiance à l'apprentissage automatique d'une machine, tout en nous assurant au maximum de la solidité de nos conclusions, nous avons utilisé, parallèlement, six méthodes différentes considérées comme fiables pour distinguer l'auteur d'un texte. Testées sur un corpus de trente grandes comédies en vers écrites après la mort de Corneille et Molière par six auteurs, trois de ces méthodes prédisent le bon auteur dans 100% des cas, deux dans 93% et une dans 87% des cas.

FRÉQUENCES DES MOTS

La première méthode consiste simplement à compter le nombre d'occurrences de chaque mot qui apparaît dans les textes étudiés. On considère le texte comme un « sac de mots ». Les fréquences d'apparition de chaque mot dans ce sac varient d'un texte à l'autre. La distribution de ces fréquences constitue une signature : à chaque auteur ses préférences quant au vocabulaire qu'il utilise.

On peut parfois considérer que certaines variations autour d'un même mot ne sont pas significatives. Utiliser « galère » ou « galères », « vois » ou « voit » peut paraître indifférent, ce qui compterait étant de dire « galère » plutôt que « navire », d'utiliser le verbe « voir » plutôt que « regarder ». La conjugaison, le pluriel, etc. ne seraient qu'accidentels par rapport au terme choisi. Dans ce cas, on travaille non pas sur le mot, mais sur le *lemme* : un usage de « vois », « voit » ou « voyons » est alors considéré dans le calcul comme un usage de « voir », un usage de « galères » comme un usage de « galère », et ainsi de suite. Mais cette ➤

> deuxième méthode consistant à étudier les lemmes, intéressante en soi, s'est révélée un peu moins fiable (93%) lors de notre phase de calibration que l'étude des mots.

Étudier le vocabulaire dans son intégralité peut toutefois générer certains biais. Notre emploi de certains mots dépend bien sûr de qui nous sommes, mais aussi de ce dont nous parlons. Si l'on étudie des textes d'auteurs au style proche, les thèmes abordés peuvent engendrer plus de variations dans le vocabulaire que les différences de style des écrivains. Or la langue du théâtre classique est extrêmement codifiée, en particulier pour les pièces écrites en vers. Dans notre cas, la crainte de variations liées aux sujets n'est donc pas infondée.

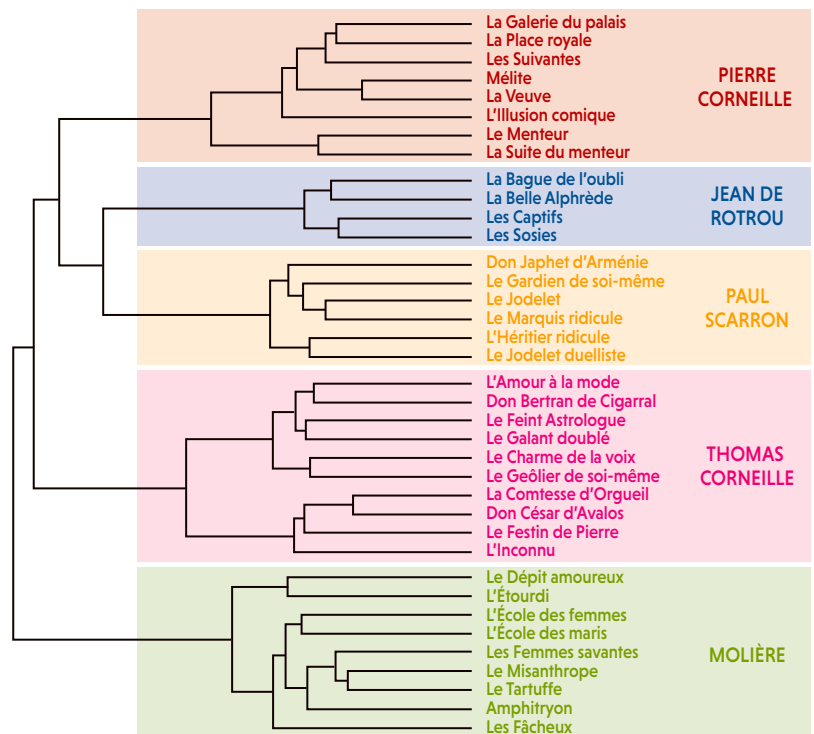
Deux solutions contraires sont alors envisageables. On peut choisir de ne s'intéresser qu'aux mots choisis avec le plus grand soin : ceux qui sont placés à la rime. Ils demandent une réflexion particulière et sont souvent très révélateurs de l'auteur d'un texte. Cette méthode axée sur les rimes a donné sur notre corpus test des résultats corrects à 87%. À l'inverse, on peut se focaliser sur les mots choisis le plus inconsciemment. Les *mots outils* («et», «de», «or», «mais»...) sont peu susceptibles de varier en fonction du sujet traité et ils sont particulièrement fréquents, ce qui permet d'obtenir des statistiques fiables. Ils ont longtemps été considérés comme l'indice le plus pertinent pour déterminer l'auteur d'un texte et, pour notre corpus test, les résultats étaient corrects à 93%.

ORDRE DES ÉLÉMENTS DANS LES PHRASES, SYNTAXE

Une cinquième méthode s'intéresse à l'ordre des éléments d'une phrase. Les phrases «Belle marquise, vos beaux yeux me font mourir d'amour» et «D'amour mourir me font, belle Marquise, vos beaux yeux» comportent exactement les mêmes mots, les mêmes lemmes, les mêmes mots outils. Les méthodes précédentes les considèrent donc comme parfaitement identiques. Pourtant, il est indéniable que le style diffère notablement entre ces deux tournures.

Pour conserver de l'information sur l'enchaînement des mots, on peut considérer le texte non comme un «sac de mots», mais comme un sac de «n-grammes de caractères». Cette dernière expression désigne simplement un enchaînement d'un nombre *n* de caractères. Supposons que l'on décompose les phrases précédentes en 5-grammes de caractères. En notant _ chaque espace, nous obtiendrions pour la première: Belle_ elle_, lle_m, le_ma, e_mar, _marq, etc. Et pour la seconde, nous aurions: D'amo_, amou_, mour_, etc. On réplique alors nos calculs en travaillant non sur des mots, mais sur ce type de séquences.

Dernière option explorée: la syntaxe des auteurs. Écrire des phrases uniquement selon



une structure simple ou, à l'inverse, selon des tournures particulièrement alambiquées, constitue un trait distinctif du style d'une personne – l'un des plus fiables avec les mots outils. Pour mesurer ces variations, on annote la nature grammaticale de chaque mot du texte, grâce à une intelligence artificielle spécialement entraînée sur des pièces de l'époque. Puis on étudie à quelle fréquence des successions de trois natures grammaticales (du type «pronon personnel / verbe conjugué / nom commun») apparaissent.

SI QUATRE ET QUATRE SONT HUIT, CORNEILLE N'A PAS ÉCRIT MOLIÈRE

Nous avons appliqué ces six méthodes à un ensemble de comédies en vers écrites par Molière, Corneille et leurs contemporains. L'analyse principale a porté sur 37 pièces de Pierre Corneille, son frère Thomas Corneille, Molière, Jean de Rotrou et Paul Scarron. En suivant chacune d'entre elles, toutes les pièces de Molière ont été considérées par nos algorithmes comme celles d'un même auteur, et clairement distinguées des pièces de Corneille – alors que d'autres auteurs, comme Scarron, semblent même plus proches stylistiquement de Molière.

Il paraît donc plus qu'improbable que Molière ait servi de prête-nom à Corneille ou à d'autres dramaturges célèbres de son temps. Faute de génie pouvant le détrôner dans les cœurs et les esprits, le français devrait donc rester encore pour quelque temps «la langue de Molière». ■

Cet arbre de parenté correspond aux résultats de la classification des textes en fonction de l'usage de mots outils (en vert les pièces de Molière, en rouge celles de Pierre Corneille, en rose Thomas Corneille, en jaune Paul Scarron, en bleu Jean de Rotrou). On constate que l'algorithme a parfaitement regroupé les pièces selon leur auteur putatif.

BIBLIOGRAPHIE

F. Cafiero et J.-B. Camps, **Why Molière most likely did write his plays**, *Science Advances*, vol. 5, article eaax548, 2019.

C. Labbé et D. Labbé, **Inter-textual distance and authorship attribution. Corneille and Molière**, *Journal of Quantitative Linguistics*, vol. 8(3), pp. 213-231, 2010.

D. Labbé, **Si deux et deux sont quatre, Molière n'a pas écrit Dom Juan**, Max Milo, 2009.

H. Wouters et C. de Ville de Goyet, **Molière ou l'auteur imaginaire ?**, Complexe, 1990.

ABONNEZ-VOUS À POUR LA SCIENCE

3 FORMULES
AU CHOIX

	FORMULE PAPIER	FORMULE PAPIER + HORS-SÉRIE	FORMULE INTÉGRALE
Le magazine papier 12 numéros par an	✓	✓	✓
Le magazine en version numérique 12 numéros par an			✓
Le hors-série papier 4 numéros par an		✓	✓
Le hors-série en version numérique 4 numéros par an			✓
Accès à pouirlascience.fr actus, dossiers, archives depuis 1996			✓
VOTRE TARIF D'ABONNEMENT	59€ Au lieu de 82,80€	79€ Au lieu de 114,40€	99€ Au lieu de 174,40€
	29 % de réduction *	31 % de réduction *	43 % de réduction *

BULLETIN D'ABONNEMENT

À renvoyer accompagné de votre règlement à :

Pour la Science – Service abonnements – 19 rue de l'Industrie – BP 90053 – 67402 Illkirch Cedex – email: pouirlascience@abopress.fr

PAG20STD



**OUI, je m'abonne
pour 1 an à**
**POUR LA
SCIENCE**

1 / Je choisis ma formule (merci de cocher)

☐ **FORMULE
PAPIER**

• 12 n° du magazine papier

59€
Au lieu de
82,80€

DIA59E

☐ **FORMULE PAPIER
+ HORS SÉRIE**

• 12 n° du magazine papier
• 4 Hors-série papier

79€
Au lieu de
114,40€

PIA79E

☐ **FORMULE
INTÉGRALE**

• 12 n° du magazine
(papier et numérique)
• 4 Hors-série
(papier et numérique)
• Accès illimité aux contenus en ligne

99€
Au lieu de
174,40€

IIA99E

2 / J'indique mes coordonnées

☐ M. ☐ Mme

Nom: Prénom:

Adresse:

Code postal [] [] [] [] [] [] Ville:

Téléphone [] [] [] [] [] [] [] [] [] [] [] [] [] [] [] []

Email : (indispensable pour la formule intégrale)

.....@.....

J'accepte de recevoir les offres de Pour la Science

☐ OUI ☐ NON

3 / Je choisis mon mode de règlement

☐ Par chèque à l'ordre de Pour la Science

☐ Carte bancaire

N° [] [] [] [] [] [] [] [] [] [] [] [] [] [] [] []

Date d'expiration [] [] [] [] Clé (Les 3 chiffres au dos de votre CB) [] [] []

Signature obligatoire :

* Réduction par rapport au prix de vente en kiosque et l'accès aux archives numériques. Durée d'abonnement : 1 an. Délai de livraison : dans le mois suivant l'enregistrement de votre règlement. Offre valable jusqu'au 31/03/2021 en France métropolitaine uniquement. Pour un abonnement à l'étranger, merci de consulter notre site <https://boutique.pouirlascience.fr>. Photos non contractuelles.

Les informations que nous collectons dans ce bulletin d'abonnement nous aident à personnaliser et à améliorer les services que nous vous proposons. Nous les utiliserons pour gérer votre accès à l'intégralité de nos services, traiter vos commandes et paiements, et vous faire part notamment par newsletters de nos offres commerciales moyennant le respect de vos choix en la matière. Le responsable du traitement est la société Pour la Science. Vos données personnelles ne seront pas conservées au-delà de la durée nécessaire à la finalité de leur traitement. Pour la Science ne commercialise ni ne loue vos données à caractère personnel à des tiers. Les données collectées sont exclusivement destinées à Pour la Science. Nous vous invitons à prendre connaissance de notre charte de protection des données personnelles à l'adresse suivante : <https://rebrand.ly/charte-donnees-pls>. Conformément à la réglementation applicable (et notamment au Règlement 2016/679/UE dit « RGPD ») vous disposez des droits d'accès, de rectification, d'opposition, d'effacement, à la portabilité et à la limitation de vos données personnelles. Pour exercer ces droits (ou nous poser toute question concernant le traitement de vos données personnelles), vous pouvez nous contacter par courriel à l'adresse protection-donnees@pouirlascience.fr.